Influence of mRNA decay rates on the computational prediction of transcription rate profiles from gene expression profiles

Chi-Fang Chin[1,*], Arthur Chun-Chieh Shih[2] and Kuo-Chin Fan[1,3]

[1]*Department of Computer Science and Information Engineering, National Central University, Chung-Li, Taiwan*
[2]*Institute of Information Science, Academia Sinica, Taipei, Taiwan*
[3]*Department of Informatics, Fo Guang Univeristy, Ilan, Taiwan*

**Corresponding author (Fax, 886-3-4227426; Email, annking@iis.sinica.edu.tw)*

The abundance of an mRNA species depends not only on the transcription rate at which it is produced, but also on its decay rate, which determines how quickly it is degraded. Both transcription rate and decay rate are important factors in regulating gene expression. With the advance of the age of genomics, there are a considerable number of gene expression datasets, in which the expression profiles of tens of thousands of genes are often non-uniformly sampled. Recently, numerous studies have proposed to infer the regulatory networks from expression profiles. Nevertheless, how mRNA decay rates affect the computational prediction of transcription rate profiles from expression profiles has not been well studied. To understand the influences, we present a systematic method based on a gene dynamic regulation model by taking mRNA decay rates, expression profiles and transcription profiles into account. Generally speaking, an expression profile can be regarded as a representation of a biological condition. The rationale behind the concept is that the biological condition is reflected in the changing of gene expression profile. Basically, the biological condition is either associated to the cell cycle or associated to the environmental stresses. The expression profiles of genes that belong to the former, so-called cell cycle data, are characterized by periodicity, whereas the expression profiles of genes that belong to the latter, so-called condition-specific data, are characterized by a steep change after a specific time without periodicity. In this paper, we examine the systematic method on the simulated expression data as well as the real expression data including yeast cell cycle data and condition-specific data (glucose-limitation data). The results indicate that mRNA decay rates do not significantly influence the computational prediction of transcription-rate profiles for cell cycle data. On the contrary, the magnitudes and shapes of transcription-rate profiles for condition specific data are significantly affected by mRNA decay rates. This analysis provides an opportunity for researchers to conduct future research on inferring regulatory networks computationally with available expression profiles under different biological conditions.

[Chin C-F, Shih A C-C and Fan K-C 2007 Influence of mRNA decay rates on the computational prediction of transcription rate profiles from gene expression profiles; *J. Biosci.* **32** 1251–1262]

1. Introduction

At the steady state of the expression level of a gene, the rate of mRNA synthesis and mRNA degradation maintain an equilibrium (Raghavan and Bohjanen 2004). The mRNA decay rate, therefore, determines how quickly the expression level of an mRNA species can move to a new steady-state level after a change in the transcription level (Khodursky and Bernstein 2003). Several experimental studies of the mRNA decay rate in *E. coli* (Bernstein *et al* 2002), yeast (Wang *et al* 2002) and mammals including human (Raghavan and Bohjanen 2004; Yang *et al* 2003) have been reported recently. Within the same species, the decay rates of mRNAs belonging to different functional categories can vary from several times to over 100-fold. For example, the half-lives of mRNAs are from 3 min to 469 min in yeast (Wang *et al* 2002), while those in mammalian cells are from 20 min to 24 h (Fan *et al* 2002). These data all suggest that

Keywords. Cell cycle data; condition-specific data; gene expression profile; glucose-limitation data; mRNA decay rate; transcription-rate profile

mRNA decay rates significantly contribute to determining mRNA expression levels (Khodursky and Bernstein 2003).

Besides, reconstructing gene regulatory networks is a challenging task in computational biology. Recently, several quantitative studies of the regulatory networks from expression profiles have been proposed (Chen *et al* 1999, 2004; Nachman *et al* 2004; Sasik *et al* 2002). In these studies, the dynamics of a gene transcription process was modelled by a first-order non-linear differential equation in which the rate of change of the expression level is the difference between the transcription rate and degradation of the mRNA transcript (Chen *et al* 2004). According to mRNA expression profiles and experimentally measured decay rates, the transcription rate profiles of genes in yeast can then be determined by either a linear combination of time-dependent switch functions (Chen *et al* 2004) or a non-linear Michaelis–Menten form (Nachman *et al* 2004).

In this paper, we present a systematic method based on the dynamic gene regulation model, in which the mRNA decay rates are effectively under control and therefore the transcriptional rate profiles can be calculated accurately from available gene expression profiles. B-spline curve fitting is utilized to approximate the continuous representations of the standardized expression time-course data. It is worth noting that B-spline curve fitting can deal with the sampling errors and missing value problems that occur frequently in expression data. After the continuous representations of expression profiles have been approximated, the dynamic gene regulation model is used to calculate the transcription rate profiles by considering both expression profiles and mRNA decay rates.

An expression profile can be regarded as a representation of a biological condition because a different cellular state, such as a developmental state or a cellular response to circumstantial changes, is reflected in the alteration of gene expression profiles. The expression profiles of genes that are associated with the cell cycle, called cell-cycle genes, are characterized by periodicity because they are regulated by periodic fluctuation. The expression profiles of genes associated with environmental stresses, called condition-specific genes, are characterized by a steep rise or drop without periodicity. Therefore, different expression profiles including simulated datasets and real datasets are examined in this paper for understanding the influences of mRNA decay rates on the computational predictions of transcription rate profiles from gene expression profiles. Two simulated datasets of expression profiles are examined, of which one is characterized by periodicity, which is the same as cell-cycle genes and another is characterized by a steep rise or drop, which is the same as condition-specific genes. Considering the real datasets, the cell-cycle genes are examined since their expression profiles are regulated by periodic fluctuation. Moreover, glucose-limitation genes are regarded as condition-specific genes to be examined because their expression profiles undergo a steep rise or fall after glucose deprivation.

The rest of this paper is organized as follows. In § 2, we describe the framework of our system. The three important processes in the system, including the dynamic gene regulation model, B-spline curve fitting and transcription rate profile calculation, are also highlighted and presented in this section. Experiments were conducted on two datasets of simulated data and two datasets of real data to examine whether the presented system was useful in understanding the influence of mRNA decay rates on computational prediction of transcription rate profiles from gene expression profiles. The results are illustrated in § 3. Finally, concluding remarks are given in § 4.

2. Materials and methods

Generally, microarray experiments measure the expression level of each DNA sequence by the ratio of signal intensity between the test sample and the control sample. In this paper, unless explicitly stated, we refer to each profile as it has been transformed by the logarithm of its raw values. Furthermore, each profile has been standardized by subtracting its mean profile and then dividing the difference by its standard deviation such that each profile has a mean of zero and a variance of one. In the following contexts, we assume that the expression profiles have been manipulated under a proper logarithmic transformation and standardized with zero-mean and unit-variance.

