Jurnal Teknologi

Full paper

A Method to Construct Gene Regulatory Networks to Estimate and Calculate Time Delays

Chai Suk Phin[a], Mohd Saberi Mohamad[a*]

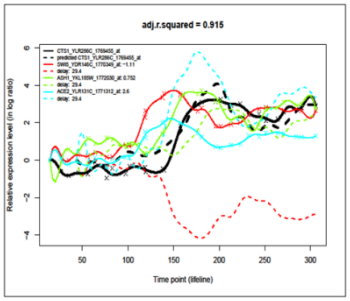
[a]*Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia*

*Corresponding author: saberi@utm.my

Article history

Received :16 November 2011
Received in revised form :
1 January 2013
Accepted :15 April 2013

Graphical abstract



Abstract

In general, the motive of this research is to infer gene regulatory network in order to clarify the basis consequences of biological process at the molecular level. Time course gene expression profiling dataset has been widely used in basic biological research, especially in transcription regulation studies since the microarray dataset is a short time course gene expression dataset and have lots of errors, missing value, and noise. In this research, R library is implemented in this method to construct gene regulatory which aims to estimate and calculate the time delays between genes and transcription factor. Time delay is the parameters of the modeled time delay linear regression models and a time lag during gene expression change of the regulator genes toward target gene expression. The constructed gene regulatory network provided information of time delays between expression change in regulator genes and its target gene which can be applied to investigate important time-related biological process in cells. The result of time delays and regulation patterns in gene regulatory network may contribute into biological research such as cell development, cell cycle, and cell differentiation in any of living cells.

Keywords: Gene expression; gene regulatory network; time delay linear regression model; time delay; time-related biological processes; saccharomyces cerevisiae

Abstrak

Secara umumnya, kajian ini dijalankan adalah untuk mengkaji dan menjelaskan process biologi yang berlaku di tahap molekul. Data gen profiling akan digunakan dalam penyelidikan biologi asas, terutamanya dalam kajian transkripsi regulasi kerana data mikroarray mempunyai kelemahan seperti terkandung data yang kesalahan nilai dan bilangan sample adalah berbeza dengan bilangan gen yang hendak dimodelkan. Dalam kajian ini, R library akan digunakan dalam membina rangkaian gen yang bermatlamat untuk menganggarkan dan mengira masa tunda diantara gen sasaran dan factor transkripsi. Masa tunda adalah parameter dalam model "*Time delay linear regression*" dan juga merupakan masa tunda semasa perubahan ekspresi diantara gen sasaran dan gen regulator. Rangkaian gen yang dibina menyediakan maklumat masa tunda antara gen sasaran dan gen regulator selain berfungsi dalam menyiasat proses biologi yang berkaitan dengan masa dalam sel-sel hidup. Hasil daripada masa tunda dan corak interasi gen dalam rangkaian akan menyumbang dalam penyelidikan biologi seperti pembangunan sel, kitaran sel, dan pembezaan sel dalam mana-mana sel hidup.

Kata kunci. Pengekspresan gene; rangkaian gen; *time delay linear regression* model; masa tunda; proses hayat sel; saccharomyces

© 2013 Penerbit UTM Press. All rights reserved.

1.0 INTRODUCTION

Microarray datasets is the result from a multiplex technology, DNA microarray [2]. With the fast development of microarray technology, time course gene expression profiling data have been generated successfully thus improved the research on gene regulatory networks. However, the step to reverse engineering was very challenging since the combinatorial nature of the problem and the fact, that the available dataset is often few and inaccurate.

In general, all existing and available models and approaches suffer to various limitations such as overfitting, high computational complexity, dimensionality, and experimental measurement as to logical models, continuous models and single-molecule level models. Logical models modeled by the most basic and simplest modeling methodology such as discrete and logic-based [4]. Continuous models using real-valued parameters over a continuous timescale which allow a straightforward comparison of the global state and experimental dataset and can theoretically be more accurate. Single molecule

level models explain the relationship between stochasticity and gene regulation.

Moreover, some software also available and been used for gene network inference such as ARACNE and Banjo. ARACNE (Algorithm for the Reconstruction of Accurate Cellular Networks), well known software of information-theoretic approaches that used to scale up complexity of gene regulatory network by microarray network in mammalian cells, besides address limitation of network deconvolution [5]. Banjo, another software that developed by group of Hartemink that used to infer gene regulatory network [9]. In general, gene regulatory network is the network that shows molecular species and their interactions, which together control gene product abundance [3]. Furthermore, gene regulatory network can be defined as nonlinear differential equations when time delays estimated during gene expression change [8].

This journal shows gene regulatory network is constructed based on interaction between time course gene expression profiling dataset of *S. Cerevisiae* (*Saccharomyces Cerevisiae*) and transcription factors or known as regulator genes. This network is constructed based on two parameters which are time delay and regulation coefficient. The function of gene regulatory network is to provide information about time delays which can be applied to investigate important time-related biological process such as cell development and cell cycle.

■2.0 MATERIALS AND METHODS

In this research, used the dataset from yeast dataset named as *Saccharomyces. Cerevisiae (S. Cerevisiae)* cell cycle data which included both gene expression profiles of wild type cells and cyclin mutant cells which also known as target gene in this research. Besides, a candidate pool of potential regulator genes or known as of target genes also applied in this research in order to analysis the interaction between regulators and target gene.

Table 2.1 The details of dataset

S. Cerevisiae cell cycle dataset	Number of genes
Wild type cells	3725
Cyclin mutant cells	2029
Regulator genes	34

In general, time course gene expression profiling dataset or known as target gene is interpolated by a B spline interpolation followed by modeled time delay linear regression models and lastly construct gene regulatory network. There are 2 different methods to construct gene regulatory as discussed in more details in Section 2.4.1, and 2.4.2.

