Computational Models for Efficient Reconstruction of Gene Regulatory Network

ZHANG, Qing

A Thesis Submitted in Partial Fulfilment
of the Requirements for the Degree of
Doctor of Philosophy
in
Biology

The Chinese University of Hong Kong
September 2011

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

Thesis/Assessment Committee

Professor LAM Hon Ming (Chair)

Professor SUN Sai Ming (Thesis Supervisor)

Professor GUO Dian Jing (Thesis Supervisor)

Professor CHAN Ting Fung (Committee Member)

Professor TANG Lei Han (External Examiner)

Declaration

I, Zhang Qing, declare that this thesis represents my own work, except where due acknowledgement is made and that it has not been previously included in a thesis, dissertation or report submitted to this university or to any other institution for degree, diploma or othter qualifications.

Signed Zhang Qing

ZHANG QING

Abstract of thesis entitled:

Computational Models for Efficient Reconstruction of Gene Regulatory Network

Submitted by ZHANG, Qing

for the degree of Doctor of Philosophy

at The Chinese University of Hong Kong in July 2011

The transcriptional regulation of genes plays vital roles in the processes of cellular responding to internal and external stimuli, differentiation and morphogenesis of living organisms. Microarray technology have been developed to detect the gene expression level in large scale, which enables systematic studies of gene regulatory networks to reveal the mechanisms that underlie the cellular processes. Reconstruction of gene regulatory network (GRN) is an important computational strategy used for knowledge discovery in system biology.

This thesis focuses on gene regulatory network reconstruction from high throughput biological data.

In the first part of this thesis, I develop a novel method-DBoMM{Difference BIC(Bayesian Information Criteria) of Mixture Model} to fit the gene expression profiles into a mixture Gaussian model and estimate the 'similarity' or 'distance' of two genes by comparing the likelihood scores of different models. I show that the DBoMM top-performed other 3 existing methods, including Pearson Correlation(COR), Euclidean distance(EUC) and

Mutual Information(MI), using the synthetic dataset. The performance was comparable to MI using the *E.coli* dataset. DBoMM can also identify condition-dependent regulatory interactions and is robust to noisy data.

I then extend the mixture distribution model used for gene network inference to a quantitative model with predictive function, which is in need by both wet-lab experimental design and synthetic biology. By inferring the conditional distribution of related gene expression, we can predict the gene expression profiles under a wide range of experimental conditions, e.g., gene knock-out, gene over-expression, and transcriptional network rewiring. Also, by linking a new experimental condition to the known conditions, the model can be used to reveal the possible functional relationships between different conditions.

In the second part of my thesis, I propose a sub-space greedy search method for efficient Bayesian Network inference. Bayesian Network (BN) has been successfully used to infer the regulatory relationships of genes from microarray data. However, one major limitation of BN approach is the computational cost because the calculation time grows more than exponentially with the dimension of the dataset. This method limits the greedy search space by only selecting gene pairs with higher partial correlation coefficients. Using both synthetic and real data, we demonstrate that the proposed method achieved comparable results with standard greedy search method, and yet saved $\approx 50\%$ of the computational time. We believe that sub-space search method can be widely used for efficient BN inference in systems biology.

摘要

轉錄調控在細胞分化、形態發生及生物體對內外刺激的響應上起重要的作用。生物微陣列芯 片技術能夠檢測大量基因在不同實驗條件下的表達情況。其產生的海量數據使得生物學家可以對基因的調控網絡進行系統的研究。而基因調控網絡的重構作為一種工具能夠從大量的基因表達數據中挖掘出最有意義的生物信息。

本論文的主要目的是研究和發展基因調控網絡的方法。

在本論文的第一部分，我通過將基因表達譜擬合進一個混合高斯模型並比較不同模型的似然度，從而發展了一個用於評估基因表達譜的相似性的新方法(DBoMM)。當我們用擬合的基因表達譜數據比較DBoMM和常用的方法皮爾遜相似度，歐幾里得距離及互信息的表現時，DBoMM的好於其它三種方法。對於真實的大腸桿菌數據集，DBoMM有不差於互信息的表現，且明顯強於另外兩種方法。 DBoMM對基因芯片數據中的噪聲具有魯棒性並且能夠檢測基因調控發生的條件。

之後，我將這個用於推導基因調控網絡的混合模型擴展為一個數量化模型。通過計算相關基因表達譜的條件分佈概率，我們能夠預測在不同實驗條件下的基因表達譜，比如基因敲除，基因過表達或調控網絡重構等。而且，通過將新的實驗條件與已知的條件相關聯，這個模型還能夠揭示不同實驗條件的功能相關性。

在本論文的第二部分，我提出了一個子空間貪婪搜索算法以提高貝葉斯網絡推導的效率。貝葉斯網絡已經被成功的用於從基因芯片數據推導基因之間的相關性。然而，最大的限制在於貝葉斯網絡的計算量隨著基因的增多而成指數增長。而我所提出的方法通過選擇具有高偏相似係數的基因對組成搜索空間，減少了搜索量。相比標準方法，新方法在獲得同樣預測精度的情況下可以節省約50%的計算時間。

Acknowledgements

First, I would like to deeply thank my two graduate supervisors, Prof. SUN Sai-ming Samuel and Prof. GUO Dianjing Diane, for providing me with the opportunity to conduct the bioinformatics research, and especially for their support and guidance on my research project throughout my doctoral study. The understanding and freedom they gave me made my graduate school time enjoyable. My life and career benefit much from their sharing of wisdom, knowledge and happiness.

I would like to express my sincere thanks to the members of my thesis committee, Prof. LAM Hon Ming, Prof. NGAI Sai Ming and Prof. CHAN Ting Fung for their guidance and comments throughout my doctoral study. Meanwhile, I sincerely thank my external thesis examiner, Prof. TANG Lei Han from Hong Kong Baptist University for his critical comments on my thesis.

I thank very especially, Prof. Fan Xiao Dan from the statistic department, for his support and advice on the statistical part of my project.

I also like to thank all of the former and current labmates, Chan Yiuman, Wang Wei, Qi Yan, Wu Wei, Yu penwen, Wang Yejun, Sun Mingan, Wu Ting, Cheng Hai, Wang

Jingxue and Ma Dongming, in G94 for their support and understanding in daily life.

Finally, I would like to express my deepest gratitude to my wife, Wei Ping, for her lasting love, patience and support to me. Thanks also go to my parents for their understanding, support and encouragement in my life.

Contents

List of Figures

List of Tables

Part I

Model-based Reconstruction of Gene Regulatory Network

Chapter 1

Introduction

1.1 Computational methods for gene regulatory network inference

The genome of a living organism often contains a large number of protein coding genes. The amounts of gene products and their temporal/spatial expression pattern are crucial to maintain the normal cellular functions and the survival of the living organism. The expression of genes can be regulated at various stages, including chromatin domains, transcription, post-transcriptional modification, translation and mRNA degradation etc. Among these, the transcriptional regulation is the major regulatory machinery for most eukaryotes and prokaryotes.

At the transcription level, the expression of a gene is directly controlled by the transcription factors. A living organism responds to the internal or external cues by tuning the expression of certain genes. For example, arabidopsis plant can respond to various abiotic and biotic stress by turning on/off the expression of many genes [1]. Yeast cell in a sugar solution can turn on enzyme coding genes necessary to process the sugar to alcohol [2]. These genes and their regulatory proteins form a gene regulatory network(GRN). For multicellular organisms, by regulating the expression of genes in different cells, GRN help to shape the body of the organism [3].

Several notable examples have set the stage for adopting GRN models in daily laboratory practice. The unprecedented link between protein mistranslation and the reaction to reactive oxygen species in response to antibiotics treatment was unveiled by combining network inference with experimental evidence in *E.coli* [4]. Similar approaches were used to unravel the complex network regulating host pathogen interactions in *Salmonellaenterica subsp. enterica serovar Typhimurium* [5] and to chart the transcriptional network of the *archeon Halobacterium salinarum* for the first time [6]. Computationally inferred interactions therefore offer a useful resource for putting experimental findings into a more global context, by finding novel interactions and by unfolding links between the pathway under investigation and other cellular processes [7].

Figure 1.1 demonstrates the basic structure of a gene regulatory network. The signal (A/B) from cell/environment interact with the receptor proteins and change the function of these receptor proteins by altering their conformations. In general, the receptor proteins cause a cascade of interacting kinase proteins or other molecules to active/inactive the transcription factors, which then bind/unbind to the DNA sequence and influence the expression of genes. The control process of a gene regulatory network is illustrated in

Figure 1.2. Figure 1.3 shows the graph representation of a regulatory network, where the nodes denote proteins, their corresponding mRNAs, and protein/protein complexes, and the interactions between these molecules are represented by the edges. The arrow of the edge indicates the causal relations between two nodes and the direction of the information transmission.

To uncover the GRN, modern biological technologies have been developed to detect the expression of mRNA and to elucidate the transcriptional regulation of genes, including the classical qPCR [8–11] and EST method [12], and the high throughput microarray [13–19], CHIP-chip [20–22], CHIP-seq methods [23–25]. These technologies have produced a large number of data which enable systematic studies of gene regulatory networks and reveal the mechanisms that underlie cellular processes. The scientists are now confronted with the problem as to how to re-construct the GRN based on these massive data. In recent years, computational models have been developed to infer GRNs, and the most common modeling technique involves coupled ordinary differential equations (ODEs), Boolean(Continuous) networks, Stochastic, Clusters(introduction in next section) and Bayesian Network(introduction in next part).

1.1.1 Coupled ODEs

The algorithms based on ordinary differential equations (ODEs) can relate changes in gene transcript concentration to each other and to an external perturbation [26]. Suppose a GRN has N nodes, and we use $x_{i(t)}$ to represent the concentrations of the ith node at t time. Each ODE describes one node regulation as a function of other nodes:

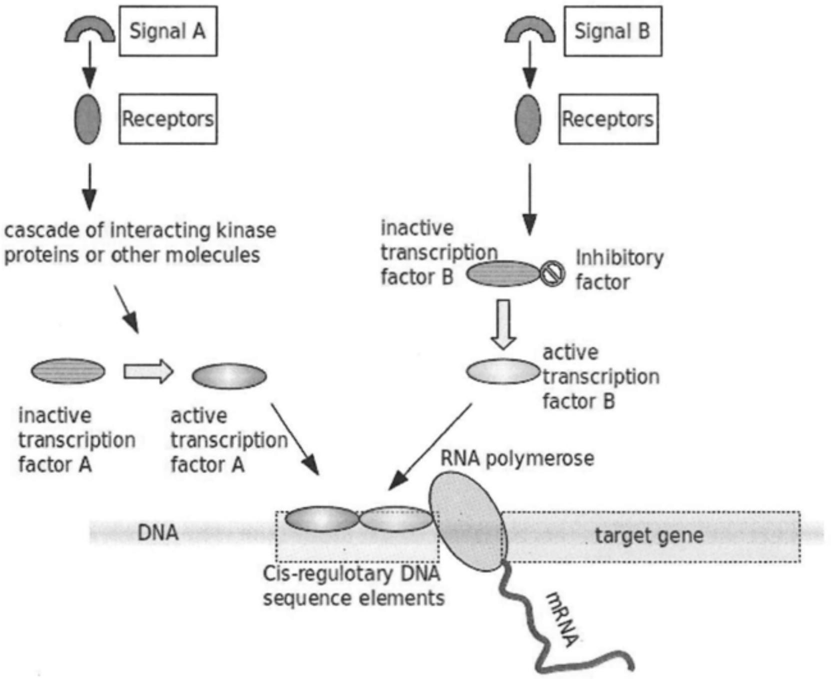

Signal A
Receptors
cascade of interacting kinase proteins or other molecules
inactive transcription factor A
active transcription factor A
Signal B
Receptors
inactive transcription factor B
Inhibitory factor
active transcription factor B
RNA polymerose
DNA
Cis-regulotary DNA sequence elements
target gene
mRNA

Figure 1.1: **An example of GRN.** This is an example that how the gene regulatory network(GRN) influence the expression profiles of genes. The signal (A/B) from cell/environment interact with the receptor proteins and change the function of these receptor proteins and other proteins. These proteins active/inactive the transcription factors, which then bind/unbind to the DNA sequence and influence the expression of genes.

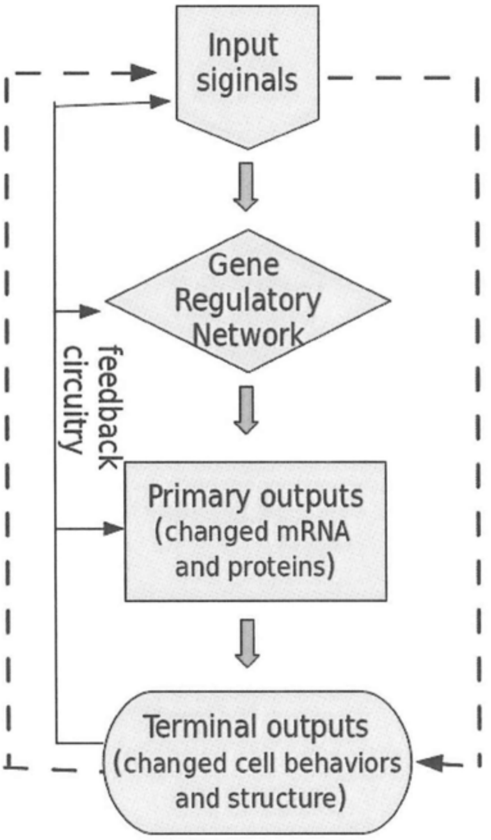


Figure 1.2: **The control process of a GRN.** The signals from cell/environment can change the expression of genes and further change the behaviors and structures of cells by influencing the gene regulatory network(GRN).

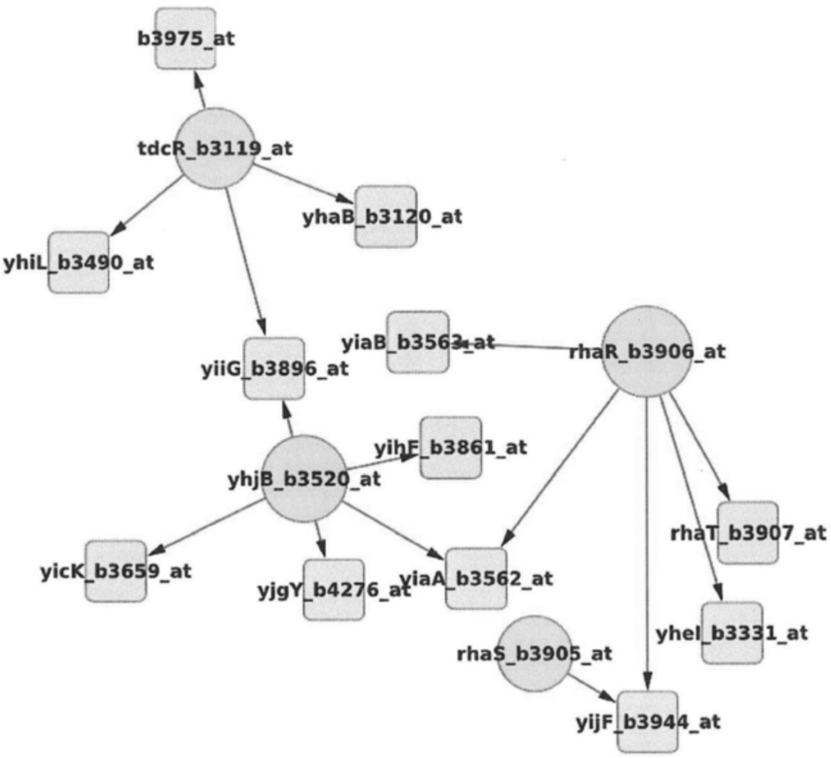


Figure 1.3: **The graph representation of a network.** The pink circle corresponds to the transcription factor and the blue square corresponds to the target gene. The direction of the arrow represents the causal relations.

$$\frac{x_{i(t)}}{dt} = f_i\left(x_1, x_2, \ldots, x_n, \theta_i\right) \tag{1.1}$$

where $\frac{x_{i(t)}}{dt}$ represents the transcription rate of gene i at time t, the function f_i expresses the dependence of x_i on the concentrations of other nodes in the GRN, θ_i is a set of parameters describing interactions among genes. Different from Bayesian Network, ODEs are deterministic method, and the interactions among genes represent causal interactions, and not statistical dependencies. To infer a gene reguatory network using ODEs, we should choose a functional form j_i and then to estimate the unknown parameters θ_i for each i from the gene expression data D using some optimisation technique [26].

