A Quantitative Framework of Transcriptional Dynamics by Integrating Multiple Sources of Knowledge

Shu-Qiang Wang*
Department of Manufacturing Engineering and Engineering Management
City University of Hong Kong,
Hong Kong
*E-mail:shuqiwang2@student.cityu.edu.hk

Han-Xiong Li
Department of Manufacturing Engineering and Engineering Management
City University of Hong Kong,
Hong Kong

***Abstract*--A key challenge in the post genome era is to identify genome-wide transcriptional regulatory networks, which specify the interactions between transcription factors and their target genes. In this work, a regulatory model based binding energy is proposed to quantify the transcriptional regulatory network. Multiple quantities, including binding affinity and the activity level of transcription factor (TF) are incorporated into a general learning model. The sequence features of the promoter and the possible occupancy of nucleosomes are exploited to estimate the binding probability of regulators. Comparing with the previous models that only employ microarray data, the proposed model can bridge the gap between the relative background frequency of the observed nucleotide and the gene's transcription rate. Experimental results show that the proposed model can effectively identify the parameters and the activity level of TF. Moreover, the kinetic parameters introduced in the proposed model can reveal more biological sense than some previous models can do.**

Keywords-Ttranscriptional dynamics; Sequence feature; Transcription rate; Bayesian inference;

1. INTRODUCTION

A challenge facing molecular biology is to develop quantitative, predictive models of gene regulation. The advance of high-throughput microarray technique makes it possible to measure the expression profiles of thousands of genes, and genome-wide microarray datasets are collected, providing a way to reveal the complex regulatory mechanism among cells. There are two broad classes of gene regulatory interactions: one based on the 'physical interaction' that aim at identifying relationships among transcription factors and their target genes (gene-to-sequence interaction) and another based on the 'influence interaction' that try to relate the expression of a gene to the expression of the other genes in the cell (gene-to-gene interaction).

In recent years, researchers have proposed many different computational approaches to reconstruct gene regulatory networks from high-throughput data, e.g. see reviews by Bansal et al. and Markowetz and Spang [1, 2]. These approaches fall roughly into two categories: qualitative and quantitative aspects. Inferring qualitative regulatory networks from microarray data has been well studied, and a number of effective approaches have been developed [3-10]. However, these methods are based on coarse grained qualitative models [11, 12], and cannot provide a realistic and quantitative view of regulatory systems.

On the other hand, quantitative modeling for gene regulatory network is in its infancy. Research on quantitative models for genetic regulation has arisen only in recent years, and most of them are based on classical statistical techniques. Liebermeister et al. [13] proposed a linear model for cell cycle-related gene expression in yeast based on independent component analysis. Holter et al. [14] employ singular value decomposition to uncover the fundamental patterns underlying gene expression profiles. Pournara et al. [15] and Yu et al. [16] proposed the Factor analysis model to describe a larger number of observed variables. However, these approaches are based on linear regression, and are not always being consistent with the observations in biochemical experiments which are nonlinear. Imoto et al. [17] proposed a nonlinear model with heterogeneous error variances. This model matches the microarray data well but it is not satisfying enough in revealing more biological sense. Segal et al. [18] proposed a transcription control network based model and apply their model to the segmentation gene network of Drosophila melanogaster. They reveal that positional information is encoded in the regulatory sequence and input factor distribution. However, there is still a little bit of dilemma in the model: the activity level of transcription factors is difficult to be measured or to be identified. Actually, assaying the transcription factors' activity state in a dynamic fashion is a major obstacle to the wider application of the kinetic modeling. TFs' activity levels are difficult to measure mainly due to two technical limitations: TFs are often present at low intercellular concentrations and the changes in their activity state can occur rapidly due to post-translational modifications.

Based on the above description, this paper aims to describe the transcriptional regulatory network quantitatively. In this work, a Bayesian inference based regulatory model is proposed to quantify the transcriptional dynamics. The model

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

relies on a continuous time, differential equation description of transcriptional dynamics where TFs are treated as latent on/off variables and are modeled using a switching stochastic process. Multiple quantities, including binding energy, binding affinity and the activity level of transcription factor (TF) are incorporated into a general learning model. The sequence features of the promoter and the occupancy of nucleosomes are exploited to derive the binding energy. Compared with the previous models, the proposed model can reveal more biological sense.

