GENE REGULATORY NETWORK INFERENCE VIA REGRESSION BASED TOPOLOGICAL REFINEMENT

JOCHEN SUPPER[*], HOLGER FRÖHLICH[*], ANDREAS ZELL

Centre for Bioinformatics Tübingen (ZBIT), Sand 1, 72076 Tübingen
jochen.supper@uni-tuebingen.de

Inferring the structure of gene regulatory networks from gene expression data has attracted a growing interest during the last years. Several machine learning related methods, such as Bayesian networks, have been proposed to deal with this challenging problem. However, in many cases, network reconstructions purely based on gene expression data not lead to satisfactory results when comparing the obtained topology against a validation network. Therefore, in this paper we propose an "inverse" approach: Starting from *a priori* specified network topologies, we identify those parts of the network which are relevant for the gene expression data at hand. For this purpose, we employ linear ridge regression to predict the expression level of a given gene from its relevant regulators with high reliability. Calculated statistical significances of the resulting network topologies reveal that slight modifications of the pruned regulatory network enable an additional substantial improvement.

1. Introduction

Transcriptional datasets provide valuable insights into specific cellular processes under various conditions. To control these processes the cell utilizes regulatory mechanisms, whereas for each specific process only a small fraction of the complete regulatory network is affected. Therefore, a gene regulatory network (GRN) is a large graph covering regulatory mechanisms for various stimuli. Yet for a specific observation only a small fraction of the GRN can be inferred and linked to the respective transcriptional data. Thus, the fixed topological structure of a GRN can be detached from the dynamic structure comprising a subset of the fixed topological structure associated with a quantitative observation.

Today, modeling of GRNs is guided by a rich flow of experimental data. The stream is still widened by an increasing pool of measurement techniques including mRNA microarray technology[17], chromatin immunoprecipitation (ChIP)[1], quantitative RT-PCR[7] and microarray-based immunoassays[13]. Despite of all this information, detailed knowledge regarding the topology of network models is still almost exclusively collected by biologists. They collect and integrate data, expand and refine their models and finally validate them. For our modeling efforts, we will use this qualitative knowledge provided by the biological observations to compile *a priori* topologies of the GRN. Within such topologies, we search for a subset of connections which is in good accordance to the transcriptional data and therefore prune the topological network using a reverse engineering approach.

---

[*]These authors contributed equally to this work.

2

Several of such reverse engineering approaches for GRNs have emerged during the last years. They include analytical methods such as Boolean networks[12], (non)-linear networks[23,21], S-Systems[20] and differential equations[3], but also machine learning methods such as decision trees[18] and Bayesian networks[10]. Most of these approaches have been applied solely to transcriptional data, thereby neglecting the *a priori* information available. Thus, they implicitly assume that the topological network can be connected arbitrarily, which leads to a huge number of potential regulatory mechanisms. However, some inference methods previously published consider *a priori* information. For instance Someren *et al.* compiled a validation network based on a co-citation approach[22], others have restricted the regulators to transcription factors[15], and, as in this work, the putative *cis*-regulatory elements have been employed as initial network structure[16].

Despite the use of *a priori* information, a set of transcriptional response measurements has to be available in order to reconstruct GRNs. Given this data, one of the above mentioned reconstruction methods can be employed to untangle the underlying topological structure of the interaction network. One of the principle problems thereby is the ambiguity inherent in reconstructing the GRN from given expression data, which is due to the small sampling rate along with the high noise level of the measurements. Even if the total number of microarray measurements is growing at a tremendous rate, the number of measurements utilizable for a specific observation is still limited. The bottom line is, that typical GRN inference methods produce topologies, which contain a certain fraction of false or unverifiable interactions. This fraction of spurious edges has to be traded off against the fraction of correctly identified regulatory dependencies. Husmeier *et al.*[10] systematically investigate this trade-off for a Bayesian network reconstruction of an artificially created network.