In this section, the systematic method is introduced in detail. First, the dynamic gene regulation model shows that the decay rate, expression profile and transcription rate profile are modelled in a first-order differential equation. Second, since expression profiles are sampled non-uniformly to be the time-course data usually resulting in sampling errors or missing value problems, B-spline curve fitting is used to approximate the continuous representations; this can deal with sampling errors and the problem of missing values. Finally, the transcription rate profiles for the given genes are calculated computationally from the continuous representations of the expression profiles and their corresponding decay rates based on the dynamic gene regulation model.

2.1 *Dynamic gene regulation model*

In this model, given a gene x, its regulation process can be simply formulated by a first-order differential equation (Singer and Penman 1973; Chen *et al* 1999, 2004; Sasik *et al* 2002; Yeung *et al* 2002):

$$\dot{X}_i(t) = G_i(t) - \lambda_i \cdot X_i(t) + \xi_i(t)\,, \tag{1}$$

where $X_i(t)$, $G_i(t)$, λ_i, and $\xi_i(t)$ represent the expression profile of the i-th gene at time t, mRNA synthesis rate, decay rate and noise of a specific gene i, respectively. The term $G_i(t)$ represents a complex function of the relevant biological factors including the abundance of transcription factors, enzymes, energy resources of the cell such as external physical and chemical stimuli (Sasik *et al* 2002). In general, this term can be considered as a transcription rate profile for a gene. Besides, $\xi_i(t)$ can be ignored because this item represents noise and uncertain factors.

This dynamic gene regulation model defines the changes in mRNA level at time t, $\dot{X}_i(t)$, as the difference between $G_i(t)$ and $\lambda_i \cdot X_i(t)$, which represents the production and degradation of transcripts at time t, respectively. However, the given expression profiles are time-course data instead of continuous representations resulting from the microarray experiments and the first derivative of $X_i(t)$ cannot be directly estimated from these discrete data. Therefore, a good curve-fitting technique is required to appropriate the continuous representation of $X_i(t)$ and the first derivative of $X_i(t)$, $\dot{X}_i(t)$, can then be easily estimated.

2.2 *B-spline curve fitting*

For the time-course data in microarray experiments, finding an appropriate continuous representation of each gene throughout the course is very important (Bar-Joseph 2004). Since all data of the time-course gene expression profiles are often non-uniformly sampled and some unexpected experimental problems cause sampling errors or missing data problems, using continuous curves to approximate these discrete time-course data can reduce the noise and fill most of the unexpected parts of the input data. Several approaches have been proposed to solve this so-called curve-fitting problem (Bar-Joseph *et al* 2003; Luan and Li 2003). In this paper, we adopt the B-spline curve fitting method, which uses piece-wise low-degree polynomials to approximate the continuous representation of the data for avoiding problems generated by the use of single high-degree polynomials. The normalized B-spline bases can be calculated by the following recursion formula (Bar-Joseph *et al* 2003; Rogers 2000):

$$b_{i,1}(t) = \begin{cases} 1, & p_i \le t \le p_{i+1} \\ 0 & \text{otherwise} \end{cases}, \tag{2}$$

and

$$b_{i,k}(t) = \frac{(t - p_i) b_{i,k-1}(t)}{p_{i+k-1} - p_i} + \frac{(p_{i+k} - t) b_{i+1,k-1}(t)}{p_{i+k} - p_{i+1}}, \tag{3}$$

where $b_{i,j}$ $(1 \le j \le k)$, p_i $(i = 1...m+k)$, k are a B-spline basis, a knot, and the order of the basis polynomials, respectively. For instance, k for a cubic B-spline equals 4 and the fitting curve $y(t)$ can be written as

$$y(t) = \sum_{i=1}^{m} C_i b_{i,4}(t), \tag{4}$$

where $p_4 \le t \le p_{m+1}$ and the coefficients C_i $(1 \le i \le m)$ are called control points.

Let X_n be the standardized expression profile of gene x whose observations are made at n time points $t_1,\ldots, t_n$. It can then be written as the matrix equation $X_n = BC$, where C is the vector of m control points and B is an n by m matrix in which $[B]_{i,j} = b_{j,k}(t_i)$. By using a least-square method to solve the matrix equation, if n=m, we can have $C = (B^TB)^{-1}B^TX_n$. Thus, $y(t) = \sum_{i=1}^{m} C_i b_{j,k}(t_i)$ is a continuous representation of X_n. More details can be found in Bar-Joseph *et al* (2003) and Rogers (2000).

Since the sampling errors or missing value problems often occur in microarray data, B-spline curve fitting can transfer the discrete time-course data to continuous representations leading to the reduction of sampling errors and solving of missing value problems.

2.3 *Transcription rate profile calculation*

As mentioned in subsection 2.1, the regulation process of a gene can be simply formulated by equation (1). Rearranging the terms in equation (1), we can rewrite $G_i(t)$ as being equal to the following equation:

$$G_i(t) = \dot{X}_i(t) + \lambda_i \cdot X_i(t) - \xi_i(t). \tag{5}$$

In general, we can assume that $X_i(t)$ and $G_i(t)$ are the measured expression profile and transcription rate profile, respectively. Having obtained the continuous representation of expression profiles and decay rates, we can calculate predicted transcription rate profiles directly from equation (5). Since mRNA will be degraded eventually, the decay rate will never be equal to zero. Therefore, we do not analyse the case of decay rate equal to zero in this paper.

3. Results

In this section, we analyse the influences of decay rates on the computational prediction of transcription rate profiles by examining two datasets of simulated data and two datasets of experimental data. The results are also illustrated in this section.

3.1 *Simulated expression data*

Generally speaking, gene expression profiles can be classified into two types based on the biological condition.

First, the eukaryotic cell cycle is known to be both regulated and accompanied by a periodic fluctuation in the expression levels of numerous genes (Cho *et al* 1998). In the literature, a few transcripts with regularly periodic expression profiles have been characterized with respect to a specific cell-cycle phase (Cho *et al* 1998; Spellman *et al* 1998). Therefore, for genes belonging to this type, i.e. cell-cycle genes, their expression profiles are characterized by periodicity due to the periodic fluctuation in their expression levels. Second, all cells sense and react to changes in their environment, such as changes in temperature, nutrients, pH, and so on. Appropriate responses to these stresses must induce cell survival and proliferation (Chen *et al* 2003). The genes involved in the stress response are therefore expressed transiently in an abrupt manner after environmental stress. For these condition-specific genes, biological conditions are reflected in steep changes in expression profiles instead of periodicity.