2. PROBLEM STATEMENT

A microarray experiment only measures the "observed" quantities, as shown in fig. 1, whereas the other quantities are not observed ("hidden"). The dashed oval encloses the closest quantities on the path between the TF and the target gene.

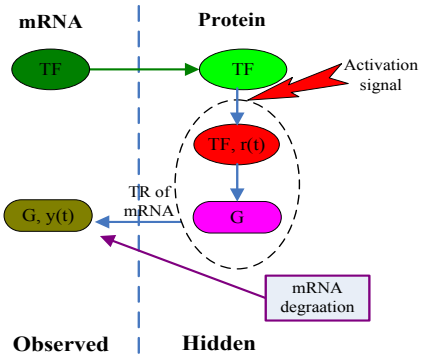

Figure 1: A qualitative molecular model of transcriptional regulation. mRNA encoding a transcription factor (TF) is translated to protein. The protein is activated and induces the transcription of a target gene at a certain rate (G). The final accumulation of G mRNA levels is determined by this transcription rate and by the rate of G's mRNA degradation.

Consider a transcriptional network of *n* genes that are regulated by *m* regulators, as well as a time-dependent external signal. Given the structure G and a set X of transcription rates of these genes in T time points, is it possible to reconstruct the time-varying activity level of *m* regulators, *R,* at all time points and the corresponding model parameters? This is an infinite dimensional problem that we tackle by placing a stochastic process prior over the activities of regulators.

Our approach relies on a continuous time, differential equation description of transcriptional dynamics where TFs are treated as latent on/off variables and are modeled using a switching stochastic process. The framework of the proposed method is shown in the fig. 2.

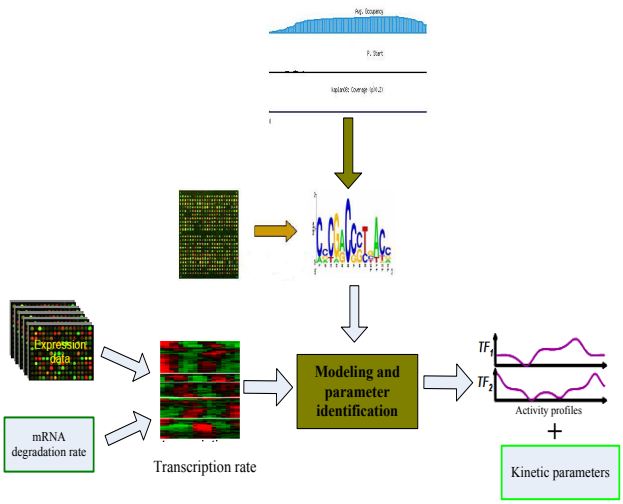

Figure 2: Flow of proposed method.

3. METHODS

3.1 Kinematic model of regulation

Compared with the gene expression level, the gene transcription rate can capture more dynamic characteristics of transcription regulation. We here employ the transcription rate to model the regulation function. We first assume:

- The derived transcription rates are average rates over a cell population.
- The speed of a TF's binding to or dissociation from its target sites is assumed to be much more rapid than the transcription process, thus rapid-equilibrium approximation can be used.

Based on the above assumptions, the transcription rate of a gene is proportional to the amount of the gene bound by its regulators in all genes of the measured cell population. We first consider the case that a gene is regulated by a single activator. The corresponding regulation function can be properly described by Michaelis–Menten equation:

$$\frac{dx}{dt} = \beta \frac{r(t)}{d + r(t)} + c - \lambda x, \qquad (1)$$

here x represents the mRNA concentration for a particular gene, *r(t)* the concentration of active TF, β and d are kinetic constants, c a baseline expression rate and λ the mRNA decay rate.

To incorporate the sequence feature and the TF binding preference into the model, we set the binding affinity $\psi = 1/d$, and (1) can be reformulated as

$$\frac{dx}{dt} = \beta \frac{k\varphi r(t)}{1 + k\varphi r(t)} + c - \lambda x, \qquad (2)$$

here *k* is a scaling parameter.