1.1.2 Boolean network

Boolean networks for modeling the gene regulatory networks are first used by Stuart Kauffman [27]. In a GRN modeled by Boolean network, the node in any one of two states: on or off, represents the gene, input or output. For a gene, "on" corresponds to the gene being expressed; for inputs and outputs, "on" corresponds to the substance being present. The edge with arrow in the network from one node to another corresponds to the causal link between the two nodes. The state of a node is the Boolean function of the states of all the parent nodes.

Continuous network models of GRNs are an extension of the Boolean networks. Nodes and edges represent the same biological events in Boolean networks. However, the states of node display a continuous range of activity levels which can capture several properties of gene regulatory networks not present in the Boolean model [28].

1.1.3 Stochastic gene networks

A stochastic process is one whose behavior is non-deterministic, in that a system's subsequent state is determined both by the process's predictable actions and by a random element. Because that gene expression can be thought as a stochastic process [29, 30], scientists used stochastic formalism to model gene regulatory network [31–34]. The first versions of stochastic models of gene expression involved only instantaneous reactions and were driven by the Gillespie algorithm [35].

Specifically, gene transcription is modelled as single step multiple delayed reactions in order to account for the time it takes for the entire process to be complete [36]. A set of reactions were proposed [37] that allow generating GRNs.

For example, basic transcription of a gene can be represented by the following single-step reaction (RNAP is the RNA polymerase, RBS is the RNA ribosome binding site, and Pro_i is the promoter region of gene i) [38]:

$$\text{RNAP} + \text{Pro}_i \overset{k_{i,bas}}{\rightarrow} \text{Pro}_i(\tau_i^1) + \text{RBS}_i(\tau_i^1) + \text{RNAP}(\tau_i^2) \tag{1.2}$$

Except for the classes described above, based on different rules, the algorithms inferring GRN can be classified into different categories. The applications, advantages and limitations of these network inference methods have been summarized in several excellent reviews [7, 26, 39].

From pure data set standpoint, Bansal et.al compared several general reverse algorithms including classic clustering algorithm, Bayesian networks, information-theoretic approaches and ordinary differential equations and showed that reverse-engineering algorithms are indeed able to correctly infer regulatory interactions among genes, at least when one performs perturbation experiments complying with the algorithm requirements [26].

Karlebach et.al [39] divided various computational models into three classes, logical models, continuous models and single-molecule level models. The advantages and limitations of these models have been discussed. Some open questions regarding the regulatory networks, including how structure, dynamics and functionality relate to each other, how organisms use regulatory networks to adapt to their environments, and the interplay between regulatory networks and other cellular processes, such as metabolism were raised.

Focused mainly on top-down network inference methods, Smet et.al classified these methods into different categories combining criteria that relates directly to the biological interpretation of outcome and reviewed the strategies of them [7].

1.2 Methods for distance measurement

Clustering is a unsupervised learning method, and a common technique for statistical data analysis used in many fields, including machine learning, data mining, pattern recognition, image analysis, information retrieval, and bioinformatics. It has been extensively used in microarray data analysis to group genes with similar expression patterns [40–44]. The underlying assumption is that co-expressed genes may share common functional tasks and regulatory mechanisms. Similar expression patterns may also provide useful insights into various transcriptional and biological processes [44–46].

The most general cluster method is hierarchical algorithms [47, 48], which find successive clusters using previously established clusters.

Different from hierarchical method, partitional algorithms typically determine all clusters at once. It include k-means clustering [49] and fuzzy clustering [50].

DBSCAN [51] and OPTICS [52] are two typical density-based clustering algorithms. This algorithm can discover arbitrary-shaped clusters that are some regions in which the density of observed data exceed a specific threshold.

Biclustering methods [53] are devised to find the function module. This algorithm can not only cluster the genes but also the conditions under which the genes show similar pattern [54].

Most clustering algorithms depend heavily on 'similarity' or 'distance' measures that quantify the degree of association between expression profiles [55]. The choice of distance measure for a successful identification of gene relations and regulatory networks is probably more important than the choice of machine learning algorithm [46, 56].

Two major class of methods are commonly used to measure the gene distance [56]. In the first method, the expression profiles of two genes are viewed as two vectors in some space and the distances are computed in a pairwise fashion. For example Pearson correlation(COR), Euclidean distance(EUC), Manhattan metric(MAN),Cosine correlation(EISEN),Spearman correlation(SPEAR), Kendall's tau correlation(TAU) [57], etc. all adopt this method. The second method ignores the natural pairing of observations, the gene expression profiles are instead assumed to be from different probability density functions. The distance of two genes is represented by calculating the distances between two distributions. Both Kullback-Leibler information (KLI) [58, 59] and Mutual information(MI) [60] belong to this class.

COR and EUC have been widely implemented to measure the similarity of gene expression profiles [61–68], because of their simple formulas and successful application in conventional data-extensive research fields, such as signal or image processing. However, these two methods bear obvious limitations. For example, COR is based on the assumption that the expression of genes are linearly related. Both COR and EUC are sensitive to noise effects and outliers of the expression profiles [55] and require complete gene expression profiles as input. This limits their widely the application due to the often missing values in microarray data.

In contrast, mutual information (MI),a well known method in information theory [60], measures the dependencies of the distributions, which are assumed to produce the gene expression profiles. It is robust to noise, outliers and missing data and can detect any kind of dependence between distributions in theory [69, 70]. MI has been widely used to analyze gene expression data [46, 66, 70–72]. However, the measure of MI requires the discretization of the continuous expression values. Most discretization methods use histogram based procedure [70, 71, 73], which is arbitrary. And these arbitrary bins also can not supply any information about the relationships between different experimental conditions.

In this thesis, I propose a method to measure the similarity of gene expression profiles by calculating the dependence of distributions to overcome the limitations of mutual information method. Specially, the difference of Bayesian Information Criterions(BIC) between joint and marginal distribution models of two genes is used to describe the similarity of these two genes. The joint and marginal distributions are assumed to follow a bivariate and two univariate mixture Gaussian distributions respectively. We named this method DBoMM(Difference BIC of Mixture Model). Because DBoMM calculates the dependence of distributions, it is not sensitive to noisy, outliers and missing data. In addition, it does not requires the linear assumption. For each gene pair, the expression patterns in the samples(experimental conditions) belonging to the same distribution are similar. It reflects the condition-dependent relationships between genes [74, 75]. The inferred statistical parameters from gene expression profile can also be used to predict the dynamics of functionally related gene. Using synthetic dataset, we show that DBoMM out-performed PCC, EUC, and MI method. The performance is better than PCC and EUC, whereas comparable to MI when using the *E.coli* microarray dataset.

Although the regulatory networks inferred by these methods provide important clues about the gene function in most cases, quantitative models that accurately predict the dynamic behavior of genes under system perturbations are required by synthetic biology, which aims to re-design biological systems with desired function by rewiring the genetic network.

1.3 Quantitative model for synthetic biology redesign

Synthetic biology is a new area of biological research that combines science and engineering. Synthetic biology encompasses a variety of different approaches, methodologies and disciplines, and many different definitions exist. What they all have in common, however, is that they see synthetic biology as the design and construction of new biological functions and systems not found in nature.

Currently, synthetic biology focuses on altering the general process flow-specifically modifications to the function and behavior of the process units (transcription [76] and translation) and the associated process streams (DNA, RNA [77], and protein). Numerous synthetic gene circuits have been created in the past decade, including bistable switches, oscillators, and logic gates [78–82], and possible applications abound, ranging from biofuels, to detectors for biochemical and chemical weapons, to disease diagnosis, to gene therapies.

Technologies and algorithms introduced in the first section have produced various interactions between genes and gene regulatory networks, which in theory can direct the design of synthetic biology. However, as an engineering discipline, synthetic biology cannot rely on endless trial and error methods driven by verbal description of biomolecular interaction networks. The challenge facing synthetic biologist is then to reduce the enormous volume and complexity of biological data into concise theoretical formulations with predictive ability, ultimately associating synthetic DNA sequences to dynamic phenotypes [83].

To redesign the transcriptional regulation network, we need a quantitative model able to predict the gene dynamics. In this part, I developed a computational model for quantitative prediction of gene expression profiles based on Gaussian mixture models. In this model, a regulatory network inferred from various reverse engineering methods or experiments is first decomposed into different modules; each consisting of related genes (participate in the same pathway or regulatory gene pairs, etc.). In each module, the gene expression profiles are trained to fit a mixture multivariate Gaussian distribution and the estimated parameters are used to represent the gene expression levels and the gene relations. By calculating the conditional distribution of multivariate normal distribution, the model infers the expression values of gene given that of other related genes. In addition, by comparing the expression profiles of genes under a new condition vs. known conditions, the model assigns the new condition (treatment, mutant or redesign) into a group of known conditions, allowing the researchers to estimate the functional relationships among these conditions. We demonstrate that the proposed mixture model out-performed other multiple linear regression(MLR) based method developed by Carrera et.al [84]. This model can also be easily extended as a benchmark synthetic dataset generator for evaluation of network inference algorithm because the complex transcription process is represented by certain estimated statistical parameters.

Using this model, the *E.coli* transcriptome profiles under knockout and over-expression of master regulatory genes, as well as network rewiring, were accurately estimated. This model may serve as a useful tool to guide both experimental design and genome-wide redesign of transcription regulation in synthetic biology.

The first part of my thesis research has two objectives: 1. to develop a new gene similarity measure method based on a mixture Gaussian model to overcome the limitations of mutual information method; 2. to extend this method into a quantitative model for synthetic biology transcription regulation redesign. The mixture Gaussian model is used to solve the two problems because of two reasons: 1. mixture Gaussian model has been successfully applied in many fields; 2. it is more flexible to describe the complex transcription regulatory relations between genes.

□ **End of chapter.**

Chapter 2

Methodology

2.1 Data sets

The gene expression dataset consists of a compendium of 445 *E.coli* Affymetrix Antisense2 microarray data monitoring the expression profiles (http://m3d.bu.edu/) of 4345 genes [85]. The samples were collected under different experimental conditions, e.g. PH changes, growth phases, antibiotics, heat shock, different media, varying oxygen concentrations and numerous genetic perturbations. The data was normalized using RMA method [86] in bioconductor package.

The gene regulation data is extracted from RegulonDB version 7 [87]. Of all the interactions, we removed these genes that do not match the probe sets and self-regulation interactions, leaving a reference network consisting of 1531 nonredundant genes and 3774 experimentally confirmed regulatory interactions. Based on the topological structure of this network, 1531 genes and 3774 interactions were classified into 1156 modules.

SynTReN [88] is used to generate a simulated data sets with varying number of conditions for a synthetic transcription regulatory network with 1000 genes.

Dataset for noisy estimating:

SynTReN is used to generate 5 simulated data sets with 100 conditions and 500 genes. The 5 simulated data sets include different portions of biological and experimental noise, 0%, 20%,40%,60% and 80%.

2.2 Softwares

The R [89] codes for the inferring process are available in Appendix A.3. The R package *mclust* [90, 91] was used to train the expression profiles of genes to fit a mixture Gaussian distribution. The TCA pathway figure was drawn using *Cytoscape* [92] with plugin *KGMLreader*.

2.3 Model selection

To decide the number of components in the Gaussian mixture model, Bayesian Information Criterion (BIC) [93] is used to find a proper compromise between the likelihood and the number of parameters of the model. More specifically, it is defined as

$$BIC = -2lnL + kln(n)$$

where $n =$ the number of data points;

$k =$the number of free parameters to be estimated;

L =the maximized value of the likelihood function of the model.

2.4 Multivariate Gaussian mixture model

We use multivariate Gaussian mixture model to describe the gene relations. The joint probability of gene expression value is

$$p(\mathrm{g}_1, \cdots, \mathrm{g}_D) = \sum_{k=1}^{K} \pi_k N(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

where $\{\mathrm{g}_1, \cdots, \mathrm{g}_D\}$ is a set of genes, π_k is the weight of the kth component and $\sum_{k=1}^{K} \pi_k = 1$, $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ are the mean vector and the covariance matrix of the kth component respectively .

2.5 Parameter estimation

Suppose we have a set of genes $\{\mathrm{g}_1, \cdots, \mathrm{g}_D\}$ and there are N expression data points $\{\boldsymbol{e}_1, \ldots, \mathbf{e}_N\}$, this data set can be represented as an $N \times D$ matrix $\boldsymbol{E}$. We assume the N data points are independent from the same multivariate Gaussian mixture distribution. The log-likelihood of the observed data is :

$$\ln p(\boldsymbol{E}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k N(\boldsymbol{e}_n | \boldsymbol{u}_k, \boldsymbol{\Sigma}_k) \right\}$$

The training process is to find the maximum likelihood estimate of $(\pi, \mu, \sum)$. An elegant and powerful method for handling this task is the $Expectation - Maximization$ (EM) algorithms [94].

E step: Evaluate the responsibilities using the current parameter values

$$\gamma(z_{nk}) = \frac{\pi_k N(\boldsymbol{e}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j N(\mathbf{e}_n|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}$$

$\gamma(z_{nk})$is the posterior probability that component k is responsible for generating $\boldsymbol{e}_n$.

M step: Re-estimate the parameters using the current responsibilities

$$\boldsymbol{\mu}_k^{\text{new}} = \frac{1}{N_k}\sum_{n=1}^{N}\gamma(z_{nk})\boldsymbol{e}_n$$

$$\boldsymbol{\Sigma}_k^{\text{new}} = \frac{1}{N_k}\sum_{n=1}^{N}\gamma(z_{nk})(\boldsymbol{e}_n - \boldsymbol{\mu}_k^{\text{new}})(\boldsymbol{e}_n - \boldsymbol{\mu}_k^{\text{new}})^{\text{T}}$$

$$\pi_k^{\text{new}} = \frac{N_k}{N}$$

where

$$N_k = \sum_{n=1}^{N}\gamma(z_{nk})$$

2.6 Dependence BIC of mixture model(DBoMM)

The joint distribution of expression profile of two genes are assumed to follow a bivariate mixture Gaussian distribution. Therefore the marginal distribution of one gene's expression profile follows a univariate mixture Gaussian distribution. The mixture distribution can be described as:

$$p(\mathbf{x}) = \sum_{k=1}^{K}\pi_k N(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

where π_k is the weight of the kth component, $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ denote the mean vector and the covariance matrix of the kth component respectively.

Bayesian information criterion(BIC) [93] is used to estimate the number of distribution automatically. Expectation-maximization algorithms(EM) [94] is used to find the maximum BIC.

Then the similarity of two genes' expression profiles can be written as:

$DBoMM(X,Y) = BIC(M_{\mathbf{xy}}) - BIC(M_{\mathbf{x}}) - BIC(M_{\mathbf{y}})$

where $M_{\mathbf{xy}}$ is the joint distribution model with minimal BIC of genes x and y, $M_{\mathbf{x}}$ and $M_{\mathbf{y}}$ respective are marginal distribution models with minimal BIC of gene x and gene y respectively.

2.7 A model-based clustering method for gene similarity measurement

2.7.1 Similarity measurements

The Euclidean distance, Pearson correlation, and mutual information (MI) are commonly used measures in gene expression analysis. These measures quantify a pairwise distance between expression profiles over n conditions that are represented by the two vectors $\mathbf{x} =(x_1, \ldots, x_n)$, and $\mathbf{y} =(y_1, \ldots, y_n)$.

2.7.2 Euclidean Distance and Pearson Correlation

The Euclidean distance between two expression profiles is given by

$$E(\mathbf{x},\mathbf{y}) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

The Pearson correlation coefficient between two expression patterns is defined as

$$R(\mathbf{x},\mathbf{y}) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})\sum_{i=1}^{n}(y_i - \bar{y})}}$$

where $\bar{x}$, $\bar{y}$ denote the average patterns level.

We used commands *euc*() and *cor.dis*() in package *bioDist* under *R* platform [?, 95] to calculate the Euclidean distance and Pearson correlation coefficient.

2.7.3 Mutual Information

Given two random variables X, Y with respective ranges $x_i \in A_x, y_j \in A_j$ and probability distributions functions $P(X = x_i) \equiv p_i, P(Y = y_j) \equiv p_j$, the Mutual information between two expression patterns, represented by random variables X and Y, is given by

$$I(X;Y) = \sum_{i}\sum_{j} p_{ij} log \frac{p_{ij}}{p_i p_j}$$

The gene expression profiles were divided into different bins and then the mutual information is computed. The data was treated as if they are discrete. We used $mutualInfo()$ in package *bioDist* [96] and the default number of bins(10) to calculate the mutual information of two genes.