2.1 B Spline Interpolation

In this research, time points which obtain from experimentation of biological process too complicated and dummy. Thus, B spline interpolation process as shows in Figure 2.1 is done to get interpolated time points dataset within the range of discrete set of known time points in order to get potential time points.

B spline interpolation function to select potential time points from candidate pool of time course gene expression profiling dataset or known as target gene is.used to figure the expression of 100 time points rather than 30 time points in the existing time course dataset.

```
data(wt.expr.data)

wt.bspline.data<- ts.bspline(wt.expr.data, ts.point= as.numeric(colnames(wt.expr.data)),
data.predict=100)
```

Figure 2.1 R command for b spline interpolation

2.2 Time Delay Linear Regression Model

Basically, time delay linear regression models is optimized by a computational algorithm AIC (Akaine Information Criterion) forward selection procedure which assist to plot the linear regression of the expression change between regulatory genes and target genes.

Time delay linear regression models are modeled based on the interpolated expression data from B spline interpolation and candidate pool of regulators. Figure 2.2 displays the R command to model time delay linear regression model. Both single.adj.r.squared and multiple.adj.r.squared indicates the cutoffs of adjusted R squared of single regulator regression and multiple regulator regression. Whereby tf.list represented regulator genes, min.coef and max.coef means the minimum and maximum regression coefficient respectively.

```
data(tf.list)

dir.create('wt.models')

setwd('wt.models')

wt.models<-timedelay.lm.batch(bspline.data=wt.bspline.data, expr.data=wt.expr.data,

regulator.list=tf.list, target.list=rownames(wt.bspline.data),

single.adj.r.squared=0.8, multiple.adj.r.squared=0.9, maxdelay=ncol(wt.bspline.data)*0.1,

min.coef=0.25, max.coef=4, output=T, topdf=T,

xlab='Time point (lifeline)', ylab='Relative expression level (in log ratio)')
```

Figure 2.2 R command to model time delay linear regression models

2.3 Computational Algorithm

An AIC (Akaine information criterion) computational algorithm is used to select possible regulatory genes with certain time. This algorithm used iteratively during the selection of possible regulatory genes from a candidate pool of regulatory genes.

There are two major parameters in this formula which essential in selecting regulatory genes among in candidate pool. The k_p is the number of parameters in the statistical model, and L is the maximized value of the likelihood function for the estimated model.

$$AIC = 2\ k_p - 2\ln(L) \quad (2.1)$$

The computational procedure of this algorithm has two steps. First step is the sorting of regulatory genes based on their relevance with target gene. Second step was optimization of the time delay linear regression model with forward selection

procedure for regulatory genes and time delays. Potential regulatory genes toward certain target gene with time delays either in activation or regression expression are constructed in this step. From the equation, tf''_i define the transcription factor and M define as the smallest AIC of regulator genes. The possible regulator genes can be expression by the following equation:

$$S = \{ tf''_1, tf''_2, \ldots, tf''_i \} \; (1 \leq i \leq M) \qquad (2.2)$$

If 1 < 2, tf_1 has smaller AIC than tf_2 and thus tf_1 can be considered better than tf_2.

2.4 Construction of Gene Regulatory Network

The gene regulatory network is constructed based on the time delay linear regression models as discuss in Section 2.2. In this research, two distinct methods are implemented to construct gene regulatory network as discussed in Section 2.4.1, and 2.4.2.

2.4.1 Method to Construct Gene Regulatory Network for a Single Target Gene

This method as shows in Figure 2.3 is used to construct gene regulatory network which specific to a single target gene.

Firstly, time course gene expression profiling dataset is interpolated by B spline interpolation followed by model time delay linear regression model from interpolated dataset. Lastly, construct gene regulatory network for a single target gene.

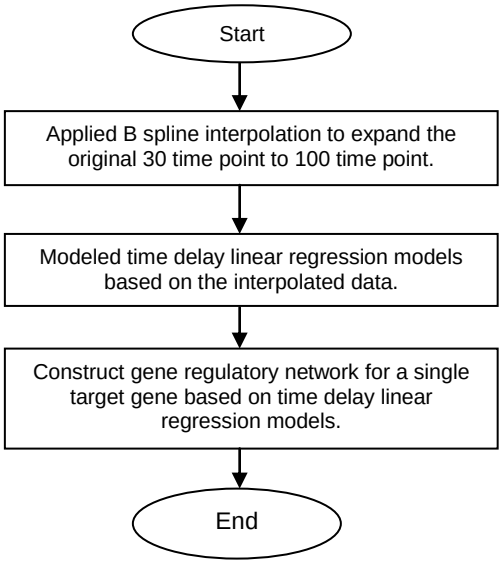


Figure 2.3 Flow chart to construct gene regulatory network for a single target gene

Figure 2.4 shows the R code that function to construct gene regulatory network for a single target gene such as CTS1_YLR286C_1769455_at. This R code can construct gene regulatory network of another target gene by displaced the name of desire target gene.

```
require(igraph)

node<-unique(c(as.character(data[,1]),as.character(data[,2])))
node.num<-1:length(node)-1; names(node.num)<-node
g<-
graph(rbind(node.num[as.character(data[,1])],node.num[as.character(data[,2])]),directed=TRU
E)

vertex.color<-rep('white',length(node)); names(vertex.color)<-node
vertex.color[unique(as.character(data[,1]))]<-'lightblue'

pdf('CTS1_YLR286C_1769455_at.pdf',width=5,height=5)
plot.igraph(g,vertex.label=node,vertex.color=vertex.color,vertex.size=2,
layout=layout.fruchterman.reingold)
dev.off()
```

Figure 2.4 R code to construct gene regulatory network for a single target gene

2.4.2 Method To Construct Activation and Regression Gene Regulatory Network

In this method as displays in Figure 2.5 there is an additional step of regulator genes filtration before construction of gene regulatory network. Firstly, time course gene expression profiling dataset is undergoes B spline interpolation. Secondly, model time delay linear regression models from interpolated dataset. Thirdly, filter regulator genes according to activation and regression coefficient followed by construct both activation and regression gene regulatory network separately.