This motivated us to take an "inverse" approach to the usual GRN reconstruction methods: based on a given network topology from literature or by putative *cis*-regulatory elements (which may also contain wrong interactions), our approach is first, to identify the part of this network, which is relevant for the experimental data and second, to eventually modify this "pruned" network modestly to better fit the data. The basic assumption behind our approach is, that today, in many cases, a good starting network topology, which subsequently has to be refined in order to fit the experimental measurements, can be obtained from public databases or by the rich biological literature. In this work, the refinement is achieved through a machine learning based approach using linear ridge regression[9], which resembles the framework proposed by Soinov *et al.*[18]. In favor of our approach, we are able to show on publicly available Yeast genome datasets, that the prediction accuracy of gene expression levels is significantly higher in our fitted topology than in the original or a random network.

The remainder of this paper is organized as follows: in the next section, we introduce the datasets used and explain our approach in detail. In Section 3 we present and discuss the results obtained on our investigated datasets. Finally, in Section 4 the conclusions are drawn.

2. Materials and Methods

2.1. *Data*

2.1.1. *Budding Yeast Cell Cycle*

The biological model used for this research is the budding yeast cell cycle. This model displays a small subset of the whole network, where it is known that the chosen genes play an important role in the respective processes. The dataset was taken from Spellman *et al.*[19] and Cho *et al.*[4] , who measured the cell cycle of the budding yeast. The budding yeast cell cycle is known at a high level of detail. Spellman *et al.* and Cho *et al.* measured the progression of the cell cycle with different synchronization techniques. Overall they provided four records of time series measurements, which can be used for modeling purposes. The sampling number of these datasets ranges from 14 to 24, the samples are taken at equidistant time-steps for each series. However the time-steps are not equal for the different series, they range from 7 to 30 minutes. In this study we use the α factor, cdc 15 and cdc 28 arrest, as well as the elutriation time series. This results in 73 time point measurements altogether. These measurements have been taken on microarrays[5], each consisting of 6178 data genes. We imputed missing values in the gene expression measurements by the SVD method described[6].

2.1.2. *Chen Dataset*

Of the above described 6178 genes we chose a subset according to Chen *et al.*[2] , who used a set of differential equations to define the topology of the GRN. In addition to the interactions provided by the differential equations we searched TRANSFAC[24], Entrez Gene[14] and the *Saccharomyces* Genome Database (SGD[11]) for known dependencies between a pair of genes.

The entire network contains 56 interactions and is depicted in Figure 1. It will serve as our first *a priori* network for the reconstruction.

2.1.3. *Cis-Regulatory Elements*

The major control in transcriptional gene regulation is mediated by transcription factors (TFs) that bind to the promotor region of a gene. We used these connections between TFs and genes to construct the second validation network. In this network, the inferred genes are still restricted to the genes from the Chen *et al.* dataset, whereas the TFs are not. To establish these TF-gene connections, we extracted the *cis*-regulatory elements from the SCPD (The Promoter Database of *Saccharomyces cerevisiae*) database. This database was developed by Zhang *et al.*[25] and contains experimentally mapped TF binding sites as primary data entries and predefined putative regulatory elements using matrix and consensus methods. To extract the binding sites, we restricted the search to the 500 bp upstream sequence and searched for the consensus patterns contained.

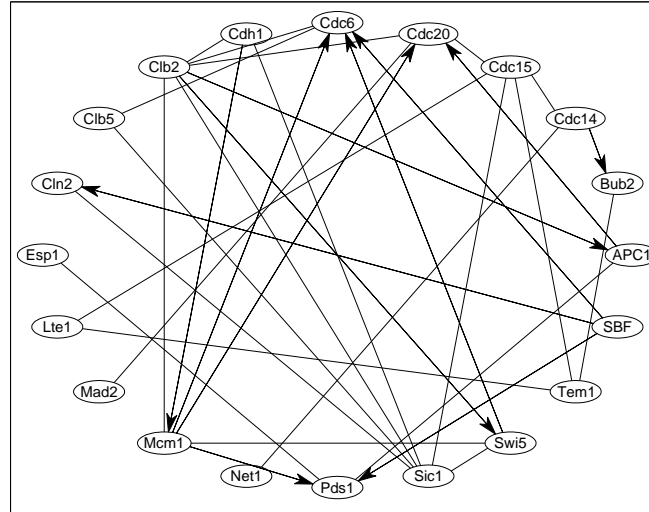The entire network contains 145 interactions and is depicted in Table 1.