2.7.4 Measure the performance of different methods

To evaluate the performance, we computed the precision and recall of the inferred networks by comparing the inferred network to the reference network. Precision is the fraction of predicted interactions that are correct [TP/(TP + FP)], and recall is the fraction of all known interactions that are discovered by the algorithm [TP/(TP + FN)], where TP is the number of true positives, FP is the number of false positives, and FN is the number of false negatives. Precision and recall were computed over a range of pruning thresholds; interactions with scores below the pruning threshold were removed from the inferred network. For *E.coli* dataset, we constrained the resulting network maps to include only the genes available in our RegulonDB control set.

2.8 Redesign of transcription regulation using a mixture model

2.8.1 Dataset separation Process

In a total of 445 microarray samples, 45 (10%) were randomly selected as the test dataset and the remaining 400 samples were used for model training. We repeated the dataset separation process (bootstrap) 9 times to ensure the results were not dependent on the test set.

2.8.2 Decomposing the pathway or regulation network into different modules

The pathway and regulatory network can be represented by an undirected graph and a directed acyclic graph (DAG) respectively, in which the nodes and edges represent genes and the relationship between genes respectively. As proposed by Segal et.al [54], a module is a set of genes co-regulated to respond to different conditions. Here we used module to represent a set of functionally related genes. Specifically, for the pathway network, the genes involved in the same pathway belong to the same module; and for the regulatory network, the target gene and all its regulator genes belong to the same module. In each module, we assume the gene expression values follow a mixture of multivariate Gaussian distribution.

2.8.3 Comparison between predictive and experimental expression value

In theory, the model predicts the distribution of gene expression values, which is not readily comparable to the experimental data. To get a concrete value for comparison purpose, the expected value (mean) of the mixture distribution was used as the predictive value. Relative error(RE) was used to validate the performance of the model.

$$\mathrm{RE} = |(\mathrm{e}_p - \mathrm{e}_e)/\mathrm{e}_e|$$

where e_p and e_e correspond to the predictive and experimental expression value respectively.

2.8.4 Prediction of transcriptome profiles

Updating the parameters

In one module, we suppose the expression values of some genes are known and we want to predict the expression values of other genes. Actually, this process is to infer the conditional distribution of the unknown genes given the values of the known genes. The expression values of two sets of genes were represented as $\mathbf{e}^{known}$ and $\mathbf{e}^{unknown}$. So we can write the equation:

$$p(\mathbf{e}^{unknown}|\mathbf{e}^{known}) = \frac{p(\mathbf{e}^{unknown}, \mathbf{e}^{known})}{p(\mathbf{e}^{known})}$$

$p(\mathbf{e}^{unknown}, \mathbf{e}^{known})$ can also be written like this:

$$p(\mathbf{e}^{unknown}, \mathbf{e}^{known}) =$$

$$\sum_{k=1}^{K} \pi_k N(\mathbf{e}_k^{unknown}|b_k, A_k) N(\mathbf{e}_k^{known}|\boldsymbol{\mu}_k^{known}, \Sigma_k^{known})$$

where

$$b_k = \boldsymbol{\mu}_k^{unknown} + \left[\Sigma_k^{(known,unknown)}\right]^T \left[\Sigma_k^{known}\right]^{-1} (\mathbf{e}^{known} - \boldsymbol{\mu}_k^{known})$$

$$A_k = \Sigma_k^{unknown} - \left[\Sigma_k^{(known,unknown)}\right]^T \left[\Sigma_k^{known}\right]^{-1} \Sigma_k^{(known,unknown)}$$

here b_k and A_k are the new mean and variance of unknown genes given the values of known genes in the kth distribution. In section 2.8.5, we give the detail inference process how to get b_k and A_k.

so the expression of $p(\mathbf{e}^{unknown}|\mathbf{e}^{known})$ becomes

$$p(\mathbf{e}^{unknown}|\mathbf{e}^{known}) =$$

$$\frac{\sum_{k=1}^{K} \pi_k N(\mathbf{e}_k^{unknown}|b_k, A_k) N(\mathbf{e}_k^{known}|\boldsymbol{\mu}_k^{known}, \Sigma_k^{known})}{p(\mathbf{e}^{known})}$$

which is equivalent to

$$p(\mathsf{e}^{unknown}|\mathsf{e}^{known}) = \sum_{k=1}^{K} \pi_k^{\text{new}} N(\mathsf{e}^{unknown}|b_k, A_k)$$

with

$$\pi_k^{\text{new}} = \frac{\pi_k N(\mathsf{e}^{known}|\boldsymbol{\mu}_k^{known}, \Sigma_k^{known})}{p(\mathsf{e}^{known})}$$

then we can get a set of newly updated parameters b_k,A_k,and π_k^{new}, which describe the distribution of unknown genes' values given the values of known genes.

The π_k^{new} is a vector describing the probabilities that this new sample belongs to each component. And we assign the new sample into the component with maximal probability.

Expected values of unknown genes

After updating the parameters, a new mixture of Gaussians which describe the expression profiles of unknown genes are obtained. To get a concrete expression values of the unknown genes, we calculate the expected expression values of the unknown genes using this equation,

$$E(\mathsf{e}^{unknown}) = \sum \boldsymbol{\pi}^{\text{new}} \boldsymbol{b}$$

with $\boldsymbol{\pi}^{\text{new}}$=$\{\pi_1^{\text{new}}, \ldots, \pi_K^{\text{new}}\}$ and $\boldsymbol{b} = \{b_1, \cdots, b_k\}$.

2.8.5 Marginal and conditional distributions of multivariate normal distribution

Assume an n-dimensional random vector

$$\mathbf{x} = \begin{bmatrix} \boldsymbol{x}_1 \\ \mathbf{x}_2 \end{bmatrix}$$

has a normal distribution $N(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ with

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} \; \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} \; \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

where $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are two subvectors of respective dimensions p and q with $p + q = n$. Note that $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^T$, and $\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}_{21}^T$.

The joint density of $\mathbf{x}$ is:

$$f(\mathbf{x}) = f(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} exp[-\frac{1}{2}Q(\mathbf{x}_1, \mathbf{x}_2)]$$

where Q is defined as

$$Q(\boldsymbol{x}_1, \mathbf{x}_2) = (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$$

$$= [(\mathbf{x}_1 - \boldsymbol{\mu}_1)^T, (\mathbf{x}_2 - \boldsymbol{\mu}_2)^T] \begin{bmatrix} \boldsymbol{\Sigma}^{11} \; \boldsymbol{\Sigma}^{12} \\ \boldsymbol{\Sigma}^{21} \; \boldsymbol{\Sigma}^{22} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 - \boldsymbol{\mu}_1 \\ \mathbf{x}_2 - \boldsymbol{\mu}_2 \end{bmatrix}$$

$$= (\mathbf{x}_1 - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}^{11} (\mathbf{x}_1 - \boldsymbol{\mu}_1) + 2(\mathbf{x}_1 - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}^{12} (\mathbf{x}_2 - \boldsymbol{\mu}_2) + (\mathbf{x}_2 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{22} (\mathbf{x}_2 - \boldsymbol{\mu}_2)$$

here

$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} \; \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} \; \boldsymbol{\Sigma}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{\Sigma}^{11} \; \boldsymbol{\Sigma}^{12} \\ \boldsymbol{\Sigma}^{21} \; \boldsymbol{\Sigma}^{22} \end{bmatrix}$$

because

$$\mathbf{\Sigma}^{11} = (\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{12}^{T})^{-1} = \mathbf{\Sigma}_{11}^{-1} + \mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}(\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{12}^{T}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^{-1}\mathbf{\Sigma}_{12}^{T}\mathbf{\Sigma}_{11}^{-1}$$

$$\mathbf{\Sigma}^{22} = (\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{12}^{T}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^{-1} = \mathbf{\Sigma}_{22}^{-1} + \mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{12}^{T}(\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{12}^{T})^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}$$

$$\mathbf{\Sigma}^{12} = -\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}(\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{12}^{T}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^{-1} = (\mathbf{\Sigma}^{21})^{T}$$

Substituting $\mathbf{\Sigma}^{11}$, $\mathbf{\Sigma}^{12}$ and$\mathbf{\Sigma}^{22}$ into to $Q(\boldsymbol{x}_1, \mathbf{x}_2)$ get:

$$Q(\boldsymbol{x}_1, \mathbf{x}_2) = (\mathbf{x}_1 - \boldsymbol{\mu}_1)^T \left[\mathbf{\Sigma}_{11}^{-1} + \mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}(\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{12}^{T}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^{-1}\mathbf{\Sigma}_{12}^{T}\mathbf{\Sigma}_{11}^{-1}\right](\mathbf{x}_1 - \boldsymbol{\mu}_1)$$

$$-2(\mathbf{x}_1 - \boldsymbol{\mu}_1)^T \left[\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}(\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{12}^{T}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^{-1}\right](\mathbf{x}_2 - \boldsymbol{\mu}_2)$$

$$+(\mathbf{x}_2 - \boldsymbol{\mu}_2)^T \left[(\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{12}^{T}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^{-1}\right](\mathbf{x}_2 - \boldsymbol{\mu}_2)$$

$$= (\mathbf{x}_1 - \boldsymbol{\mu}_1)^T\mathbf{\Sigma}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1) + (\mathbf{x}_1 - \boldsymbol{\mu}_1)^T \left[\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}(\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{12}^{T}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^{-1}\mathbf{\Sigma}_{12}^{T}\mathbf{\Sigma}_{11}^{-1}\right](\mathbf{x}_1 - \boldsymbol{\mu}_1)$$

$$-2(\mathbf{x}_1 - \boldsymbol{\mu}_1)^T \left[\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}(\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{12}^{T}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^{-1}\right](\mathbf{x}_2 - \boldsymbol{\mu}_2)$$

$$+(\mathbf{x}_2-\boldsymbol{\mu}_2)^T\left[(\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{12}^T\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}\right](\mathbf{x}_2-\boldsymbol{\mu}_2)$$

$$=(\mathbf{x}_1-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{x}_1-\boldsymbol{\mu}_1)$$

$$+\left[(\mathbf{x}_2-\boldsymbol{\mu}_2)-\boldsymbol{\Sigma}_{12}^T\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{x}_1-\boldsymbol{\mu}_1)\right]^T(\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{12}^T\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}\left[(\mathbf{x}_2-\boldsymbol{\mu}_2)-\boldsymbol{\Sigma}_{12}^T\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{x}_1-\boldsymbol{\mu}_1)\right]$$

For any vectors u and v and a symmetric matrix $A=A^T$:

$$u^TAu-2u^TAv+v^TAv=u^TAu-u^TAv-u^TAv+v^TAv$$

$$=u^TA(u-v)-(u-v)^TAv=u^TA(u-v)-v^TA(u-v)$$

$$=(u-v)^TA(u-v)=(v-u)^TA(v-u)$$

We define

$$b\triangleq\boldsymbol{\mu}_2+\boldsymbol{\Sigma}_{12}^T\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{x}_1-\boldsymbol{\mu}_1)$$

$$A\triangleq\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{12}^T\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$$

and

$$Q_1(\boldsymbol{x}_1) \triangleq (\mathbf{x}_1 - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}_{11}^{-1} (\mathbf{x}_1 - \boldsymbol{\mu}_1)$$

$$Q_2(\boldsymbol{x}_1, \mathbf{x}_2) \triangleq \left[(\mathbf{x}_2 - \boldsymbol{\mu}_2) - \boldsymbol{\Sigma}_{12}^T \boldsymbol{\Sigma}_{11}^{-1} (\mathbf{x}_1 - \boldsymbol{\mu}_1)\right]^T (\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{12}^T \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12})^{-1} \left[(\mathbf{x}_2 - \boldsymbol{\mu}_2) - \boldsymbol{\Sigma}_{12}^T \boldsymbol{\Sigma}_{11}^{-1} (\mathbf{x}_1 - \boldsymbol{\mu}_1)\right]$$

$$= (\mathbf{x}_2 - b)^T A^{-1} (\mathbf{x}_2 - b)$$

and get

$$Q(\boldsymbol{x}_1, \mathbf{x}_2) = Q_1(\boldsymbol{x}_1) + Q_2(\boldsymbol{x}_1, \mathbf{x}_2)$$

Now the joint distribution can be written as:

$$f(\mathbf{x}) = f(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{(2\pi)^{n/2} |\boldsymbol{\Sigma}|^{1/2}} exp[-\frac{1}{2} Q(\mathbf{x}_1, \mathbf{x}_2)]$$

$$= \frac{1}{(2\pi)^{n/2} |\boldsymbol{\Sigma}_{11}|^{1/2} |\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{12}^T \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12}|^{1/2}} exp[-\frac{1}{2} Q(\mathbf{x}_1, \mathbf{x}_2)]$$

$$= \frac{1}{(2\pi)^{n/2} |\boldsymbol{\Sigma}_{11}|^{1/2}} exp[-\frac{1}{2} (\mathbf{x}_1 - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}_{11}^{-1} (\mathbf{x}_1 - \boldsymbol{\mu}_1)] \frac{1}{(2\pi)^{n/2} |A|^{1/2}} exp[-\frac{1}{2} (\mathbf{x}_2 - b)^T A^{-1} (\mathbf{x}_2 - b)]$$

$$= N(\mathbf{x}_1, \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}) N(\mathbf{x}_2, b, A)$$

The marginal distribution of $\mathbf{x}_1$ is

$$f_1(\mathbf{x}_1) = \int f(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_2 = \frac{1}{(2\pi)^{n/2}|\mathbf{\Sigma}_{11}|^{1/2}} exp[-\frac{1}{2}(\mathbf{x}_1 - \boldsymbol{\mu}_1)^T \mathbf{\Sigma}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1)]$$

and the conditional distribution of $\mathbf{x}_2$ given $\mathbf{x}_1$ is

$$f_{2|1}(\mathbf{x}_2|\mathbf{x}_1) = \frac{f(\mathbf{x}_1, \mathbf{x}_2)}{f(\mathbf{x}_1)} = \frac{1}{(2\pi)^{n/2}|A|^{1/2}} exp[-\frac{1}{2}(\mathbf{x}_2 - b)^T A^{-1}(\mathbf{x}_2 - b)]$$

with

$$b = \boldsymbol{\mu}_2 + \mathbf{\Sigma}_{12}^T \mathbf{\Sigma}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1)$$

$$A = \mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{12}^T \mathbf{\Sigma}_{11}^{-1} \mathbf{\Sigma}_{12}$$

2.8.6 Multiple linear regression

To describe the genetic regulations using a linear model, the mRNA dynamics from gene y_i is given by

$$\frac{\mathrm{d}}{\mathrm{dt}} y_i = a_i + \sum_{j \in TF} b_{ij} y_i + \sum_{j \in TF} \sum_{k \in TF} b_{ijk} y_i + \cdots + \sum_{j \in TF} \sum_{k \in TF} \cdots \sum_{l \in TF} b_{ijk \cdots l} y_i - \sigma_i y_i$$

where a_i is the basal synthesis rate, b_{ij} the transcription regulatory coefficient of TF j, $b_{ijk\cdots l}$ the cooperative transcription regulatory coefficient of TFs $j, k \cdots l$ acting on the gene i and σ_i is the degradation rate. In the steady state, the formula can be written:

$$y_i = \alpha_i + \sum_{j \in TF} \beta_{ij} y_i + \sum_{j \in TF} \sum_{k \in TF} \beta_{ijk} y_i + \cdots + \sum_{j \in TF} \sum_{k \in TF} \cdots \sum_{l \in TF} \beta_{ijk \cdots l} y_i$$

where the parameters can be defined: $\alpha_i = a_i/\delta_i$, $\beta_{ij} = b_{ij}/\delta_i$ and $\beta_{ijk \cdots l} = b_{ijk \cdots l}/\delta_i$.

To estimate the model parameters $\alpha_i, \beta_{ij} \cdots \beta_{ijk \cdots l}$, multiple linear regression [97] is used, which is the result of a minimization problem (least squares) defined by

$$(\hat{\alpha}_i, \hat{\beta}_{ij}, \cdots \hat{\beta}_{ijk \cdots l}) = \arg\min \left\{ y_i - \alpha_{\mathrm{i}} - \sum_{\mathrm{j \in TF}} \beta_{\mathrm{ij}} y_{\mathrm{i}} - \sum_{\mathrm{j \in TF}} \sum_{\mathrm{k \in TF}} \beta_{\mathrm{ijk}} y_{\mathrm{i}} - \cdots - \sum_{\mathrm{j \in TF}} \sum_{\mathrm{k \in TF}} \cdots \sum_{\mathrm{l \in TF}} \beta_{\mathrm{ijk \cdots l}} y_{\mathrm{i}} \right\}$$

2.8.7 The schematic representation of the workflow

A schematic representation of the workflow can be found in Figure 2.1.