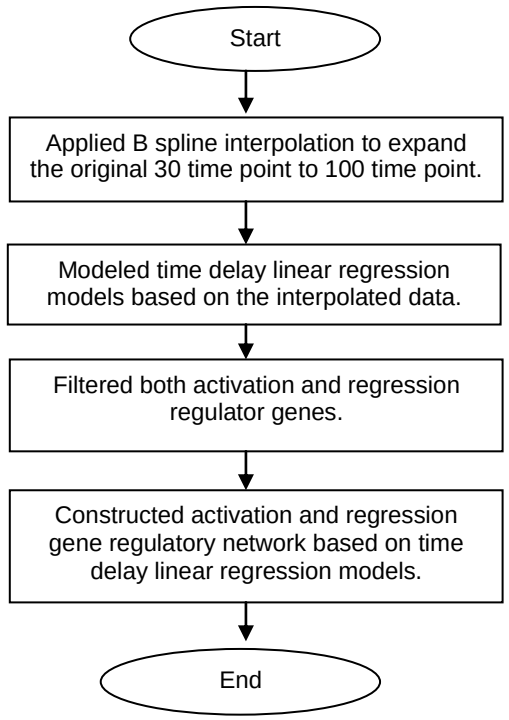


Figure 2.5 The flow chart of method to construct activation and regression gene regulatory network

Both activation and regression gene regulatory network is filtered respectively under different R code as showed in Figure 2.6 and Figure 2.7.

```
activation<-array(dim=c(3725,5))
dim(activation)
for(i in 1:3725)
{
if(wt.models[i,3]>0)
{
activation<-
data.frame(regulator=wt.models[i,1],target=wt.models[i,2],coefficient=wt.models[i,3],time_del
ay=wt.models[i,4])
i<-i+1
}
}

for(i in 1:3725)
{
if(wt.models[i,3]>0)
{
activation[i,]<-
data.frame(regulator=wt.models[i,1],target=wt.models[i,2],coefficient=wt.models[i,3],time_del
ay=wt.models[i,4])
i<-i+1
}
}
write.csv(activation,file='Activation..csv')
```

Figure 2.6 R code for filter activation regulator Genes

```
regression<-array(dim=c(3725,5))
dim(regression)
for(i in 1:3725)
{
if(wt.models[i,3]>0)
{
regression<-
data.frame(regulator=wt.models[i,1],target=wt.models[i,2],coefficient=wt.models[i,3],time_del
ay=wt.models[i,4])
i<-i+1
}
}

for(i in 1:3725)
{
if(wt.models[i,3]>0)
{
regression[i,]<-
data.frame(regulator=wt.models[i,1],target=wt.models[i,2],coefficient=wt.models[i,3],time_del
ay=wt.models[i,4])
i<-i+1
}
}
write.csv(regression,file='Regression..csv')
```

Figure 2.7 R code for filter regression regulator gene

After filtration function, gene regulatory network is constructed by the R code shows in Figure 2.8 that function to construct activation and regression gene regulatory network respectively after the filtration.

```
x<-load("repression.RData")
x
rep
require(igraph)
node<-unique(c(as.character(rep[,1]),as.character(rep[,2])))
node.num<-1:length(node)-1; names(node.num)<-node
g<-
graph(rbind(node.num[as.character(rep[,1])],node.num[as.character(rep[,2])]),directed=TRUE)
vertex.color<-rep('white',length(node)); names(vertex.color)<-node
vertex.color[unique(as.character(rep[,1]))]<-'lightblue'
edge.color<-ifelse(rep[,3]<0,'red','green')
C.label<-sapply(rep[,3],format,digits=2)
pdf('Repression_Network',width=70,height=70)
plot.igraph(g,vertex.label=node,vertex.color=vertex.color,edge.label=paste('coef:',C.label),edge
.color=edge.color,vertex.size=2,layout=layout.fruchterman.reingold)
dev.off()
```

Figure 2.8 R code to construct activation and regression gene regulatory network

3.0 RESULTS AND DISCUSSION

In section 3.0, there are discussion, analysis and evaluation of the interaction between time course gene expression profiling dataset of *S. Cerevisiae* and transcription factors or known as regulator genes in the constructed gene regulatory network. Besides, there are comparisons between constructed gene regulatory networks based on the highlighted methods in Section 2.4.1 and 2.4.2.

3.1 Interpolated Time Course Gene Expression Profiling Dataset

Figure 3.1 showed time course gene expression profiling dataset before been interpolated. Column A indicated the name of target genes while column B to column J indicated the time points. In this research, the time point indicated as time delays during expression change between regulator and target gene.

For example coefficient value for gene CTSI_YLR286C at minute 26.78193 was -0.58905 which indicated a regression gene expression which shown that gene CTS1_YLR286C was undergoes regression regulation in this specific time. Another example of activation gene expression was gene DSE4_YNR067C at minute 35.3926 with the coefficient value of 0.130154.

Coefficient value indicated regulation effect in gene expression, a positive coefficient indicated activation or promoting gene expression of target gene and negative indicated regression means inhibiting gene expression of target gene.