4

Figure 1.   Literature network for the Chen data: An arrow $a \rightarrow b$ indicates that b is regulated by a. Edges with no arrows indicate a mutual influence $a \rightarrow b$ and $b \rightarrow a$.

Table 1.   Network composed by *cis*-regulatory elements.

Gene		Regulators
Cdc14	←	GCN4, TBP, INO1, PHO4, ABF1, Swi5, STE12
Cdc15	←	GCR1, BAS2, ABF1, GCN4, ADR1, MSE, YOX1, YHP1
Cdc20	←	YHP1, GCR1, RAP1, MSE, GCN4, STE12, YOX1, Swi5
Cdc6	←	GCR1, GCN4, MCB, Mcm1, Swi5, BAS2, TBP
Cdh1	←	MSE, GCR1, ADR1, GCN4, UASPHR, REB1, MSE
Clb2	←	AP-1, GCN4, GCR1, BAS2, GCN4, ADR1, TBP
Clb5	←	ACE2, Swi5, Mcm1, GCR1, GCN4, YOX1, YHP1, MSE, BAS2, AP-1
Cln2	←	STE12, YOX1, MSE, TBP, GCN4, BAS2, YHP1
Esp1	←	GCN4, PHO4, ABF1, YHP1, GCR1, BAS2, TBP, YOX1, RAP1, TBP
Lte1	←	GCN4, BAS2, ACE2, Swi5, TBP, ADR1, YOX1, SCB, YHP1, MCB, MATalpha2
Mad2	←	GCN4, TBP, Swi5
Mcm1	←	BAS2, GCN4, YHP1, YOX1, Swi5
Net1	←	YHP1, GCN4, YOX1, ABF1, REB1, STE12
Pds1	←	CPF1, PHO4, GCN4, TBP, YHP1, GCR1, MATalpha2, MSE, MCB, BAS2, YOX1
Sic1	←	GCN4, ABF1, BAS2, ADR1, Swi5, MCB, TBP, GCN4, MATalpha2, SCB
Swi5	←	GCN4, BAS2, Mcm1, GCR1, YHP1, YOX1, ADR1, PHO4, MATalpha2, ATF
Tem1	←	GCN4, ADR1, Swi5, GCR1, AP-1, YHP1, SCB, RAP1, Mcm1, YOX1, TBP
APC1	←	ADR1, GCR1, MIG1, MSE, YHP1, YOX1, Swi5, GCN4

2.2. *Regression Based Network Refinement*

Given one of the above described literature or *cis*-regulatory networks our goal is to fi t their topology to our datasets. This is done in a framework resembling that proposed by Soinov

et al.[18]: For each gene g we know a set of possible regulators $R_g = \{r_1, ..., r_{n_g}\}$. From these regulators we would like to select the subset $\tilde{R}_g$ of R_g, which allows the highest prediction performance of the expression level of g. Thereby, predicting the expression level of g can be either done within one time step t, or from time step $t - 1$ to t. We consider both prediction tasks and in the end merge the obtained regulator subsets $\tilde{R}_g^1$ and $\tilde{R}_g^2$ to the final subset $\tilde{R}_g$. We use linear ridge regression as the prediction machinery with ridge constant $\tau = 10^{-5}$.