To clarify the analysis, we simulate these two types of expression profiles, cell-cycle genes and condition-specific genes, to examine the influence of decay rates. First, for cell-cycle genes, a sine wave function with a period of 90 minutes is simulated to be the expression profile with periodicity as shown in figure 1a. The corresponding transcription rate profiles calculated by different decay rates based on our presented systematic method are shown in figure 1b, where the curves depicted in the colours blue, green, red, cyan, magenta, yellow and black are the results with decay rates 0.001 (1/min), 0.1 (1/min), 0.2 (1/min), 0.3 (1/min), 0.4 (1/min), 0.5 (1/min) and 0.6 (1/min), respectively. It is interesting to note that these transcription rate profiles are similar in magnitude and their shapes are only slightly different. To understand the differences in shape among these transcription rate profiles calculated on the basis of different decay rates, these transcription rate profiles are shifted to the same peak time and Pearson correlation coefficients are then

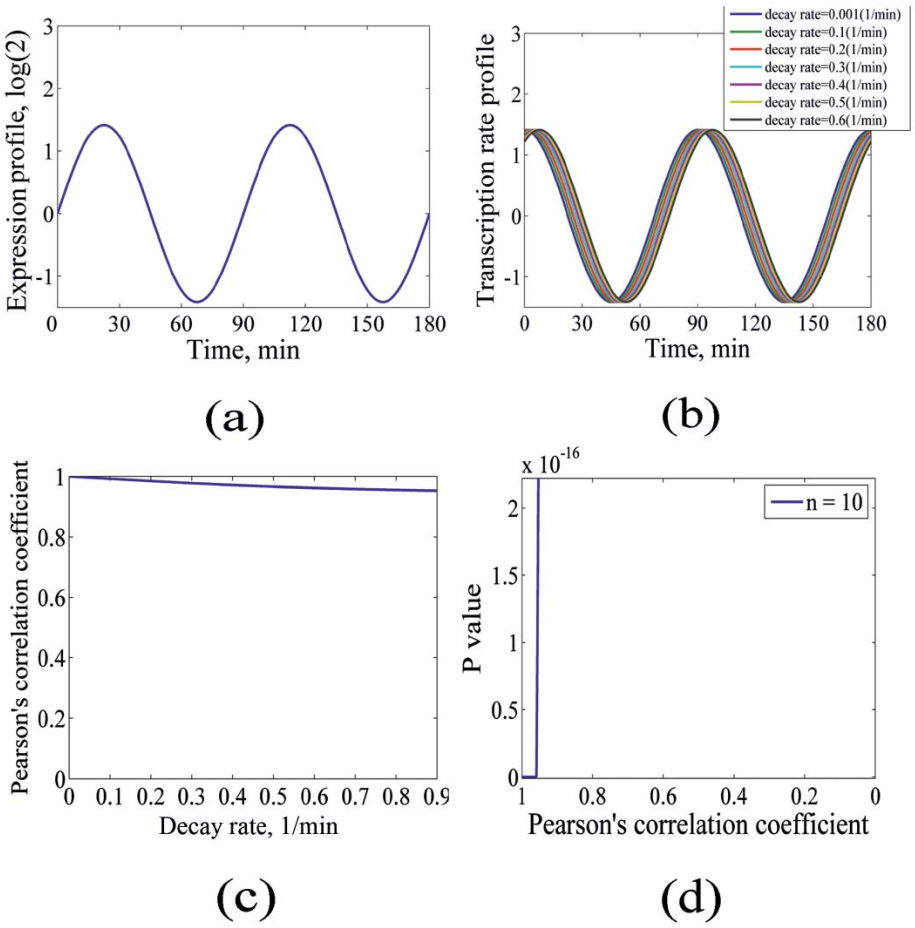

Figure 1. Simulated data for expression profile characterized by periodicity. **(a)** The expression profile. **(b)** The corresponding transcription rate profiles calculated by seven different decay rates from 0.001 (1/min) to 0.6 (1/min). **(c)** Pearson correlation coefficient calculated between transcription rate profile with a specified decay rate from 0.001 to 0.9 and one with a decay rate of 0.001. **(d)** The *P* value for Pearson correlation with 8 degrees of freedom.

calculated between the transcription rate profile with a decay rate of 0.001 (1/min), and the other transcription rate profiles with decay rates from 0.001 (1/min) to 0.9 (1/min). Pearson correlation coefficient is widely used in biology to measure the linear proximity between two profiles. This measure has the advantage of estimating the proximity between profiles depending only on the pattern. As figure 1c demonstrates, all the values of Pearson correlation coefficient are larger than 0.9, which means that these transcription rate profiles are highly correlated in shape. Moreover, the *P* values for Pearson correlation coefficients with 8 degrees of freedom, i.e. the number of pairs of data points is 10, is shown in figure 1d. Although the number of pairs of data points is 10 here, it can be sampled to infinity because the transcription rate profile is a curve. Therefore, the results clearly indicate that decay rates will not affect the computational prediction of gene expression profiles for cell-cycle genes.

Second, regarding condition-specific genes, one characteristic that distinguishes their expression profiles from that of cell-cycle genes is that there are steep rises or drops in the expression profiles. Since the curve of sigmoid function has the characteristic S-shape exhibiting a dramatic rise after a specific time, we can take advantage of this feature of sigmoid function to simulate the steep changes that often occur in condition-specific genes. That is, an expression profile of a condition-specific gene can be simulated by combining a set of sigmoid functions. A modified sigmoid function, $p(a,b,t)$, is defined as

$$p(a,b,t) = \frac{1}{1 + k \cdot e^{-a(t-b)}}, \tag{6}$$

where a, b, and k are real constants and t is the time variable, respectively.