We now take the regulation involving two regulators in to account. Denote by $r_1(t)$ and $r_2(t)$ the concentration of two regulators, ψ_1 and ψ_2 the binding affinity of two regulators from their own target sites, the regulation function can be written as below:

$$\frac{dx}{dt} = \beta \frac{k_1\varphi_1 r(t)_1 + k_2\varphi_2 r_2(t) + k_3\varphi_1\varphi_2 r_1(t) r_2(t)}{\left(1 + \varphi_1 r_1(t)\right)\left(1 + \varphi_2 r_2(t)\right)} + c - \lambda x \qquad (3)$$

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

Considering the general case, a gene is regulated by n regulators. There are 2^n different binding states in total. The n-dimension binary vector is employed to indicate a certain binding state, e.g., a 4-dimension vector (0 1 0 1) indicates that the second and the fourth regulators are bound to their own target sites while the first and the third are not bound. The regulation function can be written as:

$$\frac{dx_j}{dt} = \beta_j \frac{\sum_{s\in S_j} k_s \prod_{s_i=1;i=1,n} \varphi_{ij} r_i(t)}{\prod_{i=1}^{n}(1+\varphi_{ij} r_i(t))} + c_j - \lambda_j x_j \qquad (4)$$

where S_j denotes the set of all 2^n possible state vectors, and s_i is the i_{th} element of the state vector s.

3.1.1 Models for binding affinity

Measuring affinities of molecular interactions in high-throughput format remains problematic, especially for transient and low-affinity interactions. We here try to describe the landscape of binding affinity in the perspective of binding energy between the various DNA-binding molecules and their target genes. Binding affinity landscapes describe how each molecule translates an input DNA sequence into a binding potential that is specific to that molecule. The presented framework models several important aspects of the binding process.

By allowing molecules to bind anywhere along the input sequence, the entire range of affinities is considered, thereby allowing contributions from both strong and weak binding sites [19, 20].

- Conventional cooperative binding interactions can be explicitly modeled by assigning higher statistical weights to configurations in which two molecules are bound in close proximity.
- The cooperativity that arises between factors when both nucleosomes and transcription factors are integrated is captured automatically [21].

We first consider the simplest case that there is only one target site S_{ij} for TF i in the promoter of gene j:

$$TF_i + S_{ij} \xleftrightarrow[k_d]{k_b} [TF_i \cdot S_{ij}]$$

The site-specific binding affinity is given by

$$\varphi = C_i e^{\frac{E_{ij}}{kT}} \qquad (5)$$

where C_i is a constant , E_{ij} the binding free energy between TF i and the promoter of gene j, k and T are the Boltzmann constant and temperature, respectively.

The above case can be expanded to the general case that binding may happen in anywhere along the input sequence and the accessibility of target sites varies due to the occupancy of nucleosomes. The general binding affinity is modeled as

$$\varphi_{ij} = C_i \sum_{n=1}^{N} p_{ij}^{(n)} e^{\frac{-E_{ij}^{(n)}}{kT}} \qquad (6)$$

where $p^{(n)}_{ij}$ is the probability of transcription factor i binding to the n_{th} binding site in the promoter of gene j. Note that the derivation of $p^{(n)}_{ij}$ involves the information of the possible occupancy of nucleosomes. The nucleosomes positions can be predicted based on the nucleosome-DNA interaction model proposed by Kaplan et al [22]. Fig.3 (b) shows the occupancy of nucleosomes for the genomic sequence shown in the Fig. 3 (a).

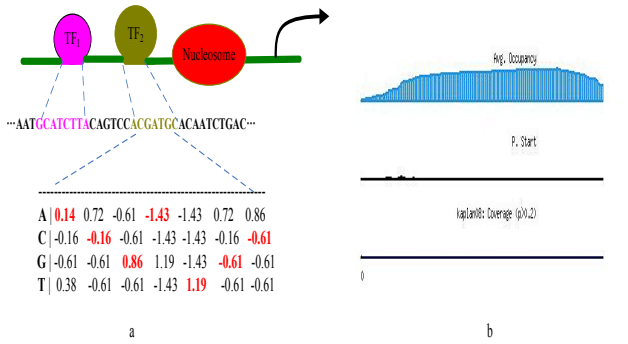

Figure 3. Employing sequence features and the occupancy of nucleosomes to estimate the binding affinity.

Since the positional weight matrices (PWM) are often used to represent the sequence motif [23, 24], we estimate the binding energy in perspective of PWM. As the background information has been taken into the PWM, the binding free energy can be approximately calculated as below:

$$E_{ij}^{(q)} = K^{(q)} \sum_{l=1}^{L} \sum_{n=\{A,C,G,T\}} J^n{}_l \left(M_L^* - M_{nl} \right)$$

where $J^n{}_l = \begin{cases} 1 & if\ n = s(l) \\ 0 & \text{otherwise} \end{cases}$

Here $K^{(q)}$ is the scaling factor, M^*_L indicates the maximum background frequency in the motif, $s(l)$ is the nucleotide in position l.