To prune the "irrelevant" regulators for g, we adopt the RFE algorithm originally proposed for SVM feature selection[8]: In linear ridge regression one estimates a hyperplane $f : \mathbb{R}^{n_g} \to \mathbb{R}, \; f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b$, where $\mathbf{w}$ is the normal vector of the hyperplane and b a bias term. The components of the normal vector $\mathbf{w}$ can be understood as weights for the individual features in $\mathbf{x}$. Therefore, we can prune the regulator, for which the component in $\mathbf{w}$ is the smallest. Then the regression function is re-estimated and the whole procedure iterated until all regulators are removed. Overall, we receive a ranking of all regulators depending on their time of removal from the regulator set R_g. The optimal number of regulators in our case is determined via 5-fold cross-validation, where we use the mean squared correlation between predicted and true gene expression values to measure the prediction performance.

After pruning the original set of regulators we allow to add one extra regulator, which was not in R_g before, if this further improves the 5-fold cross-validated mean squared correlation. This slight modification of the pruned network takes into account that there might exist interactions, which are not covered by the *apriori* provided network structure and otherwise could not be detected. Nevertheless, the selected subset $\tilde{R}_g^i$ $(i = 1, 2)$ of regulators, which typically consists only of a few genes, is completely rejected, if the mean 5-fold cross-validated correlation between the true and the predicted expression levels of g is below 60%. This ensures that only interactions with a high statistical confidence are inserted into the network.

As a last step the final regulator subset $\tilde{R}_g = \tilde{R}_g^1 \cup \tilde{R}_g^2$ is evaluated with respect to the two prediction tasks described in the beginning. This is done via 5-fold cross-validation, measuring the mean squared correlation between the predicted and the true expression level of g.

2.3. *Comparison Scheme*

The bottom line from the above described approach is that for all genes in the inferred network we obtain an estimate, how well their expression level can be predicted from their putative regulators. This can be viewed as a measure for the reliability of the network. Which allows us to compare different network topologies with respect to this measure and to compute statistical significances. More specifically, we are interested, how well a network computed with the method described in the last subsection performs relative to the following reference topologies:

- the original literature and *cis*-regulatory network

6

- a random network with the same number of connections as the literature network
- a fully connected network

Comparing the prediction performances obtained from these network topologies to our refined network allows the computation of p-values, via Wilcoxon's signed rank test. In the following section we describe the results we achieved this way for the two datasets investigated in this paper.

3. Results

3.1. *Chen Dataset*

Our method from subsection 2.2 yielded 6 genes with a nonempty regulator set. It is depicted in Figure 2a). For comparison reasons we also computed the network, which was obtained by just pruning the literature network and not introducing any further interactions. It had only 5 genes with a nonempty regulator set and is shown in Figure 2b). It is worth mentioning that both networks can differ in more than one regulator per gene: If in the modification step an additional regulator is added to the pruned regulator set, the prediction performance of the prediction model can increase such that it exceeds the minimum prescribed threshold of 60% mean squared correlation between predicted and true gene expression values. Hence, the whole regulator set is added to the pruned network at once in this case.

In Table 2 we compare the resulting average 5-fold cross-validated mean squared correlations for these genes with those obtained from the literature, the random and the fully connected network (see last section). As shown, the pruned network with additional modifications yielded a high significant improvement compared to the random and the literature network. In contrast, the pruned network without additional modifications had a much worse p-value compared to the random network and was not significantly better than the original literature network. A direct comparison of the pruned networks with and without additional modifications reveals a high statistical significant difference between both (p-value = 0.0059).

Table 2.   5-fold cross-validated correlations (r) with true expression levels for genes with nonempty regulator sets (Chen dataset). For the inferred networks (pruned+modification, pruned only) the p-values are calculated as described in subsection   2.3.