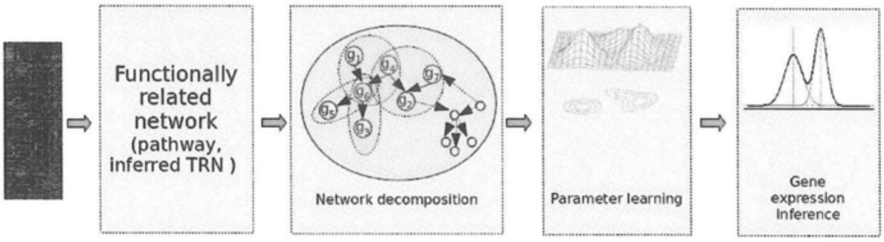

Functionally related network (pathway, inferred TRN)
g1
g4
g7
g6
g5
g2
g3
Network decomposition
Parameter learning
Gene expression Inference

Figure 2.1: **A schematic representation of the prediction workflow.**

□ **End of chapter.**

Chapter 3

A model-based clustering method for gene similarity measurement

3.1 DBoMM can distinguish the real interactions from the false ones

We applied DBoMM to both synthetic gene expression data and the real *E.coli* experimental data. To examine the ability of DBoMM in distinguishing the real and the false relations, we compared the distributions of DBoMM from the real gene interactions and the background ones respectively. We used the regulatory interaction network from RegulonDB database [87] and the pathway network from KEGG [98–100] as reference networks. In the RegulonDB regulatory network, the interactions between all transcription factors (TFs) and all target genes (excluding those real interactions), are defined as the background gene pairs. For the pathway network, the real interactions are those between genes in the same pathway, and all the interactions between genes belong to different pathways are used as the background interactions. Figure 3.1 shows the distributions of DBoMM from different interaction types. As seen, the real gene interactions obviously shifts away from the background interactions. And the difference is statistically significant(p-value of t-test between the two distributions is <2.2e10-16), indicating that DBoMM can distinguish the real and the false gene interactions based on gene expression data. A comparison of distributions using other methods (COR, EUC and MI) can be found in Table 3.1 and 3.2. The result shows MI has similar performance with DBoMM, whereas COR and EUC can not distinguish the real and background interactions.

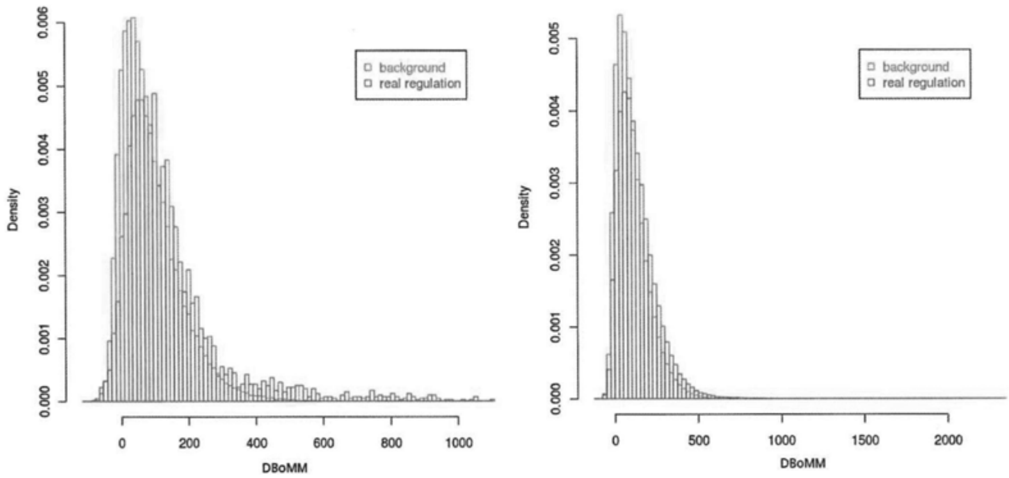


Figure 3.1: **The distributions of DBoMM from real and background interactions.** (a). the real interactions from RegulonDB; b. the real interactions from pathway genes.

Table 3.1: **The distribution of different distance scores based on RegulonDB.**

	DBoMM		MI		COR		EUC	
Regulation relations	Background	Real relations	Background	Real relations	Background	Real relations	Background	Real relations
Mean	85.71	147.64	0.19	0.25	0.76	0.69	34.47	35.55
Standard	86.86	160.67	0.69	0.14	0.17	0.22	23.94	21.14
P.value	<2.2e-016		<2.2e-016		1.0e+000		1.0e+00	

Table 3.2: **The distribution of different distance scores based on KEGG.**

	DBoMM		MI		COR		EUC	
Pathway relations	Background	Real relations	Background	Real relations	Background	Real relations	Background	Real relations
Mean	102.66	136.57	0.25	0.26	0.74	0.71	39.17	33.88
Standard	97.53	123.79	0.12	0.15	0.18	0.20	25.49	18.00
P.value	<2.2e-016		<2.2e-016		1.0e+000		1.0e+00	

3.2 A comparison with EUC, MI, and COR

The PR-curve (Precision-Recall) based on the regulatory interactions from RegulonDB is plotted(Figure 3.2). For *E.coli* dataset (Figure 3.2a), in general, the performance of DBoMM is comparable to that of MI, and both methods performance better than EUC and COR. However, EUC has better performance than DBoMM and MI when the recall is very low(0.04). For the synthetic dataset, DBoMM shows better performance than EUC and MI (Figure 3.2b). COR performed poorly using either *E.coli* or synthetic dataset. Because the regulation of gene transcription is affected by multiple factors, such as binding site affinity [101, 102], stability of the initiation complex [103, 104], cooperation among TFs, the quantity of TFs, and the time lag, etc, the direct regulatory relations between the regulators and target genes are noisy and not linear in most cases. This may explain the poor performance of similarity based COR measurement. In general, DBoMM has better performance than the other tested methods.

3.3 Motif analysis

We used DBoMM to infer a regulatory network with 60% precision using the *E.coli* dataset (Figure 3.2a). This predicted regulation network (Figure 3.3) consists of 468 genes and 741 regulatory interactions, among which 65 are included in RegulonDB. All the predictive regulatory interactions can be found in Appendix A.1. We also extracted a regulatory network consisting 407 genes and 618 regulatory interactions with 60% precision based on MI. Of the 618 regulatory interactions, 66 can be found in RegulonDB.

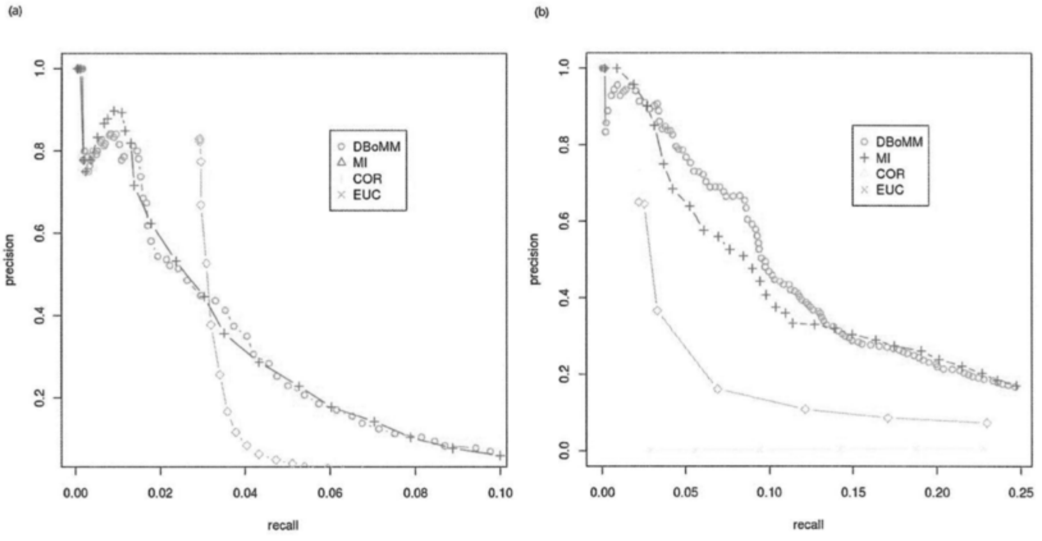


Figure 3.2: **The PR-curve based on reference network.** (a). *E.coli* dataset and the reference network from RegulonDB; (b). synthetic dataset.

424 regulatory interactions were found to be commonly shared by the two predicted networks, accounting for 57% and 68% of total interactions respectively. Interactions were only extracted from 328 known or predicted transcription factors to any of the 4,345 genes, enabling clear biological interpretation, assignment of direction (from transcription factors to non–transcription factor genes), and validation of the predictions. Interactions were also identified between transcription factors, although direction was not assigned.

Sequence analysis was conducted to detect the possible motif bound by each regulator. Not all transcription factors have enough targets to allow reliable motif detection, but for those that do, the motif provides a specific location for the regulatory interaction. All the transcription factors predicted to regulate 5 or more operons with at least

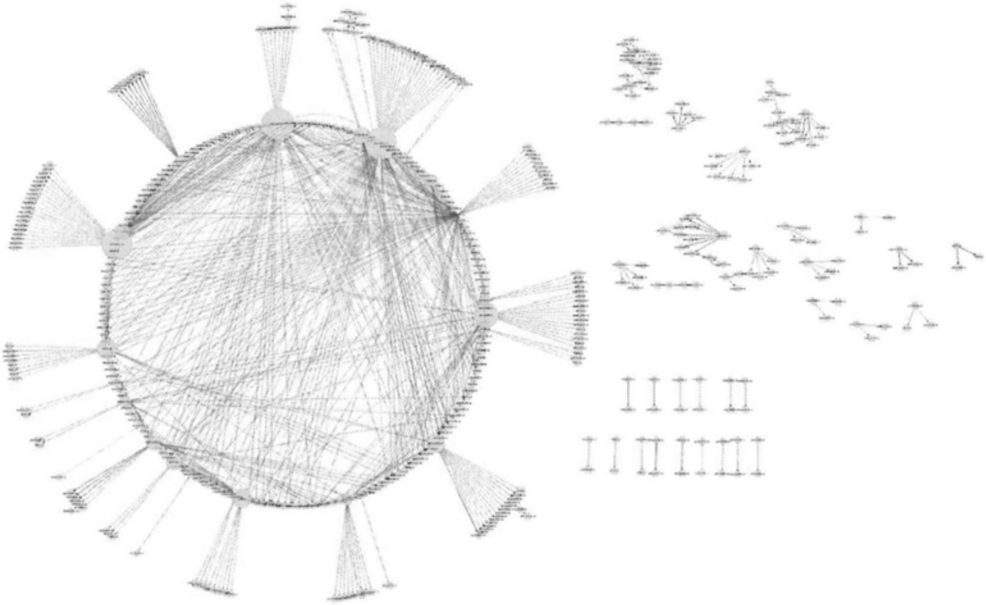


Figure 3.3: **The recovered regulation network with 60% precision.** Pink and blue circles correspond to the transcription factors and target genes respectively. The size of the circle corresponds to the out-degree of gene in this network. Green arrows represent the interactions including in RegulonDB.

a 60% confidence (28 total) were selected. For each group of operons regulated by the same transcription factor, we analyzed the sequences approximately 150 base pairs upstream of the transcription start site with the MEME multiple alignment system [105]. The promoter sequence of these genes can be found in Appendix A.2. Of these 28 regulators, 6 (*FliA*,*LexA*,*Fnr*,*DnaA*,*Nac* and *PurR*)had a known motif in PRODORIC (http://prodoric.tu-bs.de/) [106]. For 4 (*FliA*,*LexA*,*DnaA* and *Nac*) of the 6 MEME-predicted motifs (67%), motifs with best matches were identified.

FliA is a minor sigma factor responsible for the initiation of transcription at a number of genes involved in motility. Notably most of its targets are genes required for flagella synthesis. From prediction, the *FliA* protein regulates 52 genes that can be organized into 19 operons. And 40 out of the 52 genes can be validated by RegulonDB. Interestingly, all 19 operon promoters contain a highly significant motif almost identical to the known canonical *FliA* motif (Figure 3.4).

LexA represses the transcription of several genes involved in the cellular response to DNA damage or inhibition of DNA replication [107, 108] as well as its own synthesis [109]. From the predicted regulation network, *LexA* regulate 10 genes that can be organized into 9 operons. The identical *LexA* regulatory motif can be found at the 8 out of the 9 promoters(Figure 3.5) and 8 out of the 10 genes validated by RegulonDB.

DnaA is the linchpin element in the initiation of DNA replication in *E.coli*. It initiates the process of replication by binding the the origin of replication (*oriC*). From the predicted regulation network, *DnaA* regulates 7 genes that can be organized into 6 operons.

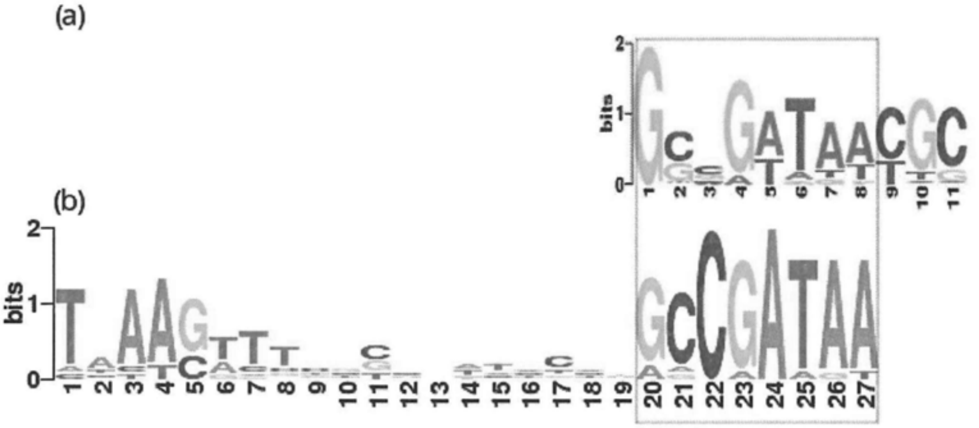


Figure 3.4: **Motifs Detected for transcription factor *FliA*.** (a). The *FliA* regulatory motif detected in the promoters of 18 out of the 19 operons inferred to be *FliA* targets; (b). The *FliA* regulatory motif from PRODORIC database [106].

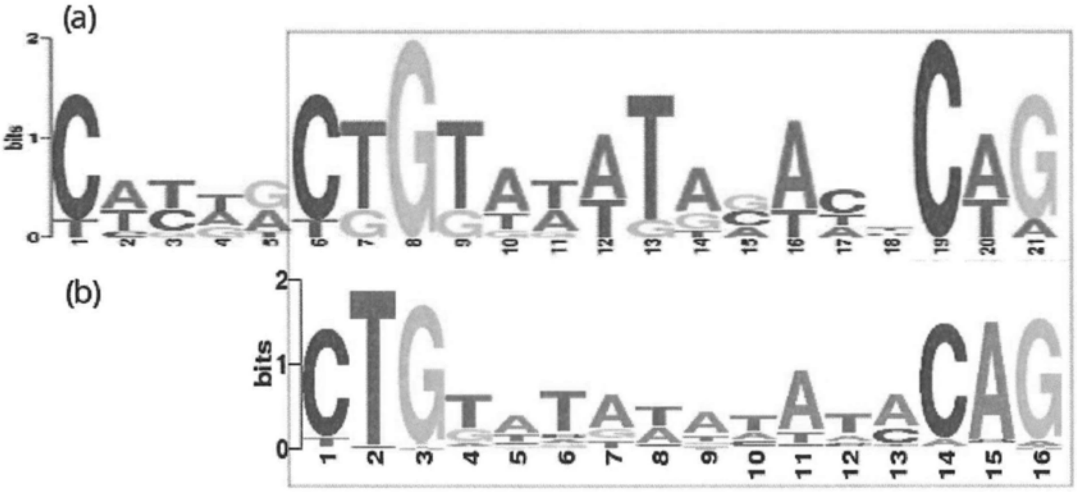


Figure 3.5: **Motifs Detected for transcription factor *LexA*.** (a). The *LexA* regulatory motif detected in the promoters of eight out of the 9 operons inferred to be *LexA* targets; (b). Bottom: The *LexA* regulatory motif from PRODORIC database [106].