3.1.2 Switching stochastic process based description of TF activity

In many biological processes, the transcription factor transit from inactive to active state as a consequence of fast post-translational modifications, so it is reasonable that we model the TF activity as a binary variable $r(t) \in \{0,1\}$.

For the regulation involving a single regulator, the TF activity can be seen as a two states Markov Jump Process [29]. Given transition rates $n_{\pm}$ for the process, the probability of the system being in a particular state at a given time is given by the following Master equation:

$$\frac{dp_1(t)}{dt} = n_+ p_0(t) - n_- p_1(t) \qquad (7)$$

$$\frac{dp_0(t)}{dt} = n_+ p_1(t) - n_- p_0(t) \qquad (8)$$

here $p_1(t)=p(r(t)=1)$ and analogously for $p_0(t)$.

For the regulation involving two regulators, the probability of the system being in a particular state at a given time is given by the following Master equations:

$$\begin{aligned}
\frac{dp_{11}(t)}{dt} &= n_+^{\,1} p_{01}(t) - n_-^{\,1} p_{11}(t) + n_+^{\,2} p_{10}(t) - n_-^{\,2} p_{11}(t) + n_+^{\,3} p_{00}(t) - n_-^{\,3} p_{11}(t) \\
\frac{dp_{10}(t)}{dt} &= n_+^{\,4} p_{00}(t) - n_-^{\,4} p_{10}(t) + n_-^{\,2} p_{11}(t) - n_+^{\,2} p_{10}(t) + n_+^{\,5} p_{01}(t) - n_-^{\,5} p_{10}(t) \\
\frac{dp_{01}(t)}{dt} &= n_+^{\,6} p_{00}(t) - n_-^{\,6} p_{01}(t) + n_-^{\,1} p_{11}(t) - n_+^{\,1} p_{01}(t) + n_-^{\,5} p_{10}(t) - n_+^{\,5} p_{01}(t)
\end{aligned} \qquad (9)$$

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

here $p_{11}(t)=p(r_1(t)=1, r_2(t)=1)$ and analogously for $p_{10}(t)$, $p_{01}(t)$ and $p_{00}(t)$, $n_{\pm}^{i}$ (i=1, 2…) indicates the corresponding transition rate.

3.2. Regulatory network modeling using dynamic Bayesian inference

To set up a Bayesian inference framework, we define the TF's switching stochastic process as the prior distribution. Notice that the stochastic process prior on the TFs implies that the mRNA concentrations are also a stochastic process, even though the relationship between *x* and r is entirely deterministic. The prior distribution then is combined with an observation model (likelihood) that relates the observed variables to the hidden variables. In this case, we model the observations $y_j(t)$ of mRNA species *j* at time *t* as normally distributed around the value of the random variable x(t):

$$y_j(t)\big|r(t) \square N(x_j(t),\ \sigma^2{}_j)$$

Given a prior model and a likelihood, we can then combine these into Bayes' theorem to obtain the posterior over the process as

$$p(r\big|y,\ \Omega)=\frac{1}{S}p(y\big|r,\ \Omega)p(r),$$

Where y denotes collectively the observations, Ω are the parameters involved in the regulation function and S a normalization constant.

3.3 Variational inference and Model optimization

We will use a variational formulation of the inference problem [25].Variational inference is a powerful inference method based on tools from optimization. Free form (i.e. unconstrained) variational inference is entirely equivalent to the general inference problem. Variational inference is used as an approximation technique: given an intractable probability distribution p, the variational approach finds an optimal approximation q within a certain family of distributions. This is usually done by minimizing the Kullback–Leibler (KL) divergence between the two distribution

$$KL[q\|p]=E_q[\log\frac{q}{p}]=\int\log\frac{q(r)}{p(r)}q(r)dr \qquad (10)$$

By selecting a suitable family of approximating distributions, the inference problem is then turned into an optimization problem. It can be shown that the KL divergence is a convex functional of q and is equal to zero iff q=p [26]. In this case, we will choose the approximating process q to be again a Markov Jump Process, so that the required KL is given by