Topology	$r \pm$std. err.	p-val. lit.	p-val. rand.	p-val. full
pruned+mod.	**72.07 $\pm$ 1.94**	**9.76 $\cdot$ 10^{-4}**	**0.0024**	0.4697
pruned only	65.74 $\pm$ 3.77	0.1953	0.0645	0.2324
literature	42.52 $\pm$ 2.76			
random	39.58 $\pm$ 2.55			
full	51.87 $\pm$ 2.83			

(a) with pruning only   (b) with pruning and modification
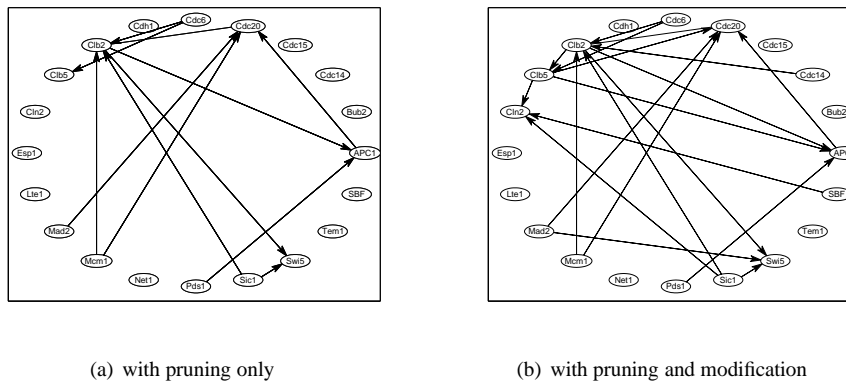
Figure 2.   Inferred networks for the Chen data.

Table 3.   Network inferred by pruning the *cis*-regulatory elements.

Gene		Regulators
Cdc14	←	Swi5, ABF1, STE12, PHO4
Cdc15	←	—
Cdc20	←	GCN4, Swi5, GCR1, MSE, YOX1, YHP1
Cdc6	←	—
Cdh1	←	UASPHR
Clb2	←	—
Clb5	←	GCN4, GCR1, BAS2, MSE, YOX1, Mcm1, AP-1, ACE2
Cln2	←	GCN4, TBP, STE12, BAS2, MSE, YOX1, YHP1
Esp1	←	—
Lte1	←	—
Mad2	←	—
Mcm1	←	—
Net1	←	—
Pds1	←	PHO4, GCR1, BAS2, YOX1, YHP1, MCB
Sic1	←	—
Swi5	←	GCN4, BAS2, ADR1, YOX1, YHP1
Tem1	←	TBP, Swi5, SCB
APC1	←	ADR1, Swi5, GCR1

3.2. *Cis-Regulatory Elements*

Our method from subsection 2.2 on this dataset yielded 11 genes with a nonempty regulator set (Table 4). Again, we also computed the network, which was obtained by just pruning the initial network network without introducing any further interactions. It had only 9 genes with a nonempty regulator set and is shown in Table 3. Like in the last subsection, in Table 4 we compared the resulting average 5-fold cross-validated mean squared correlations for these genes with those obtained from the *cis*-regulatory, the random and the fully connected network. As seen, the pruned network with additional modifications yielded a high significant improvement compared to the random, the *cis*-regulatory network and the full network. The pruned network without additional modifications in all cases had much

8

Table 4.   Network inferred by pruning the *cis*-regulatory elements network and some allowed subsequent modifications. Regulators introduced additionally to the pruned network are marked bold. Regulators, which were not in the *a priori* network are written italic as well.

Gene		Regulators
Cdc14	←	**TBP**, PHO4, ABF1, Swi5, STE12, ***RAP1***
Cdc15	←	—
Cdc20	←	GCN4, ***APC1***, Swi5, GCR1, MSE, YOX1, YHP1
Cdc6	←	—
Cdh1	←	UASPHR, ***Sic1***
Clb2	←	**GCR1, BAS2, ADR1**, ***REB1***
Clb5	←	GCN4, GCR1, BAS2, MSE, YOX1, ***Cdc6***, Mcm1, AP-1, ACE2
Cln2	←	GCN4, TBP, STE12, BAS2, MSE, YOX1, YHP1, ***AP-1***
Esp1	←	—
Lte1	←	—
Mad2	←	—
Mcm1	←	—
Net1	←	—
Pds1	←	PHO4, GCR1, BAS2, YOX1, YHP1, MCB, ***Tem1***
Sic1	←	**GCN4, TBP, ABF1, Swi5, BAS2**, ***Cdc6***, **SCB, MATalpha2**
Swi5	←	GCN4, BAS2, ADR1, YOX1, YHP1, ***ACE2***
Tem1	←	TBP, Swi5, **GCR1, AP-1**, SCB, ***Sic1***
APC1	←	ADR1, Swi5, GCR1, ***INO1***, **MSE, MIG1**