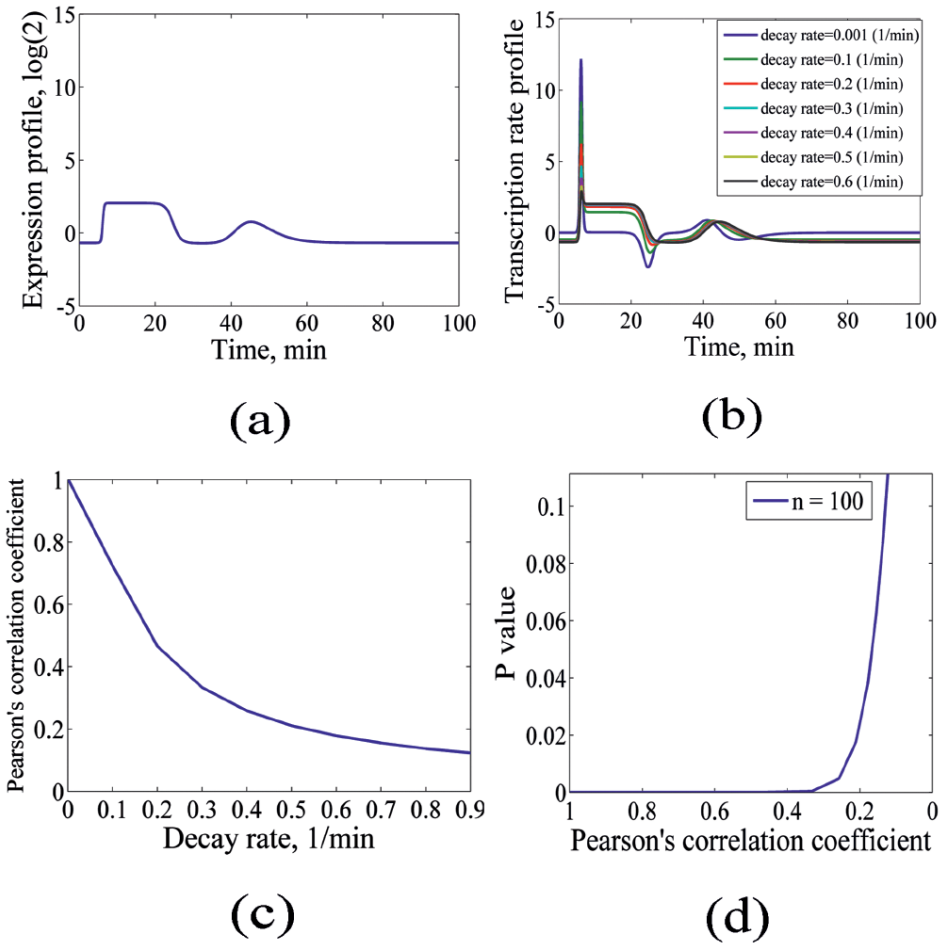

Figure 2. Simulated data for expression profile characterized by steep changes. **(a)** The expression profile. **(b)** The corresponding transcription rate profiles calculated by seven different decay rates from 0.001 (1/min) to 0.6 (1/min). **(c)** Pearson correlation coefficient calculated between transcription rate profile with a specified decay rate from 0.001 to 0.9 and one with a decay rate of 0.001. **(d)** The *P* value for Pearson correlation with 98 degrees of freedom.

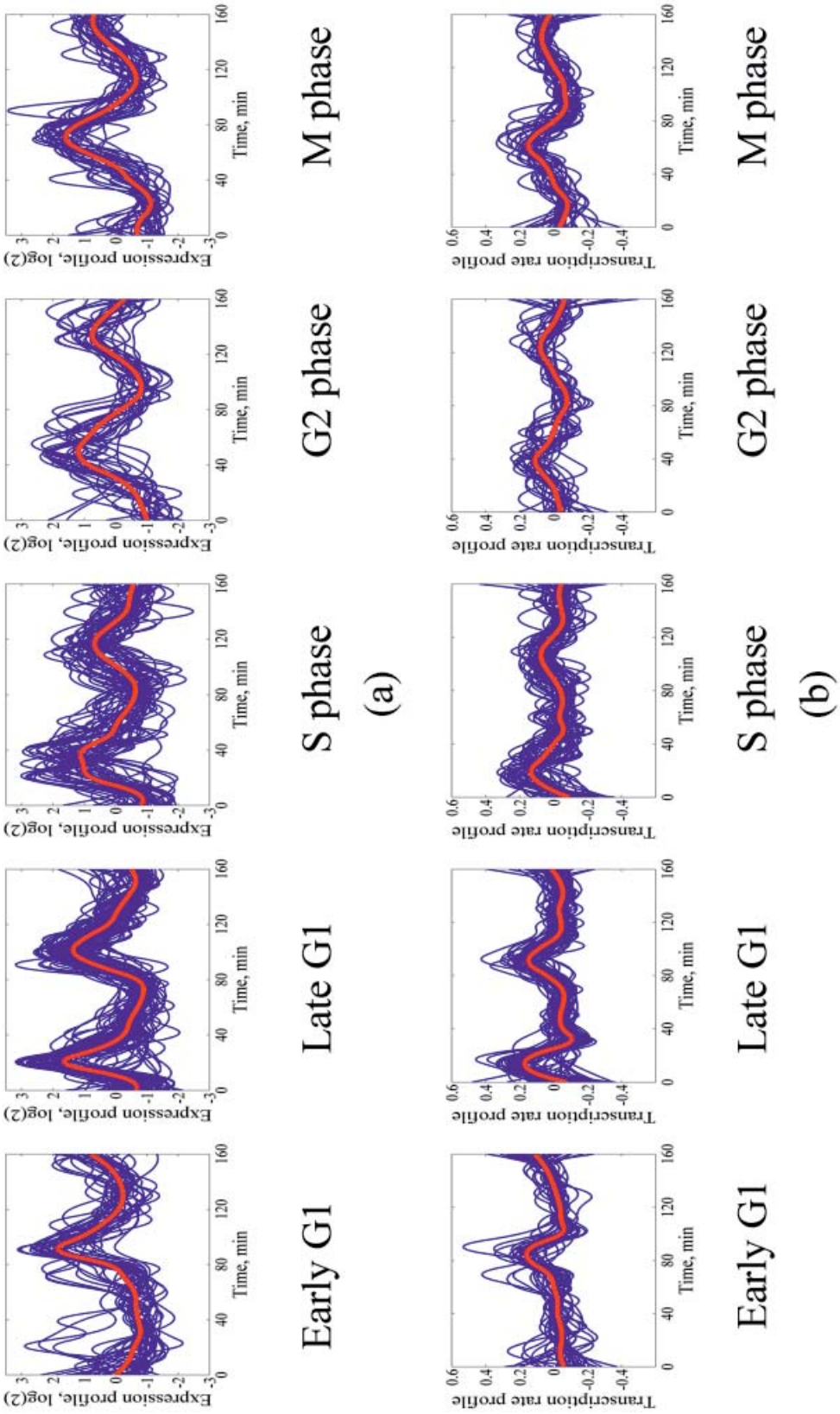

Figure 3. Yeung *et al*'s cell-cycle data including 222 genes which are sampled by 17 time points. **(a)** Expression profiles at different cell-cycle stages. **(b)** The corresponding transcription rate profiles.