$$KL[q\|p_{post}]=KL[q\|p_{prior}]+\log S-E_q[\log p(y\big|r,\ \Omega)] \qquad (11)$$

here S is a normalization constant, $E_q[\log p(y\big|r,\ \Omega)]$ the expectation of the likelihood of the observations under the approximating process. For the regulation involving a single regulator, the KL divergence between the prior process and the approximating $KL[q\|p_{prior}]$ is given by

$$KL[q\|p_{prior}]=\int_0^T\left(m_+\ln\frac{m_+}{n_+}+n_+-m_+\right)q_0(t)dt+\int_0^T\left(m_-\ln\frac{m_-}{n_-}+n_--m_-\right)q_1(t)dt \qquad (12)$$

Here m_+ is the transition rate of jumps form the 0 to 1 state for process q, and analogously for m_-. A derivation for $KL[q\|p_{prior}]$ is given in the appendix. Therefore, the inference problem can be turned into an optimization problem.

By direct computation, minimization of the KL functional (11) can be represented as the saddle point problem

$$J=\max_{\tau}\min_{q}\{KL[q\|p_{prior}]+\sum_{j=1}^{n}[\tau_j(y_j-\bar{x}(t_j))-\frac{\sigma^2}{2}\tau_j{}^2]\}, \qquad (13)$$

here τ is auxiliary variables (one for each observation). By inspection and using the properties of the KL divergence, we can find that this functional is concave in τ and convex in q. Hence we can exchange min and max. Performing the max first yields the result. This also shows that there is only a unique saddle point solution.

Solving (2) and Submitting (12) into (13), we get

$$J=\max_{\tau}\min_{q}\{\int_0^T\left(m_+\ln\frac{m_+}{n_+}+n_+-m_+\right)q_0(t)dt+\int_0^T\left(m_-\ln\frac{m_-}{n_-}+n_--m_-\right)q_1(t)dt$$
$$-\sum_{j=1}^{n}[\tau_j\int_0^t[\beta\frac{k\varphi q(s)}{1+k\varphi q(s)}+c]e^{\lambda(s-t)}ds-(\tau_j y_j-\frac{\sigma^2}{2}\tau_j{}^2)]\} \qquad (14)$$

For the regulation involving two regulators (extension to i>2 is conceptually straightforward), the KL divergence between the true posterior and the approximating process is given by

$$J=KL[q^1\|p]+KL[q^2\|p]-\sum_{j=1}^{n}\int_0^T[\beta\frac{k_1\varphi_1q^1{}_1(t)+k_2\varphi_2q^1{}_2(t)+k_3\varphi_1\varphi_2q^1{}_1(t)q^1{}_2(t)}{\left(1+\varphi_1q^1{}_1(t)\right)\left(1+\varphi_2q^1{}_2(t)\right)}+c^j]U^j(t)dt$$

with

$$U^j(t)=\sum_i\tau^j{}_ie^{\lambda^j(t-t_i)\Theta(t_i-t)}$$

And Θ is the Heaviside step (j indexes the genes, i the time points). The KL divergence between two Markov Jump Processes is given by

$$KL[q\|p]=\int_0^T[q_1(m_-\ln\frac{m_-}{n_-}+n_+)]dt+\int_0^T[(1-q_1)(m_+\ln\frac{m_+}{n_+}+n_+)]dt \qquad (15)$$

The optimization procedure is based on a forward-backward procedure, leading to ordinary differential equations which can iteratively be solved. The free energy is a functional of both the approximating processes q^1, q^2 and their transition rates m_1, m_2. However, these are not independent, but are related by the Master equation. To incorporate this constraint, we add Lagrange multipliers as

$$L(q^1,q^2,g_1,g_2)=J[q^1,q^2,g_1,g_2]+\int_0^T[\frac{dq_1{}^1(t)}{dt}+(m_{1-}+m_{1+})q_1{}^1(t)-m_{1+}]\lambda_1(t)dt$$

$$+\int_0^T[\frac{dq_2{}^1(t)}{dt}+(m_{2-}+m_{2+})q_2{}^1(t)-m_{2+}]\lambda_2(t)dt \qquad (16)$$

The Lagrange multiplier functions obeys the final condition λ(T) = 0. Now we can optimize the objective function by setting to zero its functional derivatives. The details can be found in the appendix. Estimation of the parameters A and b can be done directly by maximizing the approximate marginal likelihood $E_q[\log p(y|r, \Omega)]$.