worse p-values, especially compared to the fully connected network. A direct comparison of the pruned networks with and without additional modifications reveals a high statistical significant difference between both (p-value = 0.0036).

4. Conclusion

We introduced a method to refine a GRN topology obtained from the literature or from public databases such that it fits a given gene expression dataset. Thereby our criterion was the estimated generalization performance achieved by a linear ridge regression model, which was trained to predict the expression level of each gene in the network from the expression levels of its regulators. An algorithm was developed to find a minimal regulator subset for each gene, which allows the highest prediction rate. Thereby slight modifications of the pruned literature network were allowed in order to take into account the possible incompleteness or defectiveness of the original network topology.

We performed evaluations of our method on publicly available datasets from Yeast genome and compared our approach against the original *a priori* network, a random network with the same number of interactions as the *a priori* network and a fully connected network. We were able to show that our inferred networks on both datasets could significantly improve on the original, the random, and in case of the second dataset, also on the fully connected dataset. Furthermore, an interesting finding was that allowing slight modifications of the pruned *a priori* network in all cases lead to much better p-values than without allowing these changes.

Altogether we think that a main contribution of this work was first, the introduction of

a network refinement method from a given starting topology, and second the possibility to compute statistical significances for the inferred network, which to our knowledge has not been possible so far.

Table 5.   5-fold cross-validated correlations (r) with true expression levels for genes with nonempty regulator set (*cis*-regulatory elements).   For the inferred networks (pruned+modification, pruned only) the p-values are calculated as described in subsection   2.3.

Topology	$r\pm$std. err.	p-val. *cis*-reg.	p-val. rand.	p-val. full
pruned+mod.	$\mathbf{73.85 \pm 2.71}$	$\mathbf{5.96 \cdot 10^{-5}}$	$\mathbf{4.61 \cdot 10^{-5}}$	$\mathbf{2.14 \cdot 10^{-4}}$
pruned only	68.3 ± 4.09	0.0181	$3.86 \cdot 10^{-4}$	0.0312
cis-reg.	52.04 ± 3.32			
random	45.13 ± 2.84			
full	50.87 ± 2.81			

References

1. Michael J Buck and Jason D Lieb. ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics*, 83(3):349–60, Mar 2004.
2. Katherine C Chen, Laurence Calzone, Attila Csikasz-Nagy, Frederick R Cross, Bela Novak, and John J Tyson. Integrative analysis of cell cycle control in budding yeast. *Mol Biol Cell*, 15(8):3841–3862, Aug 2004.
3. Kuang-Chi Chen, Tse-Yi Wang, Huei-Hun Tseng, Chi-Ying F Huang, and Cheng-Yan Kao. A stochastic differential equation model for quantifying transcriptional regulatory network in Saccharomyces cerevisiae. *Bioinformatics*, 21(12):2883–90, Jun 2005.
4. R. J. Cho, M. J. Campbell, E. A. Winzeler, L. Steinmetz, A. Conway, L. Wodicka, T. G. Wolfsberg, A. E. Gabrielian, D. Landsman, D. J. Lockhart, and R. W. Davis. A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol Cell*, 2(1):65–73, Jul 1998.
5. M. B. Eisen and P. O. Brown. DNA arrays for analysis of gene expression. *Methods Enzymol*, 303:179–205, 1999.
6. S. Friedland, A. Niknejad, and L. Chihara. A simultaneous reconstruction of missing data in dna microarrays. *Linear Algebra Appl.*, 2005. to appear.
7. David G Ginzinger. Gene quantification using real-time quantitative PCR: an emerging technology hits the mainstream. *Exp Hematol*, 30(6):503–12, Jun 2002.
8. I. Guyon, J. Weston, S. Barnhill, and V. Vapnik. Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning*, 46:389 – 422, 2002.
9. A. Hoerl and R. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55 – 67, 1970.
10. Dirk Husmeier. Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks. *Bioinformatics*, 19(17):2271–2282, Nov 2003.
11. Laurie Issel-Tarver, Karen R Christie, Kara Dolinski, Rey Andrada, Rama Balakrishnan, Catherine A Ball, Gail Binkley, Stan Dong, Selina S Dwight, Dianna G Fisk, Midori Harris, Mark Schroeder, Anand Sethuraman, Kane Tse, Shuai Weng, David Botstein, and J. Michael Cherry. Saccharomyces Genome Database. *Methods Enzymol*, 350:329–346, 2002.