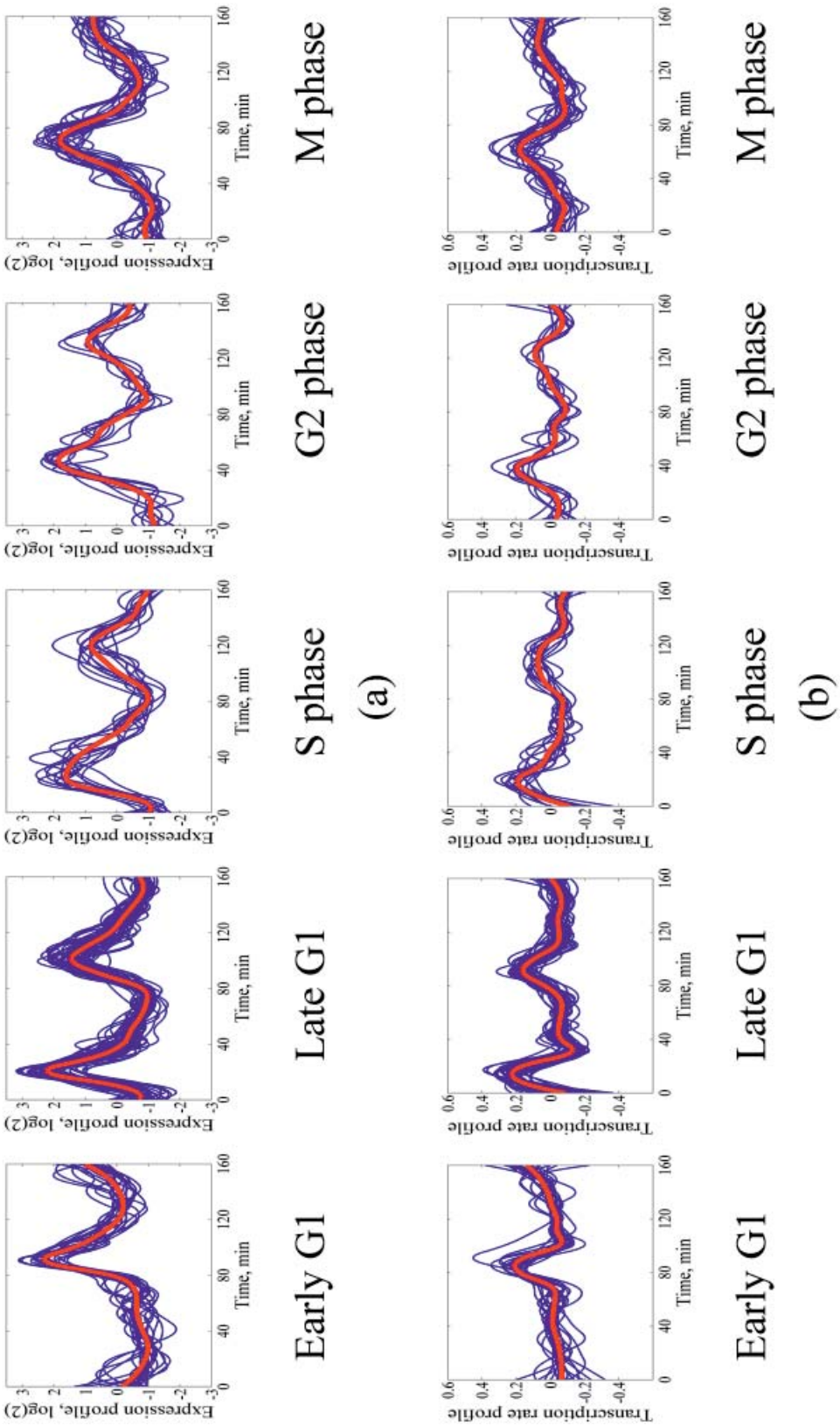

Figure 4. Refined Yeung *et al*'s cell-cycle data. (**a**) Expression profiles at different cell cycle stages. (**b**) The corresponding transcription rate profiles.

Now, let us describe an expression profile of a condition-specific gene as

$$p(5,6,t) - p(1,24,t) + p(0.5,40,t) - p(0.25,45,t) \quad (7)$$

as shown in figure 2a. Obviously, there are steep changes presented in this simulated expression profile. To calculate their corresponding transcription rate profiles, the first derivative of formula (7) has to be calculated. Since the first derivative of sigmoid function $p(a,b,t)$ will simply be $a \cdot p \cdot (1-p)$, the first derivative of formula (7) can be written as

$$5 \cdot p(5,6,t) \cdot [1-p(5,6,t)] - 1 \cdot p(1,24,t) \cdot [1-p(1,24,t)] + 0.5 \cdot p(0.5,40,t) \cdot [1-p(0.5,40,t)] - 0.25 \cdot p(0.25,45,t) \cdot [1-p(0.25,45,t)]. \quad (8)$$

Figure 2b illustrates the corresponding transcription rate profiles calculated according to seven different decay rates. The curves depicted in the colours blue, green, red, cyan, magenta, yellow and black represent the results with decay rates equal to 0.001 (1/min), 0.1 (1/min), 0.2 (1/min), 0.3 (1/min), 0.4 (1/min), 0.5 (1/min) and 0.6 (1/min), respectively. Interestingly, we find that these transcription rate profiles with different decay rates become dissimilar in both magnitude and shape. To demonstrate the influences of different decay rates for the calculated transcription rate profiles, we calculated Pearson correlation coefficients between a transcription rate profile with a decay rate of 0.001 (1/min) and the other transcription rate profiles with decay rates from 0.001 (1/min) to 0.9 (1/min) as well. As depicted in figure 2c, there is a remarkable decrease and then