	A	B	C	D	E	F	G	H	I	J
1	Time point / Target Gene	14.62687	26.78193	35.3926	45.60104	56.15834	64.42014	76.92408	83.23924	97.68981
2	CTS1_YLR286C_1769455_at	0	-0.58905	-0.82225	-0.65929	-0.74501	-0.11055	-0.95176	-0.42351	-0.72038
3	DSE4_YNR067C_1769988_at	0	-0.06041	0.130154	0.215153	0.214446	0.469136	0.014842	0.114211	0.162075
4	PST1_YDR055W_1770307_at	0	0.628727	0.756349	0.255029	-0.1052	0.030054	-0.01661	-0.21087	-0.19102
5	DSE1_YER124C_1770749_at	0	0.397707	0.443564	0.861063	0.094602	0.873591	0.191042	0.338514	0.446323
6	EGT2_YNL327W_1774263_at	0	-0.68079	-1.53372	-1.19086	-1.05659	-1.50566	-1.68936	-0.90586	-1.53798
7	HO_YDL227C_1778270_at	0	-0.05659	0.028951	-0.00758	0.266411	0.187989	0.252729	0.225554	0.423589
8	PCL9_YDL179W_1778541_at	0	0.177076	-0.19931	0.069784	-0.1086	0.062086	-0.1448	-0.12908	-0.20183
9	YPL158C_YPL158C_1778564_at	0	0.009156	0.140219	0.470988	0.433849	0.571179	0.303978	0.329201	0.361302
10	DSE2_YHR143W_1779965_at	0	-0.01773	0.003675	0.452465	0.144455	0.823948	0.495256	0.798602	1.304077
11	YLR049C_YLR049C_1771810_at	0	0.234349	-0.29379	-0.05638	0.649909	0.941304	2.482713	2.385914	3.185157
12	CDC46_YLR274W_1771390_at	0	0.327574	0.042146	0.358068	0.486082	0.775656	0.662785	0.742838	0.128468
13	YBR071W_YBR071W_1779265_at	0	0.146123	-0.07342	0.092225	0.742074	1.068339	2.562366	2.574087	2.884933
14	YOR29-17_YOR066W_1774677_at	0	-0.33726	0.179617	-0.01898	0.914771	0.960767	1.223441	1.316829	0.628907
15	GAS3_YMR215W_1773766_at	0	-0.05766	-0.31498	-0.38949	0.249097	0.217923	0.945288	0.699373	1.75599
16	ALK1_YGL021W_1776407_at	0	0.652269	0.17185	0.433324	0.07819	0.804844	0.4702	0.684221	1.791594
17	VIK1_YPL253C_1769350_at	0	0.504055	-0.09064	0.125306	-0.02374	0.258257	0.15251	0.199637	0.752934
18	PIR3_YKL163W_1772229_at	0	0.496935	-0.19901	-0.44067	-0.66615	-0.73538	-0.52786	-0.82737	-0.92968
19	CRH1_YGR189C_1776793_at	0	0.811286	-0.2497	1.070603	1.506441	2.076286	3.463108	3.474587	3.85074
20	YNR009W_YNR009W_1776088_at	0	-0.0196	0.027935	0.102228	0.5085	0.532655	1.648918	1.683574	3.00505

Figure 3.1 Original time course gene expression profiling dataset

In this research, time points which obtain from experimentation of biological process were too complicated and dummy. Thus, interpolation process was done to get interpolated time points dataset within the range of discrete set of known time points in order to get potential time points.

Figure 3.2 shows the time course gene expression profiling dataset after interpolated. As compared to Figure 3.1, the time points in Figure 3.2 were short apart for example from minute 14.62687 to minute 17.56756, which means the evaluation of gene expression was more accurate and specific than those experiment that applied on dataset without been interpolated. For example at time point of minute 14.62687, it showed a valuable coefficient value "2.33E-06" rather than showed "0" as in Figure 3.1.

	A	B	C	D	E	F	G	H	I	J
1	Time Point / Target Gene	14.62687	17.56756	20.50826	23.44895	26.38964	29.33034	32.27103	35.21173	38.15242
2	CTS1_YLR286C_1769455_at	2.33E-06	0.017145	-0.11474	-0.32946	-0.56085	-0.74272	-0.82318	-0.8198	-0.77126
3	DSE4_YNR067C_1769988_at	8.39E-07	-0.12811	-0.16511	-0.13728	-0.0709	0.007728	0.075984	0.129011	0.167351
4	PST1_YDR055W_1770307_at	-9.92E-08	0.047639	0.199731	0.403166	0.60483	0.751611	0.80088	0.760986	0.655763
5	DSE1_YER124C_1770749_at	1.58E-06	0.788315	0.970496	0.777639	0.440841	0.191194	0.209244	0.42962	0.712285
6	EGT2_YNL327W_1774263_at	1.61E-06	0.3398	0.248287	-0.1209	-0.61414	-1.07778	-1.38354	-1.52639	-1.53875
7	HO_YDL227C_1778270_at	-3.37E-07	-0.22183	-0.26368	-0.19182	-0.07248	0.028081	0.057499	0.030953	-0.01587
8	PCL9_YDL179W_1778541_at	5.45E-07	0.638243	0.768255	0.570284	0.224583	-0.0886	-0.22359	-0.20295	-0.10029
9	YPL158C_YPL158C_1778564_at	3.55E-07	0.139281	0.151791	0.092759	0.017414	-0.01902	0.025992	0.133144	0.264379
10	DSE2_YHR143W_1779965_at	9.73E-07	0.461206	0.507328	0.304216	0.017718	-0.18632	-0.17863	-0.00794	0.222951
11	YLR049C_YLR049C_1771810_at	-1.59E-06	0.5547	0.696402	0.558934	0.276124	-0.0182	-0.2115	-0.29485	-0.29077
12	CDC46_YLR274W_1771390_at	8.41E-07	0.548648	0.701716	0.594129	0.360814	0.136695	0.032426	0.040582	0.11789
13	YBR071W_YBR071W_1779265_at	-1.13E-06	0.21247	0.281514	0.249884	0.160332	0.055611	-0.02668	-0.07397	-0.08132
14	YOR29-17_YOR066W_1774677_at	3.97E-07	-1.1081	-1.33478	-0.99817	-0.41637	0.092486	0.274794	0.190764	-0.00411
15	GAS3_YMR215W_1773766_at	-1.70E-06	-0.02438	-0.02987	-0.03361	-0.05275	-0.10442	-0.19875	-0.31153	-0.4082
16	ALK1_YGL021W_1776407_at	5.01E-07	1.131979	1.436617	1.202929	0.719931	0.276638	0.10558	0.164512	0.327753
17	VIK1_YPL253C_1769350_at	9.90E-08	0.978889	1.235485	1.016322	0.56793	0.13684	-0.07626	-0.09374	-0.00565
18	PIR3_YKL163W_1772229_at	-7.18E-07	0.766819	1.001923	0.872681	0.546462	0.190635	-0.0561	-0.19447	-0.26755
19	CRH1_YGR189C_1776793_at	-8.76E-07	2.099558	2.576869	2.004868	0.956492	0.004679	-0.3851	-0.27011	0.133628
20	YNR009W_YNR009W_1776088_at	-1.48E-06	-0.07358	-0.08808	-0.06473	-0.0248	0.010489	0.024735	0.025156	0.026134