10

12. S. Liang, S. Fuhrman, and R. Somogyi. Reveal, a general reverse engineering algorithm for inference of genetic network architectures. *Pac Symp Biocomput*, 1:18–29, 1998.

13. Gavin MacBeath. Protein microarrays and proteomics. *Nat Genet*, 32 Suppl:526–32, Dec 2002.

14. Donna Maglott, Jim Ostell, Kim D Pruitt, and Tatiana Tatusova. Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res*, 33(Database issue):D54–D58, Jan 2005.

15. D. Pe'er, A. Tanay, and A. Regev. Minreg: A scalable algorithm for learning parsimonious regulatory networks in yeast and mammals. *J. Machine Learning Research*, 7:167 – 189, 2006.

16. Jianhua Ruan and Weixiong Zhang. A bi-dimensional regression tree approach to the modeling of gene expression regulation. *Bioinformatics*, Nov 2005.

17. A. Schulze and J. Downward. Navigating gene expression using microarrays–a technology review. *Nat Cell Biol*, 3(8):E190–5, Aug 2001.

18. Lev A Soinov, Maria A Krestyaninova, and Alvis Brazma. Towards reconstruction of gene networks from expression data by supervised learning. *Genome Biol*, 4(1):R6, 2003.

19. P. T. Spellman, G. Sherlock, M. Q. Zhang, V. R. Iyer, K. Anders, M. B. Eisen, P. O. Brown, D. Botstein, and B. Futcher. Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. *Mol Biol Cell*, 9(12):3273–3297, Dec 1998.

20. Christian Spieth, Felix Streichert, Nora Speer, and Andreas Zell. A memetic inference method for gene regulatory networks based on s-systems. In *Proceedings of the IEEE Congress on Evolutionary Computation (CEC 2004)*, pages 152–157, 2004.

21. J. Supper, C. Spieth, and A. Zell. Reverse engineering non-linear gene regulatory networks based on the bacteriophage lambda ci circuit. In *IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, 2005.

22. E. P. van Someren, B. L T Vaes, W. T. Steegenga, A. M. Sijbers, K. J. Dechering, and M. J T Reinders. Least Absolute Regression Network Analysis of the murine osteoblast differentiation network. *Bioinformatics*, Dec 2005.

23. D. C. Weaver, C. T. Workman, and G. D. Stormo. Modeling regulatory networks with weight matrices. *Pac Symp Biocomput*, 1:112–23, 1999.

24. Edgar Wingender. TRANSFAC, TRANSPATH and CYTOMER as starting points for an ontology of regulatory networks. *In Silico Biol*, 4(1):55–61, 2004.

25. J. Zhu and M. Q. Zhang. SCPD: a promoter database of the yeast Saccharomyces cerevisiae. *Bioinformatics*, 15(7-8):607–611, 1999.