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

4. EXPERIMENTS AND RESULTS

4.1 Case: Circadian patterns in rat liver

4.1.1 Datasets

Circadian rhythm is a daily time-keeping mechanism fundamental to a wide range of species. The basic molecular mechanism of circadian rhythm has been studied extensively. As a real example to test our approach, we considered the dynamics of the circadian patterns in rat liver. We employ the datasets from Almon RR et al [27]. This experiment was designed to examine fluctuations in gene expression in liver within the 24 hour circadian cycle in normal animals. Fifty-four male normal Wistar rats were housed in a strictly controlled stress free environment with light: dark cycles of 12 hr: 12hr. Three animals were sacrificed at each of 18 selected time points within the 24 hour cycle. RNA was prepared from livers for gene arrays. Time point designations reflect time after lights on in hours.

4.1.2 Results and analysis

To test the proposed model on the above dataset, we employ two important transcriptional regulators of which activity levels indicate the variation of heat signals in a subset of gene circadian network, hsf1 and ppara. In total, we selected 7 genes to perform posterior inference of TF activities. To ensure identifiability, we included three genes that are regulated solely by hsf1 (HSP110, HSPA8 and COL4A1), and two genes that are regulated solely by ppara (ACSL1 and HMGCS1). The remaining two genes are jointly regulated by hsf1 and ppara. These genes were chosen since they exhibit the largest variance in the microarray time course, and hence are likely to provide a cleaner representation of the output of the system.

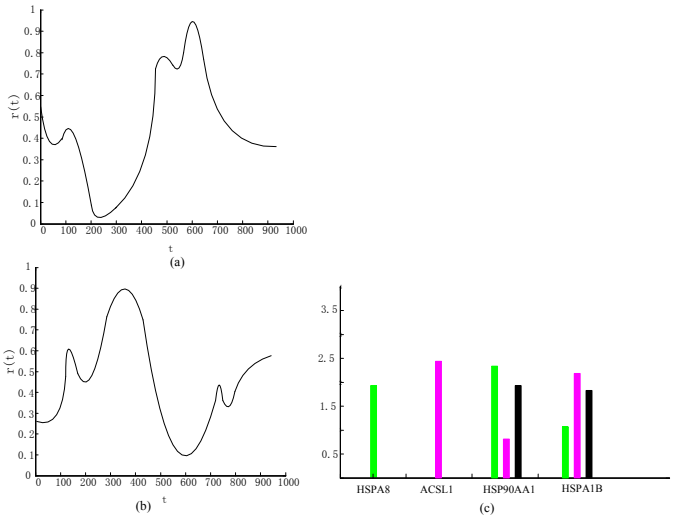

Figure 3-1. Results on circadian patterns data: (a) mean activity profile for hsf1, (b) mean activity profile for ppara, (c) bar-chart representation of the parameters k_i giving the logical structure of the interaction between two TFs.

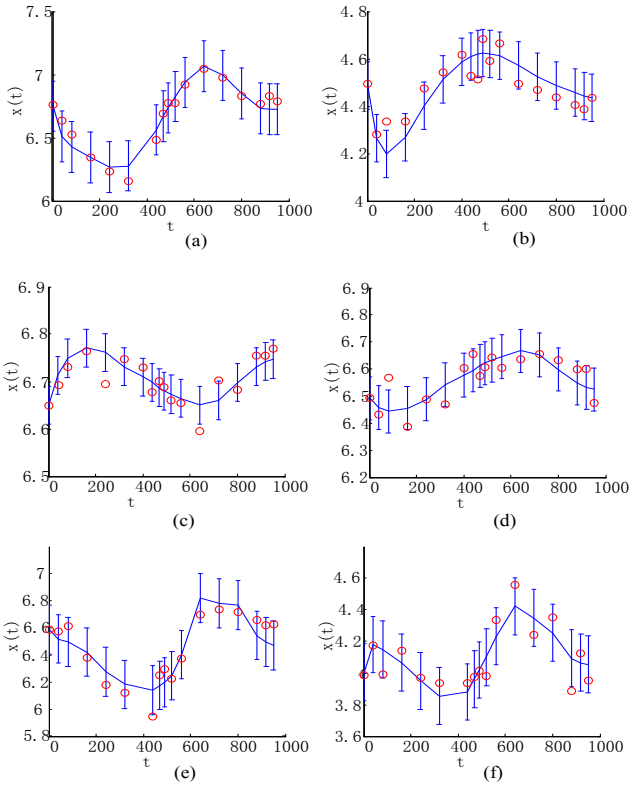

Figure 3-2. Expression profile and mean reconstruction of target genes. (a) HSPA8, (b) COL4A1, (c) ACSL1, (d) HMGCS1, (e) HSP90AA1 and (f) HSPA1B. The red circle indicates the observed value at each time-points.