Figure 3.2 Interpolated time course gene expression profiling dataset

3.2 Gene Regulatory Network for One Target Gene

A gene regulatory network for a single target gene is shows in Figure 3.3. In Figure 3.3, the target gene is regulated by three regulators ACE2_YLR131C_1771312_at, ASH1_YKL185W_1772030_at, and SWI5_YDR146C_1770349_at.

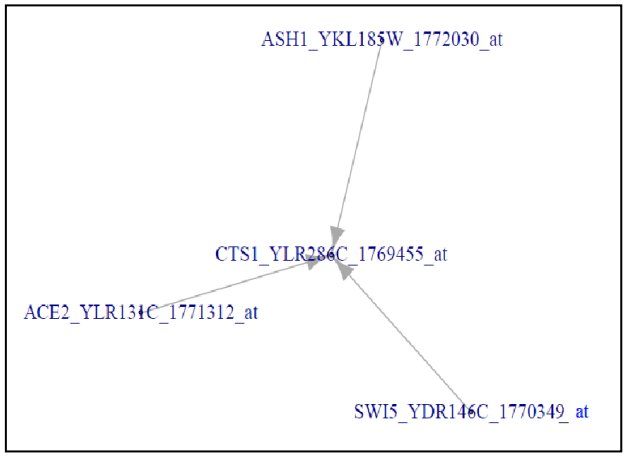


Figure 3.3 Gene Regulatory Network for CTS1_YLR286C_1769455_at

A time delay linear regression model for target gene CTS1_YLR286C_1769455_at is shows in Figure 3.4. Figure 3.4 shows expression pattern of target gene CTS1_YLR286C_1769455_at with its regulator genes. The solid lines display the interpolated values. The black dotted lines shows the values of target gene which fitted by the model. Moreover, the coloured dotted lines displayed the regulatory contribution of each regulator toward target gene CTS1_YLR286C_1769455_at. Time delays of each transcription factors also displays in Figure 3.4 as time delays for regulator genes such as SWI5_YDR146C_at, ASH1_YKL185W_at, and ACE2_YLR131C_at are the same in 29.4 minutes.

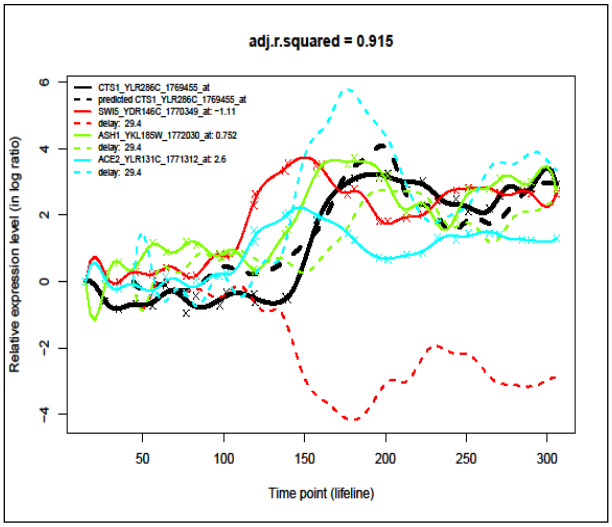


Figure 3.4 Time delay linear regression model for CTS1_YLR286C_1769455_at

The main idea to construct gene regulatory network is shows a single target gene interacted with all other possible regulator genes clearly. Thus, certain target gene can be easily analysed besides identified relationships between regulator genes and target gene. Moreover, the time delay for each regulator gene that interacts with target gene also can be calculated as the time delay is used to determine the gene expression change of the regulator genes toward target gene expression

3.3 Activatation and Regression Gene Regulatory Network

In this research, gene regulatory network also constructed into activation and regression gene regulatory networks. Figure 3.5 displays the activation gene regulatory network and Figure 3.6 displays the detail gene expression between regulator genes and target genes which highlighted in Figure 3.5. Figure 3.5 is the activation gene regulatory network that shows positive coefficient which means the gene expression only undergoes activation regulation. For example, regulator ACE2_TLR131C (in blue circle) positively regulated target genes of NIS1_YNL078W, PCL9_YDL179W, ASE1_YOP058C, and PST1_YDR055W with coefficient values of 0.76, 0.80, 0.32, and 0.25 respectively.