The identical *DnaA* regulatory motif can be found at all the promoters(Figure 3.6).

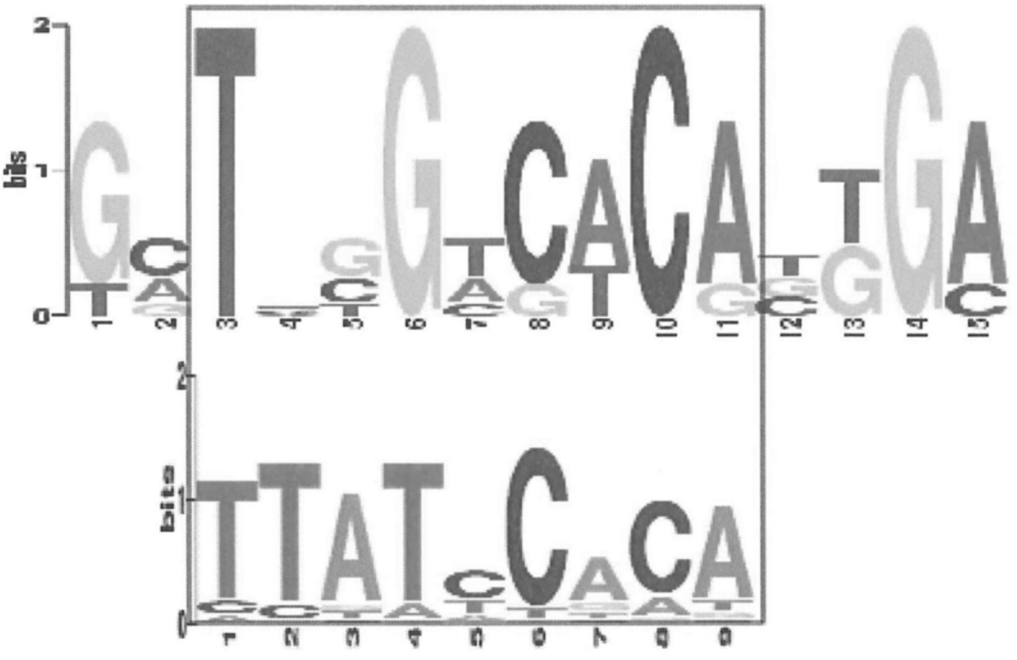


Figure 3.6: **Motifs Detected for transcription factor** *dnaA*. (a).Top: The *dnaA* regulatory motif detected in the promoters of 6 out of the 6 operons inferred to be *dnaA* targets; Bottom: The *dnaA* regulatory motif from PRODORIC database [106].

Nac, "Nitrogen assimilation control," regulates, without a coeffector, genes involved in nitrogen metabolism under nitrogen-limiting conditions [110]. From the predicted regulation network, *Nac* regulates 40 genes that can be organized into 26 operons. The identical *Nac* regulatory motif can be found at the 11 out of the 26 promoters(Figure 3.7).

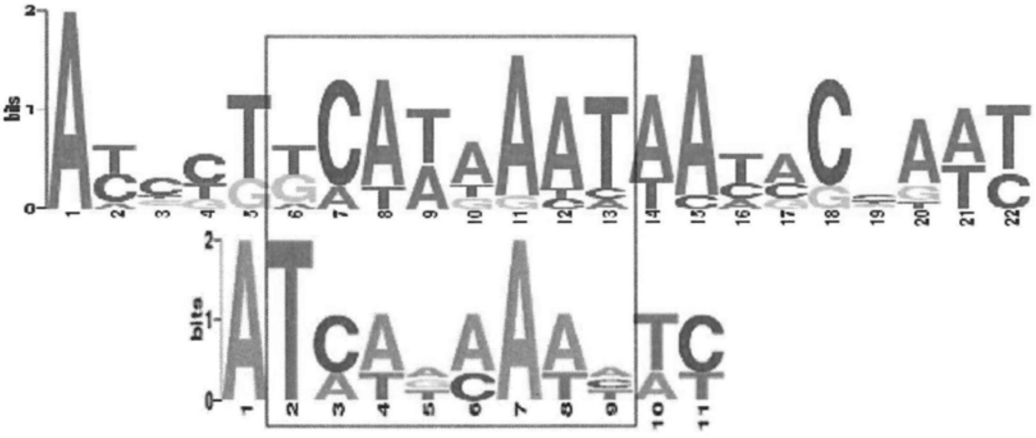


Figure 3.7: **Motifs Detected for transcription factor** ***nac*****.** (a).Top: The *nac* regulatory motif detected in the promoters of 11 out of the 26 operons inferred to be *nac* targets; Bottom: The *nac* regulatory motif from PRODORIC database [106].

3.4 DBoMM is robust to the noise

A good estimator should be robust to noisy data. Because the real gene expression profiles are from biological experiments and it is hard to estimate the noise, we used SynTReN, an artificial synthetic dataset generator, to generate simulated gene expression profiles with various levels of noise. We plot the PR-curves of DBoMM using 5 simulated datasets (Figure 3.8), and found that similar performance was achieved when using datasets with 20%,40% and 60% of noise level. The precision decreased greatly only when 80% noise was introduced.

3.5 DBoMM is able to identify condition-dependent regulatory interaction

The regulatory interactions between TFs and their target genes vary under different experimental conditions [54]. DBoMM can not only estimate the similarity of two genes, it also helps to identify the experimental conditions under which the predicted interactions take place. In the reference regulatory network, *lexA*, which is involved in the cellular response to DNA damage or inhibition of DNA replication, regulates the transcription of gene *recA* in SOS response [107, 108]. From Figure 3.9, *lexA* positively regulates *recA*, and based on the gene expression profile, the mixture model classifies the conditions into 6 clusters. For the first cluster, the values of *lexA* and *recA* are about 8.7 and 8.5(low expression). When examining the samples in this cluster, we found 2 type of conditions:

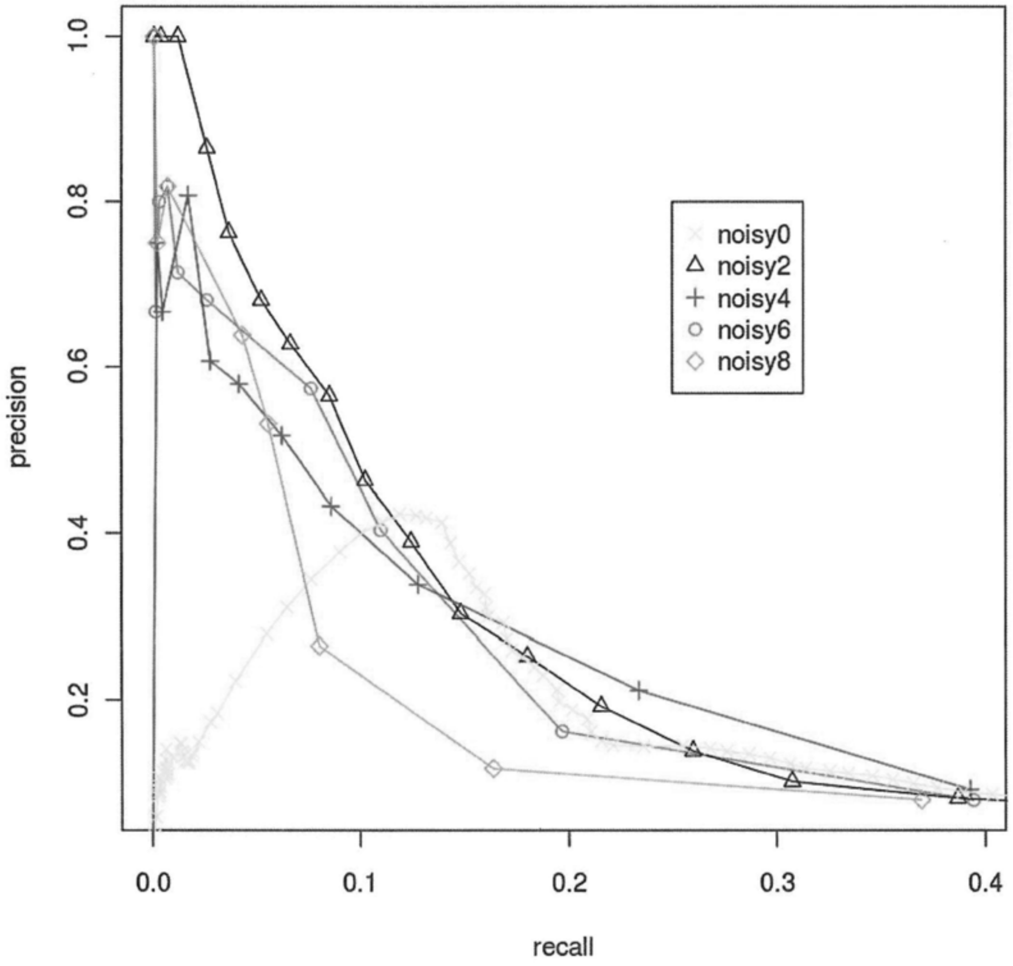


Figure 3.8: **The PR-cure of DBoMM applied to the datasets with different.**

the *recA* knock-out mutants and *E.coli* strain MG1655 at late log phase in LB with newly added glucose and MgSO4. It is reasonable that the expression of *recA* gene is low in the knock-out mutant. When glucose is added into the media at the late log phase, the DNA replication and bacteria growth resume and the expression of *lexA* and *recA* are low. We also checked the conditions of the 4th and the 5th clusters(high expression of *lexA* and *recA*) and found that they are mostly gene over-expression samples, indicating over-expression of these genes can up-regulate the expression of *lexA*, which then up-regulate the *recA*. Compared to the fourth and the fifth clusters, the expression of *recA* gene in the sixth cluster are much higher when the expression levels of *lexA* are similar. The sixth cluster includes two conditions: the *recA* over-expression mutants and cells treated with norfloxacin. This indicates that norfloxacin can active the expression of *recA* and maintain the expression of *lexA*. In fact, it is known that norfloxacin can inhibits DNA synthesis in *E.coli* and causes an accumulation of single-stranded DNA fragments capable of activating the *RecA* protein [111–113].

Based on the mixture model, the DBoMM estimates the expression profiles similarity of two genes, and the similar conditions can be clustered together to indicate the experimental conditions under which the regulatory interaction take place. This feature is useful to guide experimental design and system redesign in synthetic biology.

□ **End of chapter.**

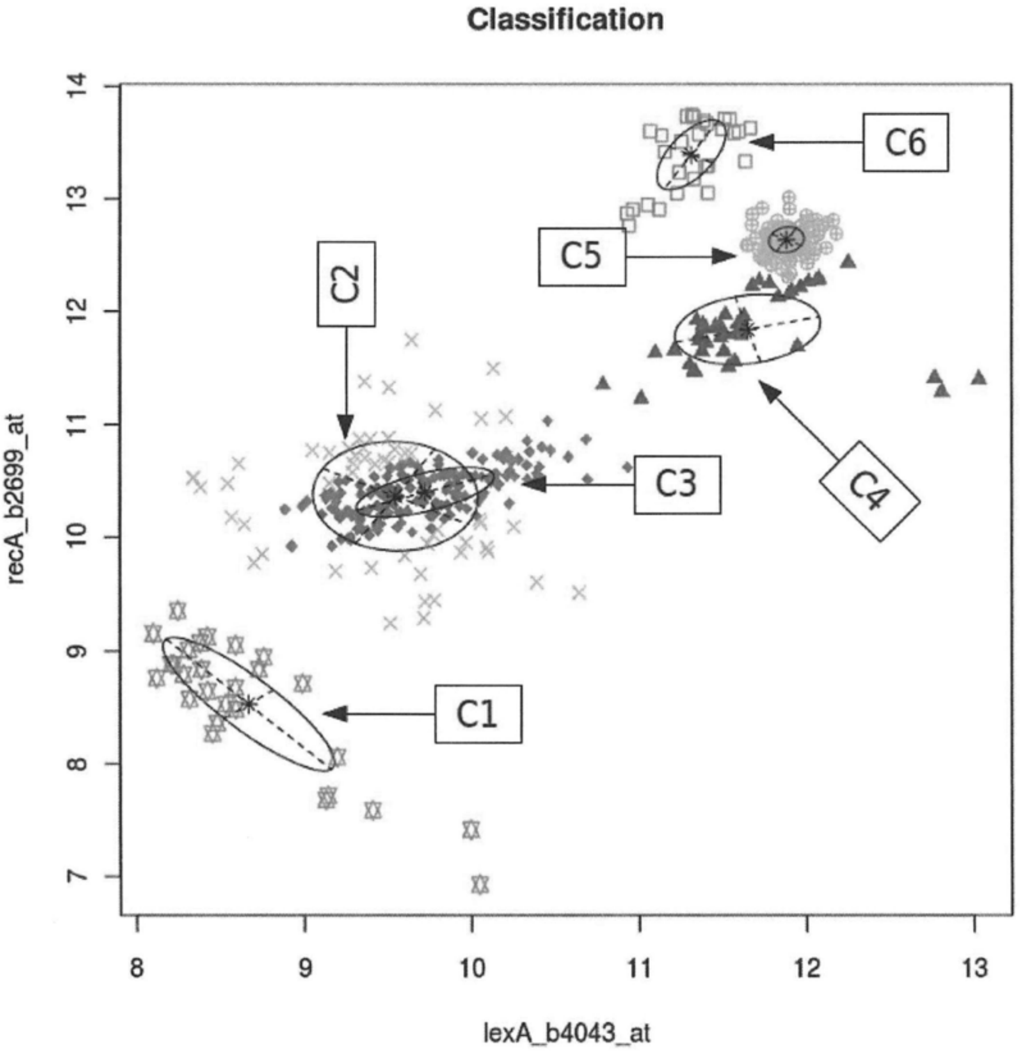


Figure 3.9: **The expression profiles clustered by DBoMM.** Cn represents the index of the cluster.

Chapter 4

Redesign of transcription regulation using a mixture model

4.1 Using inferred distribution of gene expression to represent the gene expression level

It is assumed that the inferred gene expression follows a mixture univariate Gaussian distribution. Based on this distribution, one can obtain the probability of any possible expression value for this particular gene. The inferred expression value, which refers to the expected values of the inferred mixture distribution, was assigned to a particular gene so that it is comparable with the real gene expression level. A comparison between the experimental and the inferred values for 3 randomly selected genes is illustrated in Figure 4.1. The possible expression values of gene "$dsbE$"," $hyfG$" and "$sgaH$" follow mixture univariate Gaussian distributions with 4,4 and 5 components respectively(Figure 4.1a). The inferred values (red line) were found to be similar to the experimental values(blue line). Figure 4.1b gives a visual representation of the difference between the experimental and the inferred values. Again, these two values were almost the same under most conditions, demonstrating the good performance of this model.

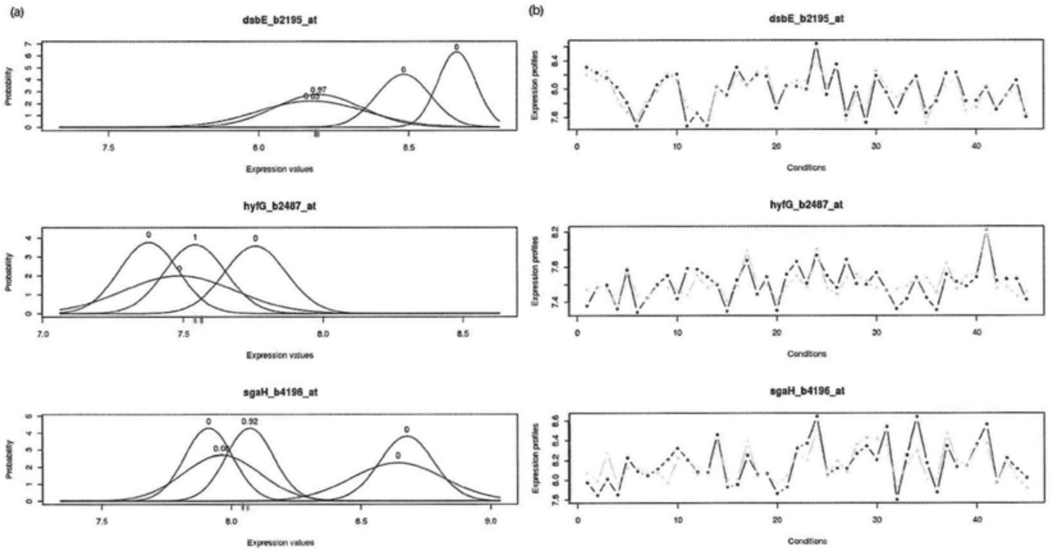


Figure 4.1: **A comparison between the experimental and the inferred gene expression values.** a. The inferred distribution of gene expression values. X-axis: gene expression values; y-axis: the probability of each value. The red and blue vertical lines represent the inferred and experimental value respectively. The values above curves are the weight of each component. b. Experimental and inferred gene expression value. X-axis: experimental conditions in the test set; y-axis: expression values. Black and grey lines represent the experimental and the inferred values respectively.