The inferred TFs' activity levels are shown in Figure 3-1 (a) and (b). Both inferred TF profiles sow a noisy periodic behavior [28]. Figure 3-1 (c) gives the values of the parameters k_i for the four selected circadian genes (HSPA8, ACSL1, HSP90AA1 and HSPA1B). The green column represents the response k_1 to hsf1 alone, the red column is the response k_2 to ppara alone and the black column represents the joint response k_{12}. It can be seen that, for gene, HSPA8, the model predicts a significant activation by hsfl alone, which is consistent with the experimental conclusion from Yan et al [28]. The black columns of HSP90AA1 and HSPA1B demonstrate that the model predicts a significant combinatorial activation which can be verified by mutagenetic techniques, i.e. by knocking out one of the TFs. Figure 3-2 shows the fit of the model to the observed data at each time-point.

5. DISCUSSION AND CONCLUSION

In this paper, a gene transcriptional regulation model is proposed in perspective of biochemical binding reactions. The sequence features of the promoter and the possible occupancy of nucleosomes are exploited to estimate the TFs' binding probability. Unlike previous methods that only employ microarray data [16, 17], the present model can reveal more biological sense.

The DBN-based model of transcription rates and regulator activity levels allows us to handle these biologically relevant quantities despite the indirect measurement of the former and

the lack of measurements of the latter. It also allows us to handle the inherently noisy measurement in a principled way. However, the proposed model still abstracts away some of the explicit processes that generate the actual observed expression data. A more explicit modeling of these will provide a more principled treatment of different sources of noise in the data. Furthermore, this model does not handle directly the upstream events that affect regulator activity. In fact, the current model is an open loop system, such that the regulation of regulator activity is itself viewed as exogenous to the system. By developing a richer modeling language we may capture more complex reaction models, model the upstream regulation of activity levels, and identify systems that involve feedback mechanisms and signaling networks.

Post-Transcriptional Modification Model (PTMM) have been previously used to model TF activities [30]; in that work, further dependencies were included between TF mRNA expression levels and their predicted activities, which enabled to predict possible post-transcriptional modifications in TFs. Naturally, it should be possible to combine both our approach and their approach to give a model capable of simultaneously inferring TF activities, combinatorial interactions and post-transcriptional regulations.

REFERENCE

[1] M. Bansal, V. Belcastro, A. Ambesi and I. Bernardo, "How to infer gene networks from expression profiles", Mol. Syst. Biol., vol. 3, no. 78, 2007.

[2] F. Markowetz, and R. Spang, "Inferring cellular networks - a review", BMC Bioinformatics, Vol. 8, S5, 2007.

[3] M.B. Eisen, P.T. Spellman, Brown P.O. and Bostein D., "Cluster analysis and display of genome-wide expression patterns", Proc. Natl Acad. Sci.USA, vol. 95, pp.14863-14868, 1998.

[4] M. Davidich and S. Bornholdt, "Boolean network model predicts cell cycle sequence of fission yeast", PLoS ONE, vol. 3, e1672, 2008.

[5] H. Lahdesmaki, I. Shmulevich, O. Yli-Harja, "On learning gene regulatory networks under the Boolean network model," Mach Learn, Vol. 52, pp.147–167, 2004.

[6] S. Imoto, T. Goto and S. Miyano, "Estimation of genetic networks and functional structures between genes by using Bayesian networks and nonparametric regression," Proc Pac Symp Biocomput, vol. 7, pp. 175–186, 2002.

[7] K. Kirimasthong and A. Manorat, "Inference of gene regulatory network by Bayesian network using Metropolis-Hastings Algorithm", ICDM, pp. 276–286, 2007.

[8] E. Segal, B. Taskar, A. Gasch, N. Friedman and Koller D "Rich, probabilistic models for gene expression," Bioinformatics, vol. 17, pp. 243–252, 2003.