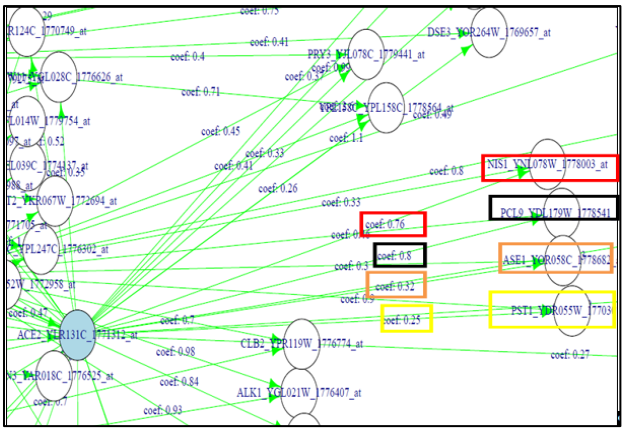


Figure 3.5 Activation gene regulatory network

	A	B	C	D
1	regulator	target	coefficient	time_delay
2	DAL81_YIR023W_1771154_at	MCM1_YMR043W_1770949_at	0.277031311	5.881389422
3	ASH1_YKL185W_1772030_at	CTS1_YLR286C_1769455_at	0.751722487	29.40694711
4	ACE2_YLR131C_1771312_at	CTS1_YLR286C_1769455_at	2.601066591	29.40694711
5	SWI5_YDR146C_1770349_at	DSE4_YNR067C_1769988_at	0.368407556	29.40694711
6	ASH1_YKL185W_1772030_at	DSE4_YNR067C_1769988_at	0.392964928	29.40694711
7	ACE2_YLR131C_1771312_at	DSE4_YNR067C_1769988_at	1.060522352	29.40694711
8	TEC1_YBR083W_1778960_at	DSE4_YNR067C_1769988_at	0.269487796	2.940694711
9	SWI5_YDR146C_1770349_at	PST1_YDR055W_1770307_at	0.898301629	29.40694711
10	ACE2_YLR131C_1771312_at	PST1_YDR055W_1770307_at	0.250298019	29.40694711
11	ASH1_YKL185W_1772030_at	DSE1_YER124C_1770749_at	0.787110861	14.70347356
12	TEC1_YBR083W_1778960_at	DSE1_YER124C_1770749_at	0.492516578	23.52555769
13	ACE2_YLR131C_1771312_at	DSE1_YER124C_1770749_at	0.515739221	29.40694711
14	ASH1_YKL185W_1772030_at	HO_YDL227C_1778270_at	1.215190994	23.52555769
15	ACE2_YLR131C_1771312_at	HO_YDL227C_1778270_at	0.986129212	26.4662524
16	SWI5_YDR146C_1770349_at	PCL9_YDL179W_1778541_at	0.463120804	29.40694711
17	ACE2_YLR131C_1771312_at	PCL9_YDL179W_1778541_at	0.805282175	29.40694711
18	SWI5_YDR146C_1770349_at	YPL158C_YPL158C_1778564_at	0.406472989	29.40694711
19	ASH1_YKL185W_1772030_at	YPL158C_YPL158C_1778564_at	0.707470688	0
20	ACE2_YLR131C_1771312_at	YPL158C_YPL158C_1778564_at	0.258384843	26.4662524

Figure 3.6 Time delays of activation regulation

Figure 3.7 shows regression gene regulatory network which displays negative coefficient which means gene expression only undergoes regression regulation. For example, regulator PUT3_YKL015W (in blue circle) negatively regulated target genes of BNI4_YNL233W, SLD5_YDR489W, ADR3_YER170W, GAS2_YLR343W, ERP4_YOR016C, and RTT105_TER104W with coefficient values of -0.32, -0.35, -0.29, -0.49, -0.61, and -0.35 respectively. Figure 3.8 displays the detail gene expression between regulator genes and target genes which highlighted in Figure 3.7.

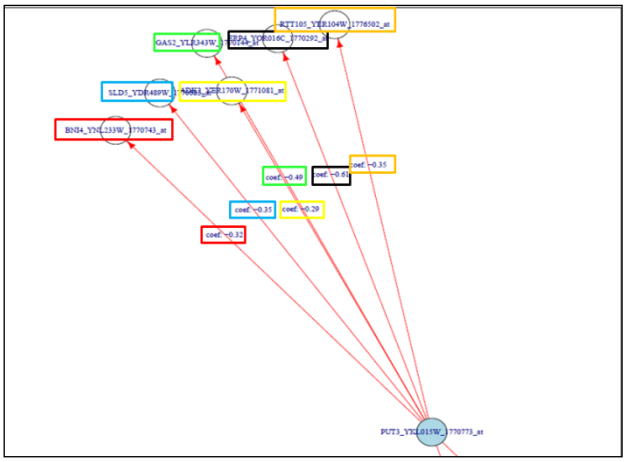


Figure 3.7 Regression gene regulatory network

	A	B	C	D
2	SWI5_YDR146C_1770349_at	CTS1_YLR286C_1769455_at	-1.114945761	29.40694711
3	PUT3_YKL015W_1770773_at	GAS2_YLR343W_1770144_at	-0.494500102	20.58486298
4	TEC1_YBR083W_1778960_at	PST1_YDR055W_1770307_at	-0.349518944	11.76277884
5	PUT3_YKL015W_1770773_at	BNI4_YNL233W_1770743_at	-0.322133856	17.64416827
6	SWI5_YDR146C_1770349_at	HO_YDL227C_1778270_at	-0.321598461	29.40694711
7	PUT3_YKL015W_1770773_at	SLD5_YDR489W_1770683_at	-0.347851839	5.881389422
8	TEC1_YBR083W_1778960_at	HO_YDL227C_1778270_at	-0.645948816	11.76277884
9	TEC1_YBR083W_1778960_at	YPL158C_YPL158C_1778564_at	-0.290427812	8.822084134
10	PUT3_YKL015W_1770773_at	RTT105_YER104W_1776502_at	-0.350644074	20.58486298
11	FHL1_YPR104C_1773575_at	YPL158C_YPL158C_1778564_at	-0.263949051	14.70347356
12	ACE2_YLR131C_1771312_at	DSE2_YHR143W_1779965_at	-0.643582038	26.4662524
13	PUT3_YKL015W_1770773_at	ADK2_YER170W_1771081_at	-0.29244256	23.52555769
14	FKH1_YIL131C_1777130_at	DSE2_YHR143W_1779965_at	-0.914881632	29.40694711
15	INO4_YOL108C_1774516_at	YLR049C_YLR049C_1771810_at	-0.580901147	0
16	FKH1_YIL131C_1777130_at	YLR049C_YLR049C_1771810_at	-0.310125289	11.76277884
17	YHP1_YDR451C_1778368_at	YLR049C_YLR049C_1771810_at	-0.569510227	29.40694711
18	CDC14_YFR028C_1779003_at	YBR071W_YBR071W_1779265_at	-0.262802619	0
19	PUT3_YKL015W_1770773_at	ERP4_YOR016C_1770292_at	-0.607208007	0
20	STB1_YNL309W_1771976_at	YBR071W_YBR071W_1779265_at	-0.567497593	26.4662524