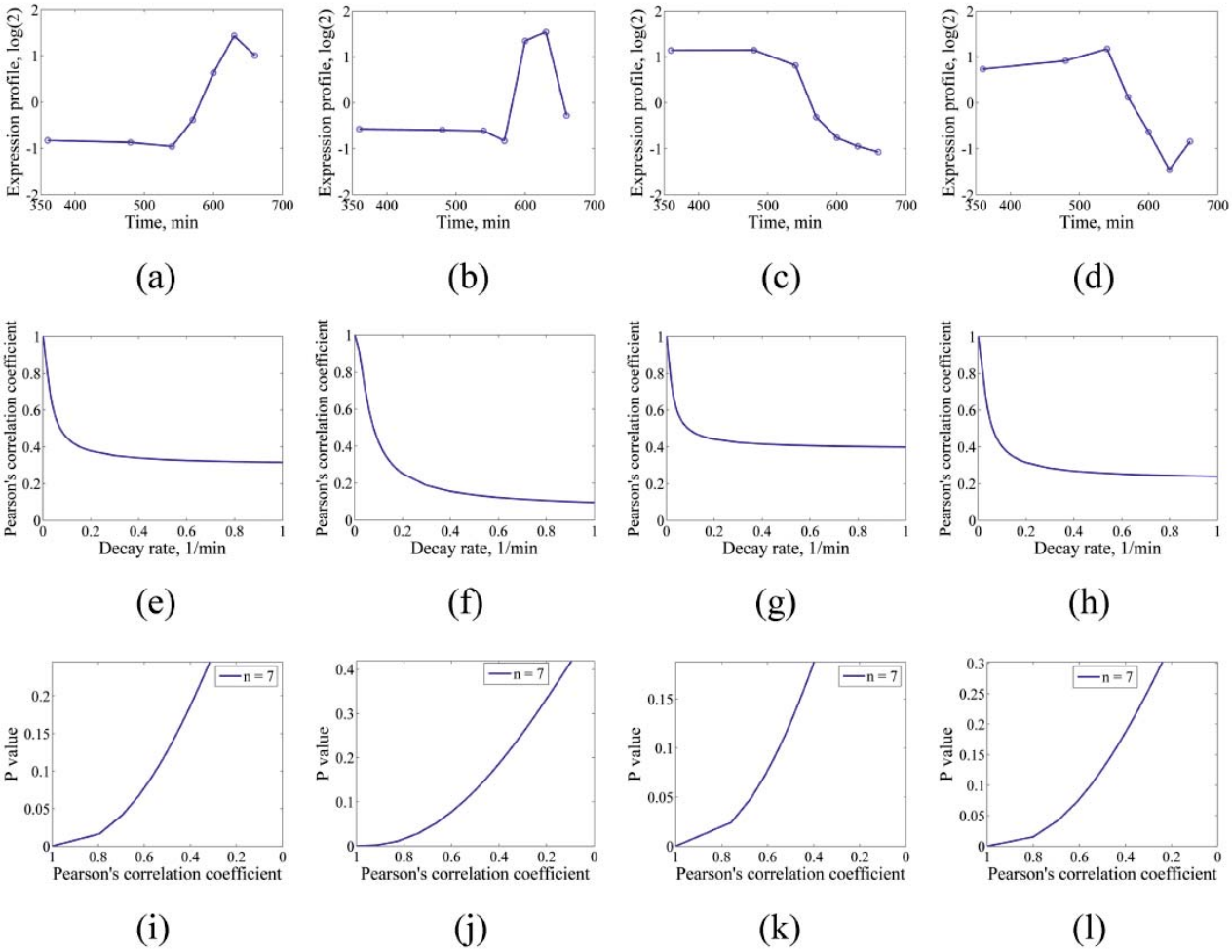

Figure 5. The plots of four genes of Brauer *et al*'s glucose-limitation data including expression profiles approximated by B-spline curve fitting with the value of the control point being set to 7 where the order of basis polynomials is set to 2, Pearson correlation coefficients between transcription rate profile predicted by a specific decay rate from 0.001 to 1 and transcription rate profile with decay rate 0.001, and the *P* value for Pearson correlation coefficients with 5 degrees of freedom. (**a**), (**e**), (**i**) are the plots of gene *PUT1*. (**b**), (**f**), (**j**) are the plots of gene *GDH3*. (**c**), (**g**), (**k**) are the plots of gene *PDC5*. (**d**), (**h**), (**l**) are the plots of gene *HXK2*.

a gradual decline to a small value for Pearson correlation coefficients, which means that the correlations decrease for these transcription rate profiles in shape when decay rates increase so that the computational prediction of transcription rate profiles are strongly related to decay rates. As shown in figure 2d, the *P* values for Pearson correlation coefficients with 98 degrees of freedom, i.e. the number of pairs of data points is 100, reveal the statistical significance for Pearson correlation coefficients.

In what follows, we examine two different types of real gene expression data in yeast including cell-cycle data and glucose-limited data to study how decay rates influence the computational prediction of transcription rate profiles from expression profiles of real data. As far as we know, cell-cycle data possess the property of periodicity (Yeung *et al* 2001) and glucose-limitation data are characterized by a steep rise or a rapid drop in expression profiles without periodicity (Brauer *et al* 2005**).**

3.2 *Refined cell-cycle data*

Here, 384 yeast cell-cycle expression profiles from Yeung *et al*'s research (2001) are considered. The 384 expression profiles are a subset of Cho *et al*'s experiment (1998) of which these genes over 17 time points taken at 10 min intervals cover almost two cell cycles. However, some of these 384 genes do not have the measured decay rates which are publicly available from Wang *et al*'s actual decay rates of 4687 yeast genes measured experimentally.