[9] Z. Hu, P. Killion and V. Iyer, "Genetic reconstruction of a functional transcriptional regulatory network," Nat. Genet., vol. 39, pp. 683–687, 2007.

[10] N. Luscombe, M. Babu and H. Yu, "Genomic analysis of regulatory network dynamics reveals large topological changes," Science, vol. 431, pp. 308–312, 2004.

[11] V. Adriano and H. Dirk, "Reconstructing gene regulatory networks with Bayesian networks by combining expression data with multiple sources of prior knowledge," Stat. Appl. Genet. Mol. Biol., vol. 6, A15, 2007.

[12] E. Segal, Y. Barash, I. Simon and N. Friedman, "From promoter sequence to expression: a probabilistic framework," RECOMB, pp. 263–272, 2002.

[13] W. Liebermeister, "Linear modes of gene expression determined by independent component analysis," Bioinformatics, vo. 18, pp. 51–60, 2002.

[14] N. Holter, M. Mitra and A. Maritan. "Fundamental patterns underlying gene expression profiles: simplicity from complexity," Proc. Natl. Acad. Sci. USA, vol. 97, pp. 8409–8414, 2000.

[15] I. Pournara and L. Wernisch, "Factor analysis for gene regulatory networks and transcription factor activity profiles. BMC Bioinformatics 8:61, 2007.

[16] T. Yu and K. Li, "Inference of transcriptional regulatory network by two-stage constrained space factor analysis," Bioinformatics, vol. 21, pp. 4033–4038, 2005.

[17] S. Imoto, S. Kim and T. Goto, "Bayesian network and nonparametric heteroscedastic regression for nonlinear modeling of genetic network," J. Bioinformatics. Comput. Biol. Vol. 1, pp. 231–252, 2003.

[18] E. Segal, T. Raveh-Sadka et al., "Predicting expression patterns from regulatory sequence in Drosophila segmentation", Nature, vol. 451, pp. 535–540, 2008.

[19] J. Gertz, E. Siggia and B. Cohen, "Analysis of combinatorial cis-regulation in synthetic and genomic promoters," Nature, vol. 457, pp. 215–218, 2009.

[20] A. Tanay, "Extensive low-affinity transcriptional interactions in the yeast genome," Genome Res., vol. 16, pp. 962–972, 2006.

[21] S.M. Johnson, F.J. Tan, H.L. McCullough, D.P. Riordan and A.Z. Fire, "Flexibility and constraint in the nucleosome core landscape of Caenorhabditis elegans chromatin," Genome Res, vol. 16, pp. 1505–1516, 2006.

[22] N. Kaplan et al., "The DNA-Encoded Nucleosome Organization of a Eukaryotic Genome", Nature, vol. 458, pp. 362-366, 2008.

[23] M.L. Bulyk et al., "Computational prediction of transcription-factor binding site locations," Genome Biol., vol.5, A201, 2003.

[24] I. Naum, D. Gary and P. Ilya, "Computational technique for improvement of the position-weight matrices for the DNA/protein binding sites," Nucleic. Acids. Research, vol. 33, no. 7, pp. 2290–2301, 2005.

[25] S. Guido, R. Andreas, O. Manfred and A. Cedric, "Switching regulatory models of cellular stress response," Bioinformatics, vol. 25, pp. 1280–1286, 2009.

[26] T. Cover and J. Thomas, Elements of information theory. Wiley, New York, 2006.

[27] R.R. Almon, E. Yang, W. Lai, I.P. Androulakis, "Circadian variations in rat liver gene expression: relationships to drug actions," J. Pharmacol. Exp. Ther., vol. 326, no. 3, pp. 700-716, 2008.

[28] J. Yan, H.F. Wang, Y.T. Liu and C.X. Shao, "Analysis of Gene Regulatory Networks in the Mammalian Circadian Rhythm," Plos Comput. Biol., vol. 4, no. 10, e1000193, 2008.

[29] L. Cai, N. Friedman and X. Sunney, "Stochastic protein expression in individual cells at the single molecule level," Nature, vol. 440, pp.580-586, 2006.

[30] Y. Shi, M. Klutstein, I. Simon, T. Mitchell and Z. Bar-Joseph, "A combined expression-interaction model for inferring the temporal activity of transcription factors," J. Comput. Biol., vol. 16, pp. 1035–1049, 2009.

978-1-4577-1666-9/11/$26.00 ©2011 IEEE