Figure 3.8 Time delays of regression regulation

These two different gene regulatory network (activation and regression) constructed in order to increase the efficiency of investigate and evaluate those activation and regression gene expression in biological processes. Thus, more accurate and precise identification of gene expression between regulator genes and target gene able to done.

3.4 Analysis Gene Expression Profiles of Wild Type Cells and Cyclin Mutant Cells

In this research, *S. Cerevisiae* cell cycle dataset was used to construct gene regulatory network. Two dataset such as wild type cells and cyclin mutant cells of *S. Cerevisiae* were applied to analysis the biological process.

Figure 3.9 showed activation expression data of wild type cells and cyclin mutant cells. In wild type cells, ALK_YGL021W target gene was regulated by SWI5_YDR146C and YHP1_YDR451C regulator genes respectively. Besides for cyclin mutant cells, ALK1_YGL021W target gene was regulated by five regulator genes such as YHP1_YDR451C, DIG2_YDR480W, MET4_YNL103W, CDC14_YFR028C, and ACE2_YLR131C.

Wild Type Cells				Cyclin Mutant Cells			
regulator	target	coefficien	time_delay	regulator	target	coefficien	time_delay
DAL81_YIF	MCM1_YM	0.277031	5.881389	SWI5_YDR	YPL158C_Y	0.956646	15.79841
ASH1_YKL	CTS1_YLR2	0.751722	29.40695	ACE2_YLR	YPL158C_Y	1.632384	0
ACE2_YLR	CTS1_YLR2	2.601067	29.40695	YOX1_YMI	YLR049C_Y	0.367727	9.874004
SWI5_YDR	DSE4_YNR	0.368408	29.40695	CIN5_YOR	YLR049C_Y	0.721434	15.79841
ASH1_YKL	DSE4_YNR	0.392965	29.40695	FKH1_YIL1	YLR049C_Y	1.256383	0
ACE2_YLR	DSE4_YNR	1.060522	29.40695	DIG2_YDR	YLR049C_Y	0.809441	0
TEC1_YBR	DSE4_YNR	0.269488	2.940695	YOX1_YMI	YBR071W_	0.903657	0
SWI5_YDR	PST1_YDR	0.898302	29.40695	FKH1_YIL1	GAS3_YMI	1.116302	0
ACE2_YLR	PST1_YDR	0.250298	29.40695	MET4_YNL	GAS3_YMI	0.48906	0
ASH1_YKL	DSE1_YER	0.787111	14.70347	YHP1_YDR	ALK1_YGL	0.499927	13.82361
TEC1_YBR	DSE1_YER	0.492517	23.52556	DIG2_YDR	ALK1_YGL	0.284782	19.74801
ACE2_YLR	DSE1_YER	0.515739	29.40695	MET4_YNL	ALK1_YGL	0.286786	7.899203
SWI5_YDR	ALK1_YGL	0.979731	0	CDC14_YF	ALK1_YGL	0.349794	1.974801
YHP1_YDR	ALK1_YGL	0.381451	0	ACE2_YLR	ALK1_YGL	0.677066	1.974801
SWI5_YDR	PCL9_YDL	0.463121	29.40695	CIN5_YOR	PIR3_YKL1	0.387417	19.74801
ACE2_YLR	PCL9_YDL	0.805282	29.40695	YOX1_YMI	CRH1_YGR	0.89531	0
SWI5_YDR	YPL158C_Y	0.406473	29.40695	RME1_YGF	CRH1_YGR	0.370135	19.74801
ASH1_YKL	YPL158C_Y	0.707471	0	YHP1_YDR	CRH1_YGR	0.966346	0
ACE2_YLR	YPL158C_Y	0.258385	26.46625	CIN5_YOR	CRH1_YGR	0.342451	9.874004
ASH1_YKL	DSE2_YHR	1.120879	14.70347	YOX1_YMI	YNR009W_	0.354533	17.77321
SWI5_YDR	DSE2_YHR	0.931697	14.70347	DIG2_YDR	YNR009W_	0.889088	3.949602
TEC1_YBR	DSE2_YHR	0.757826	29.40695	MET4_YNL	YNR009W_	0.428989	0
YHP1_YDR	DSE2_YHR	0.521296	20.58486	INO4_YOL	YNR009W_	0.476867	15.79841

Figure 3.9 activation expression data of wild type cells and cyclin mutant cells

Based on Figure 3.10, which showed regression expression data of wild type cells and cyclin mutant cells. In wild type cells, target gene of YLR049C_YLR049C was regulated by three regulator genes such as INO4 _YOL108C, FKH1 _YIL131C, and YHP1 _YDR451. While for cyclin mutant cells, YLR049C_YLR049C target gene was regulated by four regulator genes such as RME1_YGR044C, MET4_YNL103W, YHP1_YDR451C, and CDC14 _YFR028C.