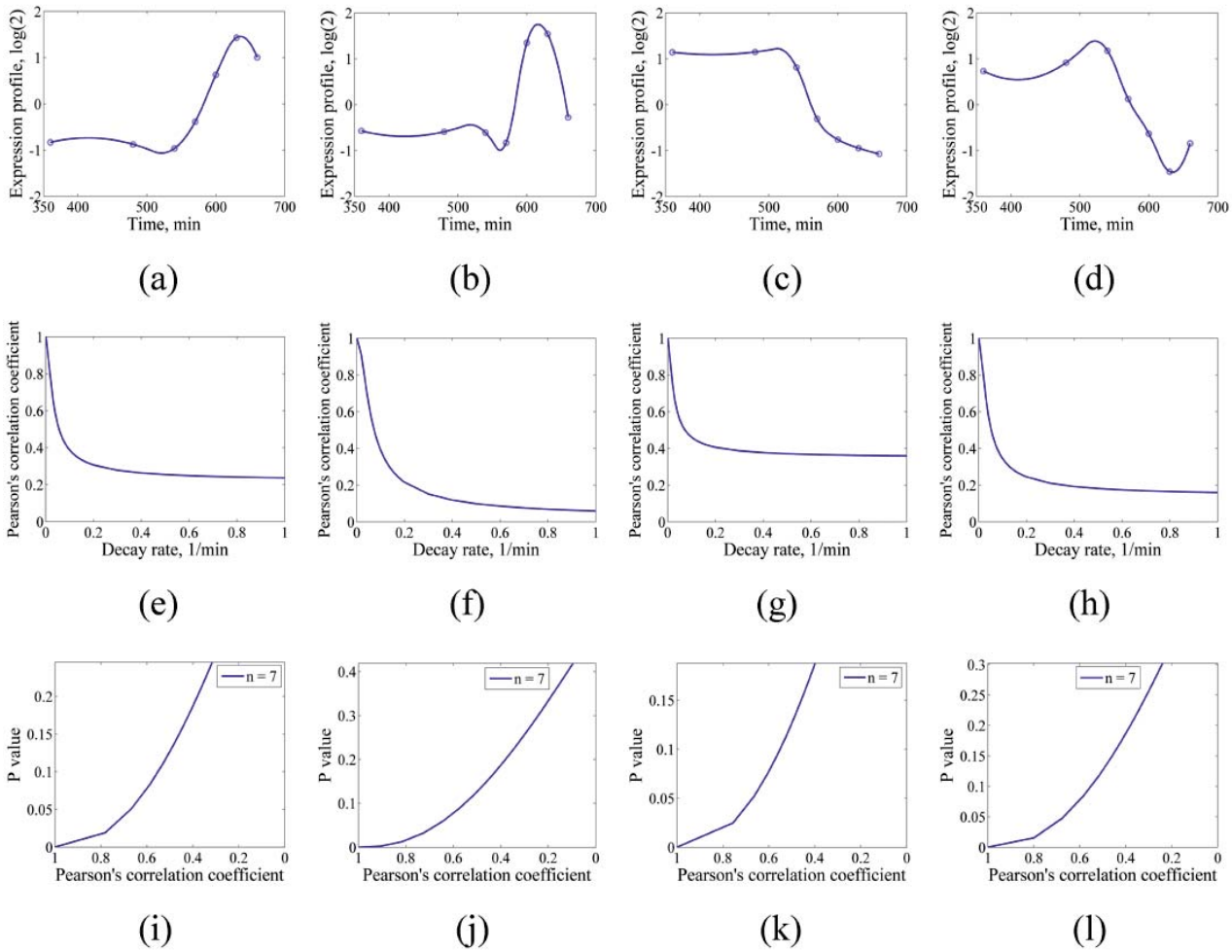

Figure 6. The plots of four genes of Brauer *et al*'s glucose-limitation data including expression profiles approximated by B-spline with the value of the control point being set to 7 where the order of basis polynomials is set to 3, Pearson correlation coefficients between transcription rate profile predicted by a specific decay rate from 0.001 to 1 and transcription rate profile with a decay rate of 0.001, and the *P* value for Pearson correlation coefficients with 5 degrees of freedom. (**a**), (**e**), (**i**) are the plots of gene *PUT1*. (**b**), (**f**), (**j**) are the plots of gene *GDH3*. (**c**), (**g**), (**k**) are the plots of gene *PDC5*. (**d**), (**h**), (**l**) are the plots of gene *HXK2*.

Among these 384 genes, we thus used 222 genes for the following analysis because the decay rates of those genes are available. According to Yeung *et al*'s result (2001), the 222 cell-cycle expression profiles are classified into five cell-cycle phase-specific groups. We utilized these genes to study if the groups of transcription rate profiles were still the same as the original groups of those genes according to their expression profiles. The five groups of expression profiles peak at different cell-cycle phases and their corresponding transcription rate profiles with the actual decay rate calculated by our presented systematic method are illustrated in figure 3a, b. For each plot, the profiles are represented by thin blue lines and the average profile is represented by a thick red line.

As shown in figure 3, some profiles are inconsistent with the peak of the mean profile for each phase-specific group in both expression profiles and transcription rate profiles. For example, several expression profiles grouped together at the early G1 phase, as shown in the extreme left of figure 3a, peak at other phases. To avoid ambiguity in our analysis, we removed these expression profiles which were inconsistent with the peak of the mean profile by using Pearson correlation coefficient. If the Pearson correlation coefficient between a profile and the mean profile was smaller than the threshold (which is set to 0.8 here), it was removed. The remaining profiles are called the refined cell-cycle data in which the decay rates of those genes are mostly less than 0.1 (1/min) or larger than 0.02 (1/min) on average. Expression profiles of the refined data at each phase-specific group and their corresponding transcriptional rate profiles are illustrated in figure 4a, b. The results tell us that the peaks of transcription rate profiles still maintain consistency for those genes originally grouped together according to their expression profiles. In addition, for those genes within a group, the peaks of transcription rate profiles remain similar in magnitude. Consequently, the results indicate that decay rates do not affect the computational prediction of the transcriptional rate profiles from the expression profiles in this case.

3.3 *Glucose-limitation data*

Unlike the feature of cell-cycle genes, i.e. periodicity, the expression profiles of glucose-limitation genes show an abrupt change at specific times when glucose is exhausted. Brauer *et al* (2005) listed a set of glucose-limited genes and examined the changes in their expression levels from fermentative to oxidative metabolism, known as the 'diauxic shift', for studying the physiological response to glucose limitation in yeast. These genes, such as tricarboxylic acid (TCA) cycle genes, vacuolar protease genes and hexokinase genes, are expressed at significant levels during the diauxic shift process. We downloaded a total of 38 genes obtained from the experimental data of the second bath culture at *http://genomics-pubs.princeton.edu/DiauxicRemodelin/home.shtml.*

However, the decay rates of these glucose-limitation genes are unavailable because Wang *et al* (2002) reported that the decay rates of genes were obtained under normal conditions instead of a specific condition, such as a glucose-limitation condition. Hence, we manually assigned values of decay rates from 0.001 to 1 to conduct the analysis. To understand the influence of decay rates on glucose-limitation data, we used different decay rates as stated previously to calculate the possible transcription rate profiles from these expression profiles. Shown in the top (a–d), middle (e–h) and bottom (i–l) panels of figure 5 are the expression profiles, the calculation of Pearson correlation coefficients and the *P* value for Pearson correlation coefficients of four different glucose-limited genes, *PUT1*, *GDH3*, *PDC5*, *HXK2* with different decay rates, respectively.

As shown in figure 5a, the expression profile of the gene *PUT1*, which functions in the assimilation of nitrogen and catabolism of amino acids, is increased significantly after the diauxic shift which is at 9.25 h (555 min in the figure). We then calculated the Pearson correlation coefficient between transcription rate profiles predicted by different decay rates from 0.001 (1/min) to 1 (1/min) and the transcription rate profile with a decay rate equal to 0.001 as shown in figure 5e. As the figure indicates, Pearson correlation coefficients decrease when the decay rate increases. Besides, there is a rapid drop in Pearson correlation coefficients when the decay rate is 0.001 (1/min) to 0.2 (1/min). Figure 5(i) shows the *P* values for Pearson correlation coefficients with 5 degrees of freedom since the number of given data points is 7. The situation is the same for gene *GDH3*, which takes carbon out of the amino acids into the TCA cycle after the diauxic shift. Its expression profile, Pearson correlation coefficients and the *P* value for Pearson correlation coefficients are shown in figure 5b, f, j. Besides, gene *PDC5* is an important transcription factor which controls the flow of metabolites to ethanol during fermentation. Its expression profile decreases threefold after the diauxic shift as shown in figure 5c. Pearson correlation coefficients illustrate the significant differences in decay rates as shown in figure 5g. Similarly, figure 5k shows the *P* values for Pearson correlation coefficients for gene *PDC5*. As figure 5d demonstrates, the expression profile of gene *HXK2* is expressed after the glucose concentration diminishes. Moreover, Pearson correlation coefficients and the *P* value for Pearson correlation coefficient as indicated in figures 5h, l illustrate that the result is consistent with other glucose-limitation genes mentioned previously. It is clear from these experiments that decay rates are important computationally in predicting transcription rate profiles from expression profiles for condition-specific genes, also known as stress-regulated genes.