4.2 The number of components has limited effects on the model performance

The number of components in the Gaussian mixture model is a very important parameter in model learning [114–117]. In this project, to balance the fitting and the complexity of the model, Bayesian Information Criterion (BIC) [93] is used to automatically determine the number of the components. Because the number of the components can affect the predictive power, we arbitrarily defined the number of component (from 1-10) during the model training process, and compared the Relative Errors (REs) calculated from the these models to that from BIC. Specifically, in each learning process, the expression profiles are classified into the components with predefined number.

Figure 4.2 shows the effect of component number on the model performance. In general, increased number of component has limited effect on the model performance when there is more than one component. We chose BIC to automatically determine the number of component because: 1. more components may cause the model's over-fitting to the training data; 2. when the training dataset is small, arbitrarily assigned component number may cause malfunction of the model. For example, we found that when the number of component was arbitrarily set to 10, some of the components became empty using a training dataset of 400 samples. In addition, the model performance was similar to BIC when more components were used (Figure 4.2).

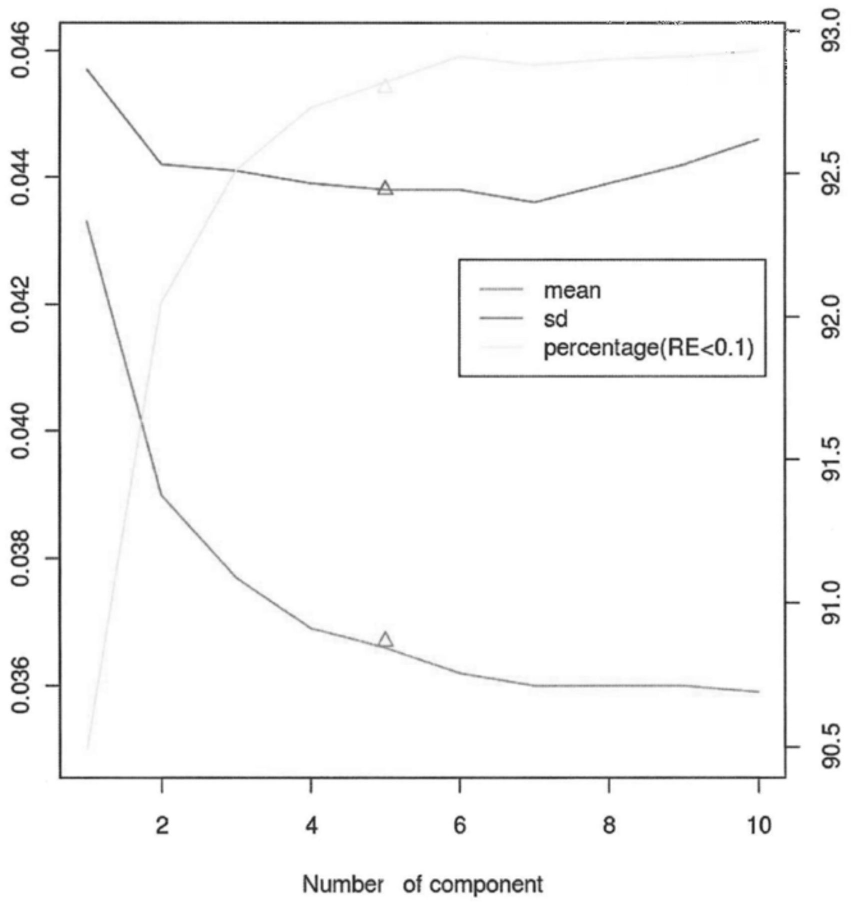


Figure 4.2: **Effect of component number on the model performance.** X-axis: the number of components arbitrarily defined in learning process. Y-axis: left, the mean and sd of REs; right, the percentage(RE<0.1) of REs. Triangles correspond to the mean, sd and percentage(RE<0.1) of REs by BIC.

4.3 The training sample size has limited effect on the model performance

To test the effect of training sample size on the model performance, 10 subsets of data containing various portions (100%, 90%, 80%, 70%, 60%, 50%, 40%, 30%, 20% and 10%) of the 400 training samples were used as the training set respectively.

From Table 4.1, best performance was achieved (RE mean and standard deviation are 0.0379 and 0.0449, respectively) when the maximum number of training samples were used. Although the performance declined slightly with reduced sample size, a sound result was achieved even when only 80 training samples were used (Table 4.1). This result indicates that the training sample size has limited effect on the model performance.

4.4 Prediction of gene expression based on its functionally related gene

In real biological systems, it is common that one gene is regulated by two or more TFs in response to different developmental or external cues [118]. To estimate the predictive power for genes regulated by multiple TFs, the 1156 modules were grouped into 6 categories based on the number of TFs in the modules. As shown in Table 4.2, the prediction accuracy increases with the number of TFs in each module. This is expected because the model uses more inputs(TFs) to limit the inferred values. However, the model performance decreased when the number of TFs was larger than 5. We presume that some

Table 4.1: **Effects of training data size on model performance.**

The first column gives the number of samples used to train the model and the portion (%) of the total samples is indicated in the brackets. The column "mean" and "sd" correspond to the mean and standard deviation of RE. The 4th column corresponds to the percentage of REs that are smaller than 0.1.

Number of trained samples(portion)	Mean	Sd	Percentage(RE<0.1)
400(100%)	0.0379	0.0449	92.23
360(90%)	0.0384	0.0455	92.03
320(80%)	0.0386	0.0458	91.98
280(70%)	0.0394	0.0467	91.66
240(60%)	0.0396	0.0471	91.61
200(50%)	0.0402	0.0477	91.35
160(40%)	0.0422	0.0501	90.40
120(30%)	0.0415	0.0488	90.87
80(20%)	0.0434	0.049	90.13
40(10%)	0.0482	0.0575	87.63

Table 4.2: **Predictive power for genes with multiple regulators.**

The column "mean" and "sd" correspond to the mean and standard deviation of RE. The 4th column corresponds to the percentage of REs that are smaller than 0.1.

Number of TFs	Mean	Sd	Percentage(RE<0.1)
1	0.0397	0.0442	0.9186
2	0.0336	0.0408	0.9367
3	0.0316	0.0421	0.9453
4	0.0294	0.0367	0.9513
5	0.0258	0.0352	0.9625
>5	0.0292	0.0404	0.9475

of these TFs may regulate the target gene expression under different experimental conditions, whereas such information is not reflected in a given regulatory network. This may be one of the reasons to explain the decreased model performance when multiple TFs are involved.

One unique feature of our model is that, because it is assumed that the expression of genes within the same module follow a joint distribution and their causal relations are neglected, then we can infer the expression profile of a given gene based on any functionally related gene, e.g TFs and their target genes. To demonstrate this feature, the inferred values were compared to the experimental values and, 95% of REs lays within interval 0 and 0.2 (Figure 4.3a).

Genes participate in the same metabolic pathway can be grouped into one module and based on the expression of only one gene, we can predict the expression profiles of all the other pathway genes. To demonstrate this useful function, 28 known genes involved in the citrate cycle(TCA cycle) pathway, eco:00020 in KEGG [98, 99, 119], were grouped into one module. Given the expression value of a randomly selected gene (*sucD*), the expression profiles of the other 27 genes were predicted. The experimental and predicted expression values for selected genes in TCA cycle were illustrated in Figure 4.3b.

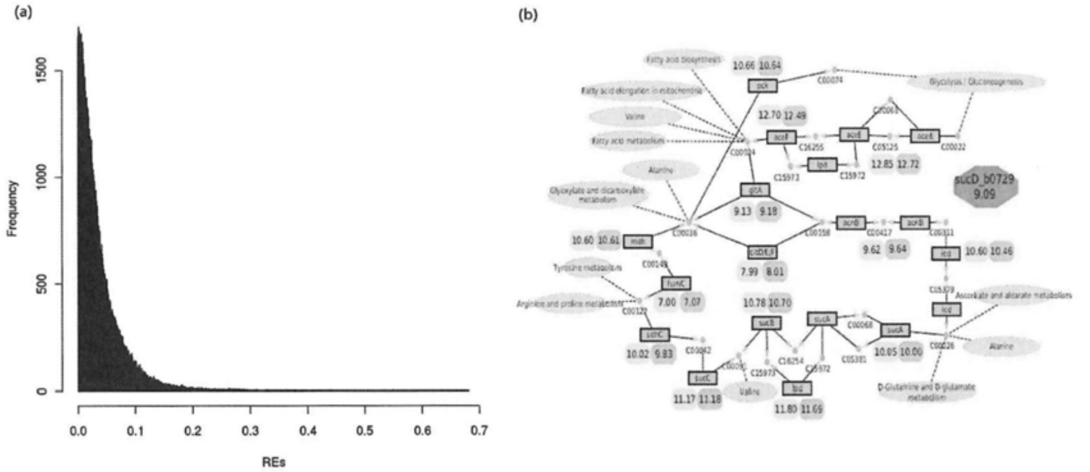


Figure 4.3: **Prediction of gene expression based on its functionally related gene.** a. Inference of TF expression profiles from their target genes. b. An application in TCA cycle genes. The framed green rectangular represents the genes in the TCA pathway. The yellow and green rectangular represent the experimental and the predicted gene expression values.

4.5 A comparison with multiple linear regression method

In the papers [84, 120], regression analysis is used to infer the gene expression profiles based on the microarray data. Carrera et.al [84] adopted a multiple linear regression

(MLR) method [97] to recover the kinetic parameters of gene relations. We compared the performance of the proposed mixture model with that of MLR using the same data. As shown in Table 4.3, the mixture model showed higher predictive accuracy compared to MLR. Similar to the mixture model, the performance of MLR increases with the number of regulators.

Table 4.3: **Predictive power of mixture model and MLR.**

The column "mean" and "sd" correspond to the mean and standard deviation of RE. The 4th column corresponds to the percentage of REs that are smaller than 0.1. The last row "overall" corresponds all analysis results including various number of TFs.

	Mean		Sd		Percentage(RE<0.1)	
Number of TFs	mixture model	MLR	mixture model	MLR	mixture model	MLR
1	0.0477	0.0518	0.0543	0.056	88.62	87.20
2	0.0412	0.0474	0.0494	0.0519	90.9	88.49
3	0.0402	0.0476	0.0506	0.0545	91.09	88.68
4	0.0388	0.0432	0.0464	0.0489	92.09	90.14
5	0.0337	0.0385	0.0406	0.0469	94.17	91.78
>5	0.0410	0.1004	0.0596	0.5572	90.85	82.56
overall	0.0416	0.0511	0.0507	0.1600	90.83	88.28

4.6 Infer the functional links among experimental conditions

In the mixture Bayesain method proposed by Ko Y et al. [75], condition-dependent regulatory interactions can be inferred by clustering the experimental conditions under which related genes show similar expression pattern into the same group. Different from their method, our model assigns a new experimental condition to the known condition groups, thereby to infer the functional links between these conditions. For each trained module, the experimental conditions (training samples) are grouped into different clusters based on expression patterns of genes contained in this module. For example, in a randomly selected module, the experimental conditions were grouped into 2 clusters (separated by dashed yellow vertical line in Figure 4.4) based on the expression patterns of the 3 genes (*appA*, *arcA* and *appY*) in this module. This clustering provides importance clue about the connections of those experimental conditions. If the expression values of the 3 genes under a new experimental condition is used as input for the trained module, the probabilities that this condition belongs to the 2 clusters can be calculated. Such information is useful for biologist to estimate the possible common functional links among these conditions.

Microarray samples usually include 3 replicates. To test the clustering power of the model, we arbitrarily set 1 of the 3 replicates as the new experimental condition and a total of 102 microarray samples were tested. Among them, 95 samples were correctly assigned into the known cluster, containing the rest of the 2 replicates (Figure 4.4).

4.7 Redesign of transcription regulation

Gene knock-out or over-expression are commonly used strategies for gene functional study. A quantitative prediction of transcriptome profile under gene knockout or over-expression

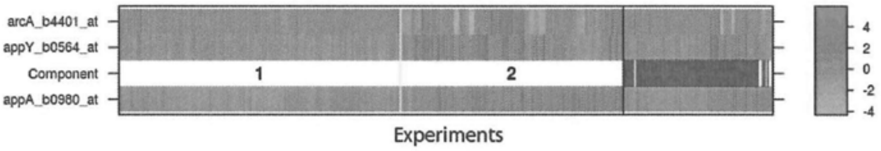


Figure 4.4: **Cluster assignment in one example module.** X-axis: samples in different experimental conditions; y-axis: genes in one module. The result from training data and test data was separated by the black line. Based on the expression profiles of the 3 candidate genes, the training dataset is classified into 2 clusters (components) separated by yellow lines. The "Component" represents the index of cluster the samples belong to. The genes above and under the "Component" are TFs and their target genes respectively. For the test data, correctly assigned samples are labeled in blue and wrongly assigned samples are labeled in white. In total, 95 out of 102 test samples were correctly assigned. Because different genes have different basal synthetic rate, z-score were used to normalize the gene expression values.

can be very useful for biological experimental design or regulatory network redesign. To demonstrate the model's predictive power in this aspect, gene knockout and over-expression test were conducted. The same reference network described before was used and the 445 samples were separated into training and test datasets based on the mutated gene. Specifically, for a particular mutated gene, the gene expression data measuring the transcriptome of this mutant was defined as the test set, and other arrays as training set. Because some genes are directly regulated and some are indirectly regulated by the mutated gene, the results were separated into two sets: one contains the directly regulated genes, and the other contains both directly and indirectly regulated genes.

In the reference network, gene *rpoD*, *crp*, *himD*, *himA*, fnr, fis and *arcA* directly regulate 779, 265, 161, 161, 146, 130 and 97 target genes respectively. However, the gene knockout/over-expression microarray data are available for only 4 of them (*crp*, fnr, fis and *arcA*). In *E.coli*, *crp* (cAMP receptor protein) is an important transcriptional dual regulator involved in various biological processes, such as osmoregulation [121], stringent response [122], virulence [121], nitrogen assimilation [123], iron uptake [124], and multidrug resistance to antibiotics [125]. Fis , "factor for inversion stimulation", encodes a small DNA-binding and bending protein, which directly modulates transcription, chromosomal replication, DNA inversion, phage integration/excision, and DNA transposition [126, 127]. Fnr and *arcA* are the primary transcriptional regulators that mediates the transition from aerobic to anaerobic growth through the regulation of hundreds of genes. The model was tested under the following system perturbation : 1.*crp* over-expression;

2.fis over-expression; 3. double knockout of $arcA\_fnr$. The number of microarray samples measuring these 3 mutants was 6, 3 and 22 respectively. The predicted results were compared to the real experimental data(Figure 4.5). As shown, the model correctly captured the expression profiles of most genes under system perturbation. This once again demonstrated the quantitative power of this model in guiding global TRN redesign, such as in the case of over-expression and knockouts of master regulators.

Except for loss-of-function and gain-of-function study, adding new regulations has also been used to study the network evolvability. For example, Isolan et al. over-expressed plasmids pairing together wild-type promoters with ORFs coding for TF that were master regulators [128]. Our model was able to predict the gene expression profile under such transcriptional rewiring particular in the case where the *rpoS* and *malT* promoters are disposed together with ORFs *ompR* and $fliA$, respectively(Figure 4.6).