Wild Type Cells				Cyclin Mutant Cells			
regulator	target	coefficien	time_delay	regulator	target	coefficien	time_delay
SWI5_YDR	CTS1_YLR2	-1.11495	29.40695	CDC14_YF	UBP9_YER	-0.46382	0
TEC1_YBR	PST1_YDR	-0.34952	11.76278	RME1_YGF	YLR049C_\	-0.69157	19.74801
SWI5_YDR	HO_YDL22	-0.3216	29.40695	MET4_YNL	YLR049C_\	-0.33792	0
TEC1_YBR	HO_YDL22	-0.64595	11.76278	YHP1_YDR	YLR049C_\	-0.62514	13.82361
TEC1_YBR	YPL158C_\	-0.29043	8.822084	CDC14_YF	YLR049C_\	-1.57028	0
FHL1_YPR	YPL158C_\	-0.26395	14.70347	HCM1_YCI	YBR071W_	-0.29247	19.74801
ACE2_YLR	DSE2_YHR	-0.64358	26.46625	RME1_YGF	GAS3_YMI	-0.60998	19.74801
FKH1_YIL1	DSE2_YHR	-0.91488	29.40695	CDC14_YF	GAS3_YMI	-1.07153	0
INO4_YOL	YLR049C_\	-0.5809	0	STB1_YNL	GAS3_YMI	-0.26019	0
FKH1_YIL1	YLR049C_\	-0.31013	11.76278	FKH1_YIL1	PIR3_YKL1	-0.58186	13.82361
YHP1_YDR	YLR049C_\	-0.56951	29.40695	CDC14_YF	PIR3_YKL1	-1.04624	3.949602
CDC14_YF	YBR071W_	-0.2628	0	DIG2_YDR	PIR3_YKL1	-0.40911	13.82361
STB1_YNL	YBR071W_	-0.5675	26.46625	RME1_YGF	PIR3_YKL1	-0.48397	3.949602
INO4_YOL	YBR071W_	-0.36134	23.52556	MET4_YNL	PIR3_YKL1	-0.41639	1.974801
MET32_YD	YOR29-17_	-1.85138	23.52556	INO4_YOL	CRH1_YGR	-0.26385	15.79841
MET4_YNL	GAS3_YMI	-1.52254	11.76278	RME1_YGF	YNR009W_	-0.476	17.77321
FHL1_YPR	GAS3_YMI	-0.77575	29.40695	CIN5_YOR	YNR009W_	-0.28051	13.82361
CDC14_YF	GAS3_YMI	-1.18067	0	SWI5_YDR	CLB1_YGR	-0.43711	15.79841
CIN5_YOR	GAS3_YMI	-0.58464	26.46625	RME1_YGF	SUR7_YML	-0.6323	17.77321
INO4_YOL	ALK1_YGL	-0.57964	0	YOX1_YMI	SUR7_YML	-0.26981	1.974801
FKH2_YNL	ALK1_YGL	-0.27471	8.822084	MET4_YNL	SUR7_YML	-0.33495	17.77321

Figure 3.10 Regression expression data of wild type cells and cyclin mutant cells

■4.0 CONCLUSION

Expression profiling technology has widely used and grown in biological field as used of expression profiling dataset in constructing gene regulatory network. With the development of this method, a new method was proposed to construct gene regulatory network which also to estimate and calculate the time delays between the transcription factors and target genes. These type of gene regulatory network not only constructed and displayed interaction between transcription factors and target genes but also showed time delays and regulation patterns either activation or regression. This approach can be used to apply in investigation of important time-related gene expression as in biological processes for example cell cycle, cell differentiation and development. Moreover also useful in research based which discovering and studying the mechanisms, pathways of both transcription factors and target genes.

Acknowledgements

We would like to thank Malaysian Ministry of Science, Technology and Innovation for supporting this research by an e-science research grant (Grant number: 06-01-06-SF1029). This research is also supported by a GUP research grant (Grant number: Q.J130000.2507.04H16) that was sponsored by Universiti Teknologi Malaysia.

References

[1] Huang, T., Liu, L., Qian, Z., Tu, K., Li, Y., and Xie, L. 2010. Using GeneReg to Construct Time Delay Gene Regulatory Networks. *BMC Research Notes*. 3(142).

[2] Huynh, H. T., Kim, J. J., and Won, Y. 2009. Classification Study on DNA Microarray with Feedforward Neural Network Trained by Singular Value Decomposition. *International Journal of Bio-Science and Bio-Technology.* 1(1).

[3] Karlebach, G., and Shamir, R. 2008. Modelling and Analysis of Gene Regulatory Network. *Nat Review Mol Cell Biol*. 9: 770–779.

[4] Kauffman, S., Peterson, C., Samuelsson, B., and Troein, C. 2003. Random Boolean Network Models and The Yeast Transcriptional Network. *Proc. Natl Acad. Sci. USA 100*. 14796–14799.

[5] Margolin, A. A., Nemenman, I., Basso, K., Chris, W., tolovitzky, G., Favera, R. D., and Califano, A. 2006. ARACNE: An Algorithm for Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context. *BMC Bioinformatics S1*. 1471–2105.

[6] Oliver, S. G. 1997. From Gene to Screen with Yeast. Genetics & Development. 7: 405–409.

[7] Wang, Y., Joshi, T., Zhang, X. S., Xu, D., and Chen, L. 2006. Inferring Gene Regulatory Network from Multiple Microarray Datasets. BMC Bioinformatics. 22: 2413–2420.

[8] Wu, F. S. 2011. Delays-Independent Stability of Genetic Regulatory Networks. *IEEE Transaction on Neural Networks.* 22(11).

[9] Yu, J., Smith, V. A., Wang, P. P., Hartemink, A. J., and Jarvis, E. D. 2004. Advances to Bayesian Network Inference for Generating Causal Networks from Observational Biological Data. *Bioinformatics.* 20: 3594–3903.