Although B-spline curving fitting is widely used, the problem of overfitting needs to be further discussed if it leads to different results during analysis. As shown in figure 5, the blue circles represent the value of setting the control point to 7, which equals the number of data points for each given gene where the order of basis polynomials is set to 2 for B-spline curving fitting in this case. In addition, figure 6 demonstrates a series of analyses of the four genes resulting from setting the number of control points to 7 where the order of basis polynomials is set to 3. Shown in figure 6a–d are the expression profiles of four genes where the blue circles represent the control points. Their corresponding Pearson correlation coefficients are shown in figure 6e–h, in which the curves exhibit similar trends as shown in figure 5e–h. From the analysis, we can conclude that the results are the same although the order of basis polynomials for B-spline curving fitting are different.

4. Conclusion

In biology, the mRNA decay rate contributes significantly to determining mRNA expression levels. Besides, reconstructing gene regulatory networks is a challenging task in computational biology. Nevertheless, how mRNA decay rates computationally affect the prediction of transcription rate profiles from expression profiles has not been well studied. In this paper, we present a systematic method to explore the influences of decay rates on the computational prediction of transcription rate profiles from expression profiles. A series of analyses including simulated expression data, real expression data including cell-cycle data and glucose-limitation data in yeast, have been carried out. As for the simulated data, the results indicate that different decay rates will not influence the computational prediction of transcription rate profiles from expression profiles whose feature is periodicity. On the contrary, the results indicate that these transcription rate profiles become dissimilar both in magnitude and shape for those characterized by steep changes. Regarding real data in the case of yeast cell-cycle data, which are regulated by periodic fluctuation in the expression levels, their corresponding transcription rate profiles still maintain consistency in peaking for expression profiles grouped within the same cell-cycle stage. As for glucose-limitation data characterized by rapid growth or decrease in their expression profiles within a fixed transition stage instead of periodic changes, the results show that the magnitudes and shapes of their transcription rate profiles are significantly dependent on the variation in decay rates. According to the findings of this study, we conclude that the decay rate plays a major role on the computational prediction of the transcription rate profiles from expression profiles characterized by aperiodicity; nevertheless, decay rates seem to play a minor role for such periodic expression data. We believe that this study suggests a good possibility for further research on the computational analysis of expression profiles under different biological conditions.

Acknowledgements

The valuable suggestions of Dr Wen-Hsiung Li, Dr Ming-Che Shih, and Dr Mei-Yeh Lu made during the discussion of this work are highly acknowledged. This work was supported in part by National Science Council of Taiwan under the grants NSC94-2213-E001-029 and the Institute of Information Science and the Genomics Research Center of Academia Sinica in Taiwan.

References

Bar-Joseph Z 2004 Analyzing time series gene expression data; *Bioinformatics* **20** 2493–2503

Bar-Joseph Z, Gerber G K, Gifford D K, Jaakkola T S and Simon I 2003 Continuous representations of time-series gene expression data; *J. Comput. Biol.* **10** 341–356

Bernstein J A, Khodursky A B, Lin P H, Lin-Chao S and Cohen S N 2002 Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays; *Proc. Natl. Acad. Sci. USA* **99** 9697–9702

Brauer M J, Saldanha A J, Dolinski K and Botstein D 2005 Homeostatic adjustment and metabolic remodeling in glucose-limited yeast cultures; *Mol. Biol. Cell* **16** 2503–2517

Chen D, Toone W M, Mata J, Lyne R, Burns G, Kivinen K, Brazma A, Jones N and Bahler J 2003 Global transcriptional responses of fission yeast to environmental stress; *Mol. Biol. Cell* **14** 214–229

Chen H C, Lee H C, Lin T Y, Li W H and Chen B S 2004 Quantitative characterization of the transcriptional regulatory network in the yeast cell cycle; *Bioinformatics* **20** 1914–1927

Chen T, He H L and Church G M 1999 Modeling gene expression with differential equations; *Pac. Symp. Biocomput.* 29–40

Cho R J, Campbell M J, Winzeler E A, Steinmetz L, Conway A, Wodicka L, Wolfsberg T G, Gabrielian A E, Landsman D, Lockhart D J and Davis R W 1998 A genome-wide transcriptional analysis of the mitotic cell cycle; *Mol. Cell* **2** 65–73

Fan J, Yang X, Wang W, Wood W H 3rd, Becker K G and Gorospe M 2002 Global analysis of stress-regulated mRNA turnover by using cDNA arrays; *Proc. Natl. Acad. Sci. USA* **99** 10611–10616

Khodursky A B and Bernstein J A 2003 Life after transcription—revisiting the fate of messenger RNA; *Trends Genet.* **19** 113–115

Luan Y and Li H 2003 Clustering of time-course gene expression data using a mixed-effects model with B-splines; *Bioinformatics* **19** 474–482

Nachman I, Regev A and Friedman N 2004 Inferring quantitative models of regulatory networks from expression data; *Bioinformatics* (*Supp1.1*) **20** i248–i256

Raghavan A and Bohjanen P R 2004 Microarray-based analyses of mRNA decay in the regulation of mammalian gene expression; *Brief. Funct. Genomic Proteomic* **3** 112–124

Rogers D F 2000 *An introduction to NURBS: with historical perspective* (San Francisco, CA: Morgan Kaufmann)

Sasik R, Iranfar N, Hwa T and Loomis W F 2002 Extracting transcriptional events from temporal gene expression patterns during *Dictyostelium* development; *Bioinformatics* **18** 61–66

Singer R H and Penman S 1973 Messenger RNA in HeLa cells: kinetics of formation and decay; *J. Mol. Biol.* **78** 321–334

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

Wang Y, Liu C L, Storey J D, Tibshirani R J, Herschlag D and Brown P O 2002 Precision and functional specificity in mRNA decay; *Proc. Natl. Acad. Sci. USA* **99** 5860–5865

Yang E, van Nimwegen E, Zavolan M, Rajewsky N, Schroeder M, Magnasco M and Darnell J E Jr 2003 Decay rates of human mRNAs: correlation with functional characteristics and sequence attributes; *Genome Res.* **13** 1863–1872

Yeung K Y, Haynor D R and Ruzzo W L 2001 Validating clustering for gene expression data; *Bioinformatics* **17** 309–318

Yeung M K, Tegner J and Collins J J 2002 Reverse engineering gene networks using singular value decomposition and robust regression; *Proc. Natl. Acad. Sci. USA* **99** 6163–6168

MS received 6 August 2007; accepted 6 November 2007

*e*Publication: 26 November 2007

Corresponding editor: Reiner A Veitia