□ **End of chapter.**

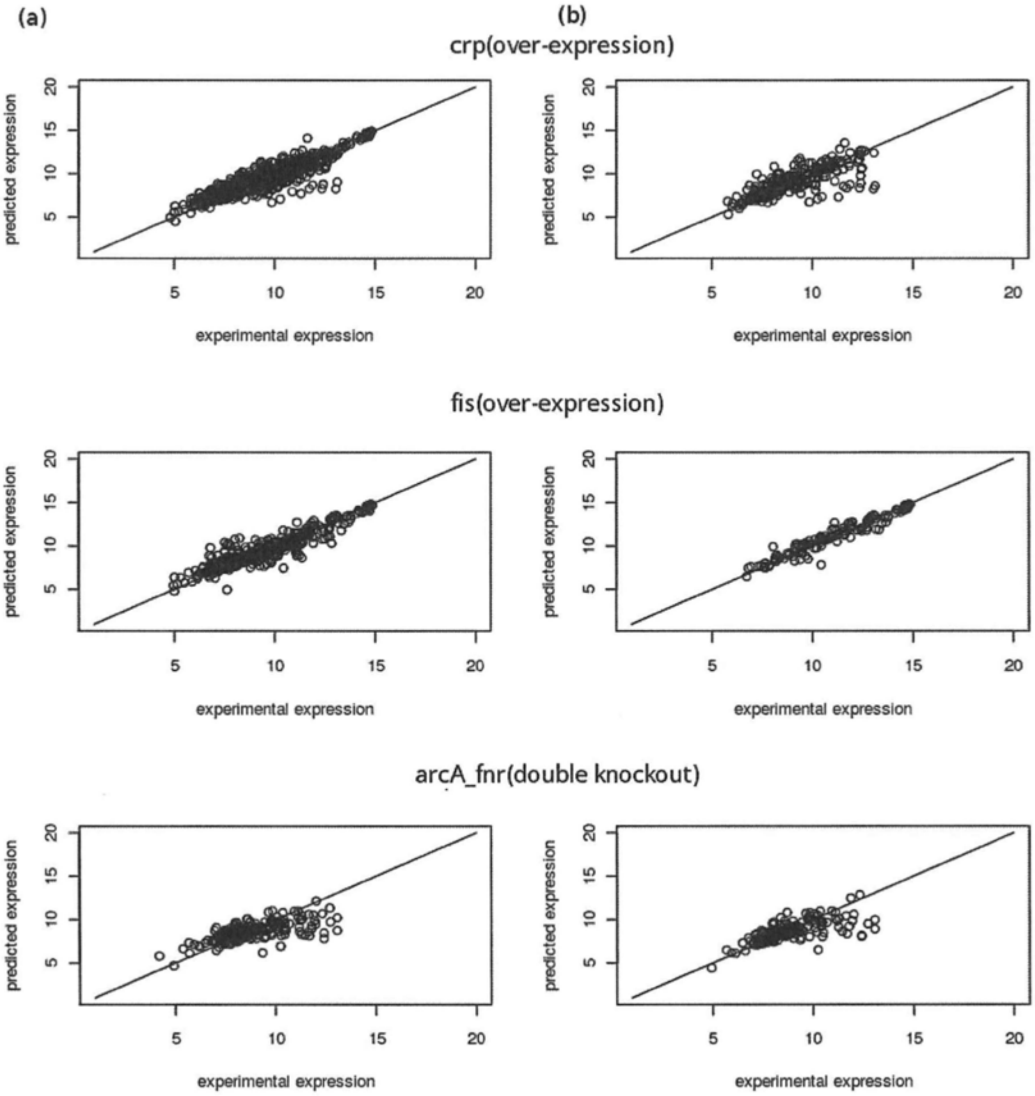


Figure 4.5: **Prediction of gene expression profiles under system perturbation.** a. Directly and indirectly regulated gene. b. Directly regulated gene.

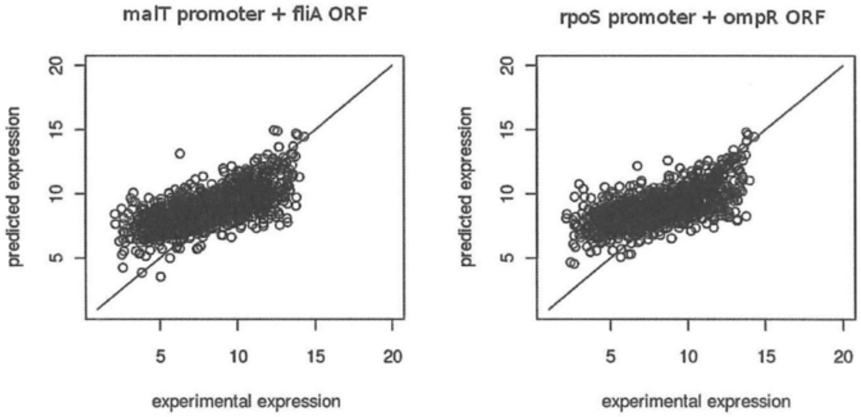


Figure 4.6: **Prediction of gene expression profiles in transcriptional rewiring of the wild-type regulatory map.** The promoters of *rpoS* and *malT* were put together with the ORFs of *ompR* and *fliA* in high-copy plasmid, respectively.

Chapter 5

Conclusion

In this part, we describe a model-based clustering method for gene similarity measurement based on their expression profiles. As proposed by Segal [43], the gene regulatory interactions can show similar or same pattern under different conditions. Based on this notion, we fit the gene expression profiles into a mixture Gaussian model. The experimental conditions, under which the pattern of regulatory interactions are similar, are assigned into different components. Because the mixture model is a fuzzy clustering and "soft" classification method, the probability of each sample belonging to the components is used to estimate the density of the components and calculate the observed probability of the samples. We used BIC to describe the fitness of gene expression profiles to the model. The difference of BIC between the joint and the marginal distribution model of expression profiles is used to estimate the similarity of genes. A Gaussian distribution is adopted to estimate the density of the cluster. The advantage of the mixture model lies in its flexibility in choosing the component distributions. For example, we can use an additional Poisson distribution to handle the noisy points. This method is also robust to the noise.

We have successfully applied DBoMM to both *E.coli* gene expression dataset and synthetic datasets, and proved that the model achieved better performance than COR and EUC. DBoMM also out-performed MI using synthetic dataset and yet the performance was comparable to MI using the *E.coli* dataset. DBoMM does not request the linear relationships between genes and can catch both the local and the global correlations. Compared to the method calculating MI from expression profiles, DBoMM uses mixture model to estimate the probability, and can infer the experimental conditions under which the predicted regulatory interaction take place.

Mixture models also can be used to calculate MI. In fact, mutual information (MI) is equivalent to the difference between the joint entropy and the conditional entropy. There are several methods to estimate the parameters of the model by using entroy of the mixture model [116, 129]. And the joint and marginal entropy of gene expression profiles under mixture distribution can be used to calculate the mutual information of genes.

Then, we extend the mixture distribution model used for gene network inference to a quantitative model with predictive function. By fitting the expression values of related genes to a mixture Gaussian distribution, the model parameterizes a given gene regulatory network inferred from various network inference methods, (e.g CLR [130], Bayesian Network [131, 132], ODE [133], ARACNE [134], mixture Bayesian network [74, 75], or even biological experiments) and then infers the conditional distribution of the expression of one particular gene, given the expression values of any related genes. Compared to the model developed by Ko Y et al. [75] , which adopts similar algorithm to infer the regulatory network, our method can quantitatively predict the gene expression profile based on the learned statistical parameters and the topological structure of the TRN. We have successfully applied the model to accurately predict the *E.coli* transcriptomic response under various experimental conditions (average REs <0.05). Furthermore, the model also performs well under genetic rewirings and over-expression/knockout of master regulators. Except for parameterizing the gene relations in a regulatory network, the model can also be used to predict the expression profiles of a particular gene based on the expression of any functionally related genes. We demonstrated that our model can correctly predict the 27 genes participate in the TCA cycle based on the expression of only one gene in the same pathway. We expect that this model can be widely used for synthetic biology system redesign and biological experimental design. This quantitative model can also be extended to simulate the gene expression data for the evaluation of network inference algorithms. For a regulatory network derived from experiments, the learned model generally infers the expression profiles of genes by specifying the values of several global regulators, which are not regulated by other genes in this network. Compared to Bulcke et.al's method [88], our algorithm calculates the conditional distribution of learned mixture model and produce expression profiles more faithfully representing the regulatory relations between genes. In addition, users can flexibly set the different expression values of global regulators to simulate the transcriptome under different conditions, which is limited by other method [135].

The model also has some limitations. For example, in principle, post-transcriptional and post-translational regulation also plays important roles in most of the cellular events. Here we neglect these effects and simply assume that the mRNA amounts measured by microarray is proportional to the protein amount and is the function of the TFs only. In addition, because the model is based on the statistical distribution of gene expression, limited amount of training samples may lead to the uncertainty of parameter estimate, which affects the predictive power. For genes not involved in the reference network, the model can not predict the values of these genes. The model is therefore more efficient when being applied to model organisms in which a large number of training samples and interaction relations are available.

□ **End of chapter.**

Part II

A sub-space greedy search method for efficient Bayesian Network Inference

Chapter 6

Introduction

Bayesian Networks (BN) have been widely adopted to infer genetic network using high-throughput dataset [136]. A Bayesian network(BN), belief network or directed acyclic graphical model is a graphical model for probabilistic relationships among a set of random variables. These relationships between variables are described by conditional probability distribution, which means the expression profiles of genes are affected by their regulators in a GRN. Based on the *Markov assumption*, that is, each variable X_i is independent of its non-descendants, given its parent in a DAG, the joint probability distribution of variables can be decomposed as the product of conditional probabilities:

$$P\left(X_1, \ldots, X_n\right) = \prod_{i=1}^{N} P\left(X_i | Pa\left(X_i\right)\right) \tag{6.1}$$

where X_i represent the variable; $Pa\left(X_i\right)$ represents the parents of X_i, the regulators of X_i in a GRN.

To infer a GRN based on Bayesian network model from gene expression data D, we must find a score function to estimate how the inferred directed acyclic graph G describes the data set D. So this structure learning process has been switched to search for the graph G that maximises the value of the score function. The score can be defined using the Bayes rule:

$$P\left(G|D\right) = \frac{P\left(D|G\right) * P\left(G\right)}{P\left(G\right)} \tag{6.2}$$

where $P(G)$ either contain *prior* knowledge on network structure, if available, or can be a constant non-informative prior, and $P(D/G)$ is a function, to be chosen by the algorithm that evaluates the probability that the data D has been generated by the graph G. The most popular scores are the Bayesian Information Criteria (BIC) [93] or Bayesian Dirichlet equivalence (BDe) [137]. Both scores incorporate a penalty for complexity to guard against overfitting of data.

It is an NP-hard problem to find the G with the maximum Bayesian score by searching all possible graphs G, because the number of all possible graphs G grows exponentially with the increasing nodes. Therefore, a heuristic search method is used, like the greedy-hill climbing approach [138], the Markov Chain Monte Carlo method [139, 140] or simulated annealing [141–143].

The most commonly used score-based Bayesian Network learning algorithm is greedy hill-climbing, which starts from a candidate network and then iteratively moves to a neighbor network that leads to the largest score improvement. During this process, the number of changes is denoted as O (n^2), where n is the number of variables. Because the number of possible networks grows more than exponentially with the number of the variables, the cost of calculation becomes acute when BN is applied to high-throughput microarray data. However, most false candidate gene pairs or networks resulted from the search process should be eliminated in reality. For example, in a small network containing 3 genes (namely X, Y, and Z), if X regulates Y and Z is not related to X and Y, the X-Z pair and Y-Z pair shouldn't be considered during the network inference. To reduce the computational cost, a measure of dependence between variables should be performed to restrict the search space before constructing the networks.

Based on this notion, mutual information has been proposed for measure of dependence between variables in network reconstruction [144]. Another simple method to infer the dependence between variables is to compute all the pair-wise correlations. However, the correlation coefficient is a weak criterion for measuring dependence because it only reflects marginal independence and indirect dependence. Partial correlation coefficient (PCC) measures the degree of association between two random variables with the effects of controlling random variables removed, and therefore provides a strong measurement of dependence [?, 145–148]. In this part, we propose a sub-space greedy search method based on partial correlation coefficient to estimate the dependence between variables and to restrict the search space. We demonstrate that our model can greatly reduce computational cost with minimum tradeoffs in network accuracy.

□ **End of chapter.**

Chapter 7

Methodology

7.1 Generation of synthetic dataset

Using simulation program SynTReN [88], we selected n genes (n=10, 15, 20, 25 and 50 respectively) and obtained independent datasets with 1000 observed samples, each contains n genes. These synthetic datasets were used to reconstruct the transcription network using the classical greedy search and the proposed sub-space search method.

7.2 Selection of Real Gene Expression Dataset and Reference Network

We adopted the microarray dataset comparing gene expression in Acute Lymphoblastic Leukemia (ALL) patients and Acute Myeloid Leukemia (AML) patients (27 ALL and 11 AML) using Affymetrix Hu6800 GeneChip$^{\mathrm{TM}}$. The chip contains 7129 gene-specific probe sets representing approximately 6817 genes [149]. Using this dataset, Amira Djebbari et al. carried out seeded BN inference to obtain a standard network containing 41 genes [150]. In this study, we used their inferred network as reference to compare the performance of proposed sub-space greedy search method to that of the classical greedy search algorithm.

7.3 Learning Bayesian Network

In graphical model representation, a Bayesian network (BN) is a directed acyclic graph (DAG) representing a joint probability distribution (JPD) of over all variables. The nodes in the DAG represent the variables and edges represent the relationship between variables. In BN, each variable is independent of its non-descendants given its parents, and the relationships between variables are described by conditional probability distributions (CPDs) denoted as p(B|A)- the probability of B given A.

The learning of a BN structure can be stated as: finding a network B that can best match D, given a dataset D$\{D_1, \ldots, D_n\}$. To assess the degree to which the resulting structure explains D, we use the score function of relative probability p (S, D). This score is also used by deal [151], a software package implemented in R [?]. This package includes several methods for analyzing gene expression data using Bayesian networks with variables of discrete and/or continuous types but restricted to conditionally Gaussian networks. BNArray [152] is another package that re-samples microarray data and construct the gene regulation network based on deal. In our study, deal package was used to calculate the CPD of gene pairs and BN scores, and BNArray package was used to re-sample datasets.

All of inferred networks by sub-space greedy search and classical greedy search are from 100 bootstrap iterations with bootstrap confidence greater than 0.5 (occurring in more than 50% of iterations).

7.4 Measure of Dependence

To measure the dependence between two genes, partial correlation coefficients of all the possible gene pairs was calculated using GeneNet [148] package implemented in bioconductor [153].

7.5 Structure learning using Sub-space Search Algorithm

In classical Bayesian Network, the greedy search algorithm explores all the candidate networks and selects the one with the highest score during iteration until the network convergence. Because the arrow deletion and turning processes are based on the DAG which has finite edges, both processes cost only limited computational time compared to the arrow addition process. We therefore proposed a method to restrict the search space in arrow adding process by selecting gene pairs with higher PCC values.

A detailed description of the algorithm is as follows:

1. Based on the partial correlation coefficient (PCC) of all the possible gene pairs, we construct a matrix (M_p) that indicates the possible regulatory relationship between genes. The rows in M_p correspond to different variables (child genes) and the columns correspond to all the potential parents of these variables. The parent genes are indexed

based on their PCC with the individual variables. For example, in an M_p containing 50 columns, the parent gene resides at the first column (M_p index 1) after the variable has the highest PCC with the variable. Similarly, the gene resides at the last column (M_p index 49) has the lowest PCC with the variable.

2. Select an initial DAG D_0, from which to start the search.

3. Calculate Bayes factor of D_0 and select networks through the following process:

(a) One arrow is added to D_0. Unlike classical greedy search that selects all the genes as the candidate parents for each child, the sub-space search method limits the search space by only selecting gene pairs with higher PCCs (parent genes with higher Mp index, e.g. 1, 2, 3, 4, 5 etc.). To avoid the possible arbitrary effects of selecting high PCC pairs only, user may choose to randomly include some low PCC pairs.

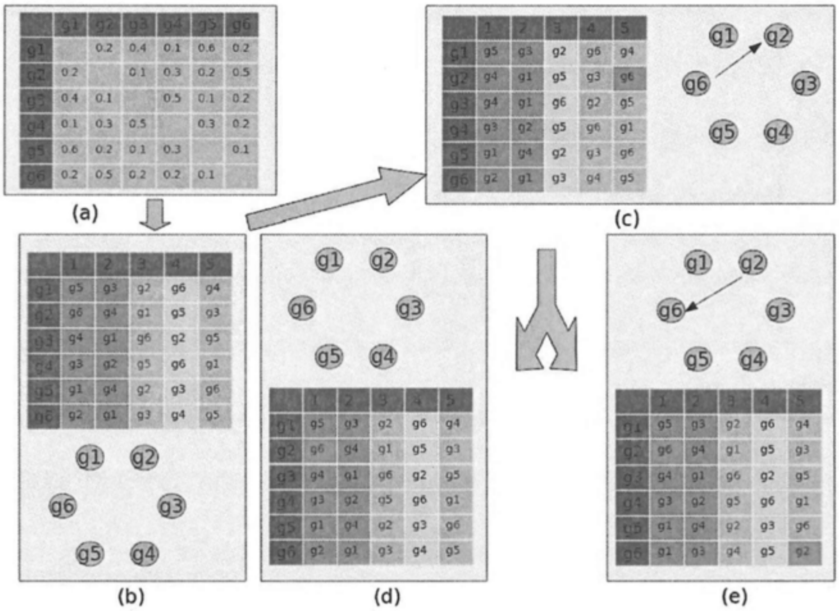
(b) One arrow in D_0 is deleted

(c) One arrow in D_0 is turned (reverted)

(d) Among all the resulted networks, select the one that increases the Bayes factor the most as candidate the DAG (D_c). If the score of D_c is higher than that of D_0, D_0 is replaced by D_c and the process is repeated from step (a). If the score of D_c is lower than that of D_0, the algorithm stops and D_0 is the final DAG.

4. If the Bayes factor is not increased, stop the search. Otherwise, let the chosen network be D_0 and repeat from step 3.

A graphical description of the sub-space search method is illustrated in Figure 7.1.



	g1	g2	g3	g4	g5	g6
g1		0.2	0.4	0.1	0.6	0.2
g2	0.2		0.1	0.3	0.2	0.5
g3	0.4	0.1		0.5	0.1	0.2
g4	0.1	0.3	0.5		0.3	0.2
g5	0.6	0.2	0.1	0.3		0.1
g6	0.2	0.5	0.2	0.2	0.1	

	1	2	3	4	5
g1	g5	g3	g2	g6	g4
g2	g6	g4	g1	g5	g3
g3	g4	g1	g6	g2	g5
g4	g3	g2	g5	g6	g1
g5	g1	g4	g2	g3	g6
g6	g2	g1	g3	g4	g5

	1	2	3	4	5
g1	g5	g3	g2	g6	g4
g2	g4	g1	g5	g3	g6
g3	g4	g1	g6	g2	g5
g4	g3	g2	g5	g6	g1
g5	g1	g4	g2	g3	g6
g6	g2	g1	g3	g4	g5

	1	2	3	4	5
g1	g5	g3	g2	g6	g4
g2	g6	g4	g1	g5	g3
g3	g4	g1	g6	g2	g5
g4	g3	g2	g5	g6	g1
g5	g1	g4	g2	g3	g6
g6	g2	g1	g3	g4	g5

	1	2	3	4	5
g1	g5	g3	g2	g6	g4
g2	g6	g4	g1	g5	g3
g3	g4	g1	g6	g2	g5
g4	g3	g2	g5	g6	g1
g5	g1	g4	g2	g3	g6
g6	g1	g3	g4	g5	g2

Figure 7.1: **A graphical representation of the sub-space greedy search algorithm.** (a) Calculation of PCC of all gene pairs using GeneNet package. (b) Construct a matrix M_p to describe the possible parents for each variable. The rows correspond to variables and the columns correspond to all their parents. The parent genes are listed in a descending order based on their PCC with the child genes (variables). Only higher ranking parents (e.g. in brown columns) are selected to form search space with the corresponding child variable. User-defined low PCC gene pairs (e.g. columns in orange) can be randomly selected in each iteration steps to avoid arbitrary effect. (c) After structure learning, if a DAG with an added arrow (g_1-g_2) is selected, the parent gene g_6 is transferred to the last column (in red) (d) If a DAG with a removed arrow (g_6-g_2) is selected, g_6 is re-transferred to the first column(in red) for the next search. (e) If a DAG with a turned arrow (g_2-g_6) is selected, then two transfer processes are done as described in (c) and (d).

7.6 Estimate of BN inference

Three types of efficiencies, precision(P), sensitivity(S) and absolute efficiency(F), were computed to compare BN inferred network and reference network. P is the fraction of predicted gene pairs that are correct: P = TP / (TP + FP) and S is the fraction of all known gene pairs that are inferred by BN: S = TP / (TP + FN) where TP is the number of true positives, FN the number of false negatives and FP the number of false positives. F thus denotes the absolute efficiency: F = 2PS / (P + S) which is the harmonic mean of precision and sensitivity.

□ **End of chapter.**

Chapter 8

Results

8.1 BN tends to select gene pairs with higher partial correlation coefficients

Using the synthetic datasets generated by SynTReN [88] , a network was reconstructed using BN inference. Comparing the PCCs of BN-inferred gene pairs with that of all the gene pairs (Figure 8.1), we found that PCCs of gene pairs resulted from BN inference follows normal distribution and the number of BN inferred gene pairs increases with increase in absolute PCC. This observation suggests that BN inference tends to select highly correlated gene pairs, which is consistent with the finding that real regulatory gene pairs often contain genes with similar expression patterns and higher PCC compared to the false ones. This result also highlights the rationale of our proposed sub-space search, which is to restrict the search space by selectively choosing gene pairs with higher PCC as an efficient alternative for BN inference.

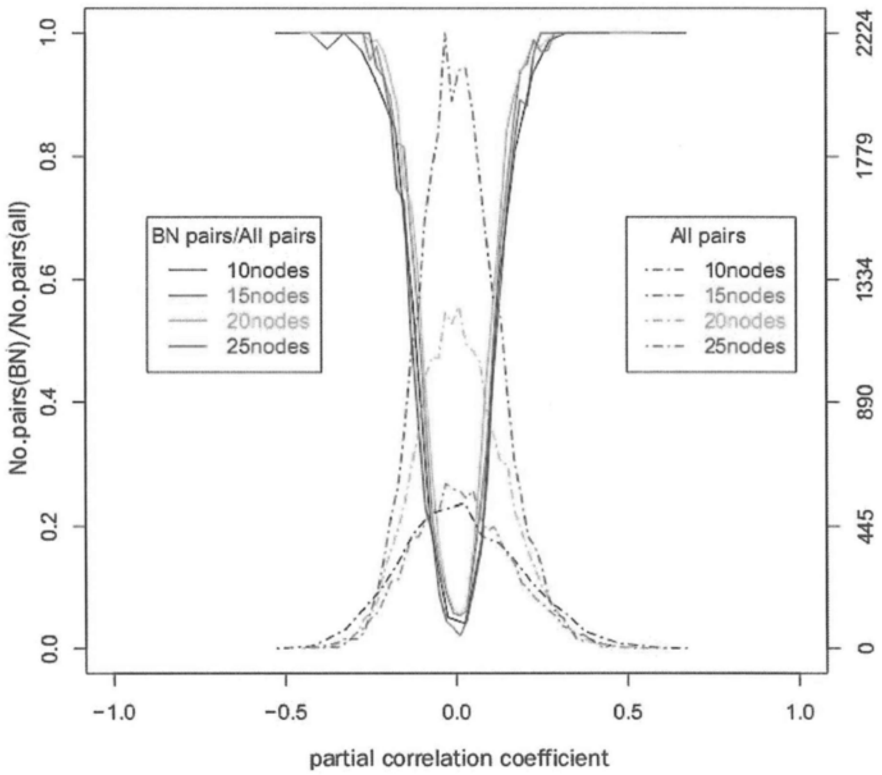


Figure 8.1: **The partial correlation values of gene pairs were plotted against the percentage of BN gene pairs.** Dashed lines: all gene pairs. Solid lines: percentage of BN gene pairs.

8.2 BN tends to infer DAGs with higher PCC in each iteration steps

Using the synthetic datasets, a matrix (M_p) based on PCC was established to indicate the possible regulatory relationship between two genes. To examine if the DAGs with highest PCCs were selected during each iteration step, DAGs with the highest score for each Mp column were collected and sorted based on their scores. As shown in Figure 8.2a, most of the DAGs contain parent genes with higher Mp index. Because only one candidate DAG with the highest score is selected for the next iteration step, we monitored the distribution of Mp index in selected DAGs and found that majority of them contain parent genes with highest Mp index (Figure 8.2b). The results again demonstrated that high PCC DAGs are also the high score DAGs inferred by Bayesian Network.

8.3 Comparison to classical greedy search method using synthetic data

A dataset containing 50 genes generated by SynTReN was used to infer the network using classical search method and the sub-space search method. By using gene pairs with various PCCs (Mp index 1-10, 1-15, 1-20, 1-25, 1-30, 1-35, 1-40, 1-45, and 1-49), the results from BN inference were compared (Figure 8.3). As shown, the consumption of computational time increased almost linearly with the increase of parent genes. However, the network score reached the highest (Table 8.1) and then remained almost unchanged

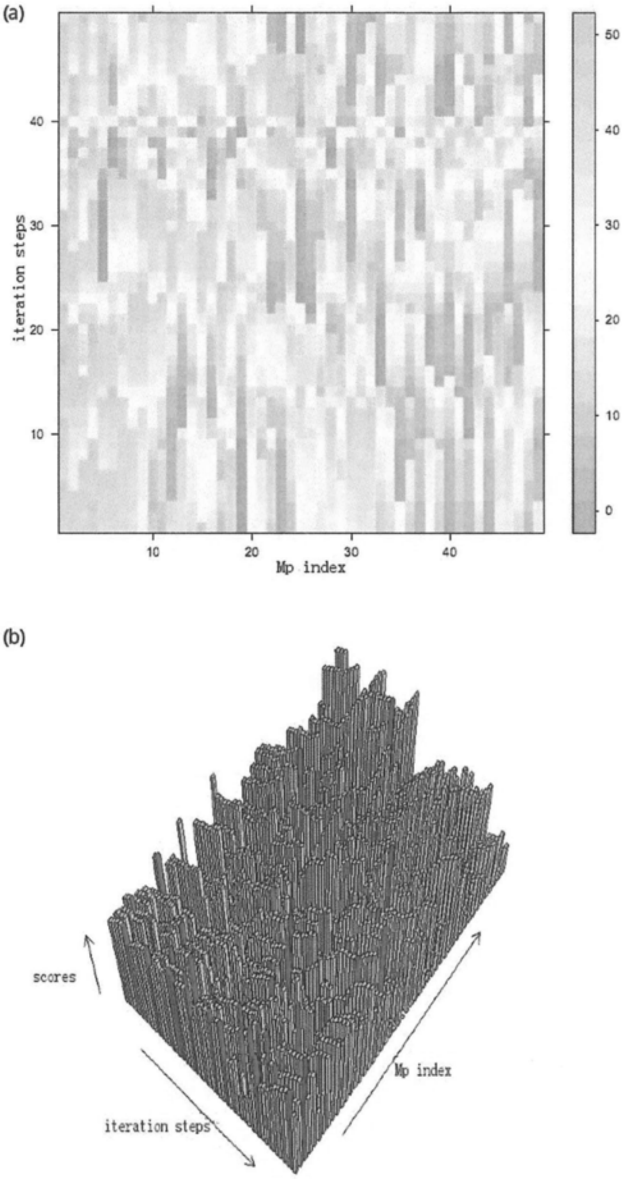


Figure 8.2: **The distribution of sorted M_p index based DAG scores and DAGs selected in each iteration step.** a. DAGs with the highest score in each column were collected in every iteration steps and sorted based on their scores. X-axis: Mp index; Y-axis: iteration steps. The color represents the Mp index of sorted DAGs. b. The heights of cuboids represent the scores of DAGs and cuboids in red means this DAG is selected for next iteration step.

after genes with Mp index 1-25 were used. This demonstrated that by including only a portion of highly correlated gene pairs, the sub-space search method achieved similar performance to the classical method in terms of network score while saved nearly half of the computational time. The highest BN score was obtained when 50% of the total gene pairs (parent gene indexed 1-25 out of 49) were included.

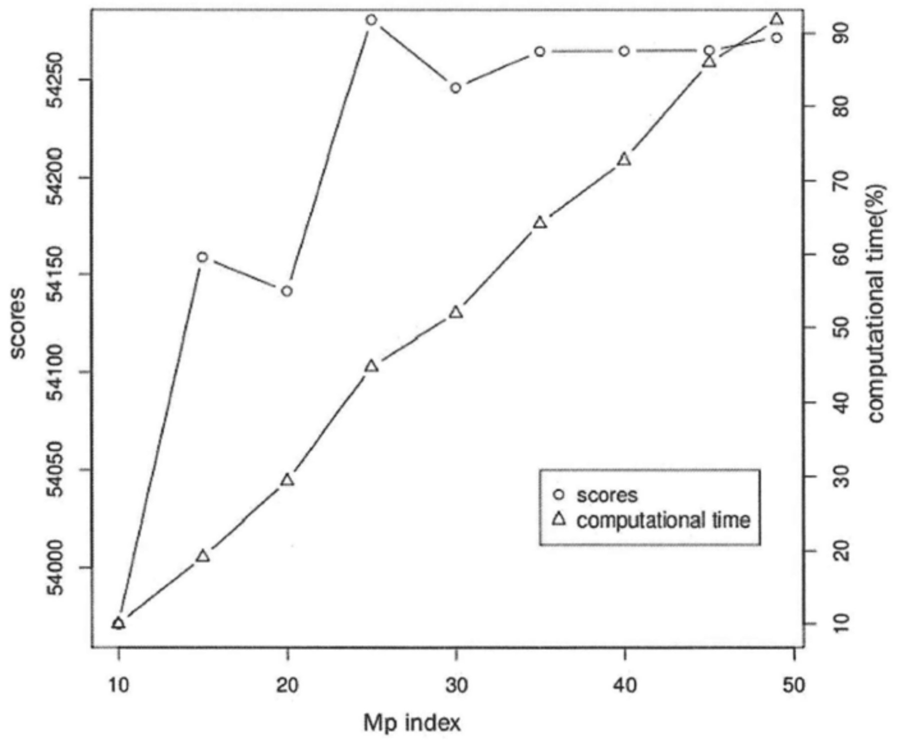


Figure 8.3: **Comparison of BN inference results using synthetic datasets.** X-axis: M_p index representing the portion of parent genes included. e.g. the number 20 means that parent gene resides at the 1st to the 20th columns after the child gene in the M_p are included; Y-axis: BN score (left) and the percentage of computational time consumed.

Table 8.1: **Comparison of standard greedy search and sub-space greedy search using synthetic dataset.**

M_p index	Score	Absolute computational time (seconds)	Relative computational time
10	3970.72	8493.05	0.23
15	4158.87	11705.13	0.31
20	4141.32	15289	0.41
25	4281.34	20711.27	0.56
30	4246.14	23234.37	0.62
35	4264.96	27524.41	0.74
40	4265.18	30530.24	0.82
45	4265.47	35182.1	0.94
49	4272.11	37239.86	1

8.4 Comparison to classical greedy search method using real dataset

A similar comparison was done using the real microarray data (see method) and the results were summarized in Figure 8.4 and Table 2. When parent genes with Mp index 1-25 were used, the inferred network achieved comparable score to that of classical greedy search, but cost only 66% of the computational time. Although the network score reached the highest when genes with Mp index 1-35 were used, the computational cost is around 90%. Considering the tradeoffs of computational cost, it is suggested to include the top 50% gene pairs in terms of PCC to obtain the maximum network efficiency.

Using the absolute efficiency (F) as an estimate, the two network generated by Amira

Djebbari et al. [150] and our sub-space search method were compared. Because the reference network was inferred based on microarray data rather than a real network validated by biological experiments, we only focus the comparison on network efficiency and computational time. Due to the limitation of BN that tends to over fit the data, low F values were observed for both networks. Despite that, the standard greedy search achieved 15% absolute efficiency using 100% consumption time. The sub-space search method, on the other hand, achieved a comparable 14% absolute efficiency at a cost of only 66% computational time. In real application, users may choose to define a degree of sub-space, or may choose to include some gene pairs with lower PCC value to avoid the possible arbitrary effects of selecting only high PCC pairs.

The advantage of restricting search space can be especially useful when large scale gene expression data is applied. In classical greedy search, the number of initial change $O(n^2)$ is first calculated and each iteration step afterwards requires $O(n)$ times new calculations. In sub-space search, however, the number of initial change is $O(k_n)$, where k is decided by user-defined number of genes. Because high-throughput microarray data is often used for BN inference, and the large number of variables (e.g. tens of thousands of genes in human genome) may cause enormous increase of computational cost. By limiting the number of gene pairs, the sub-space search can achieve efficient network inference with much less computational cost with minimum tradeoffs.

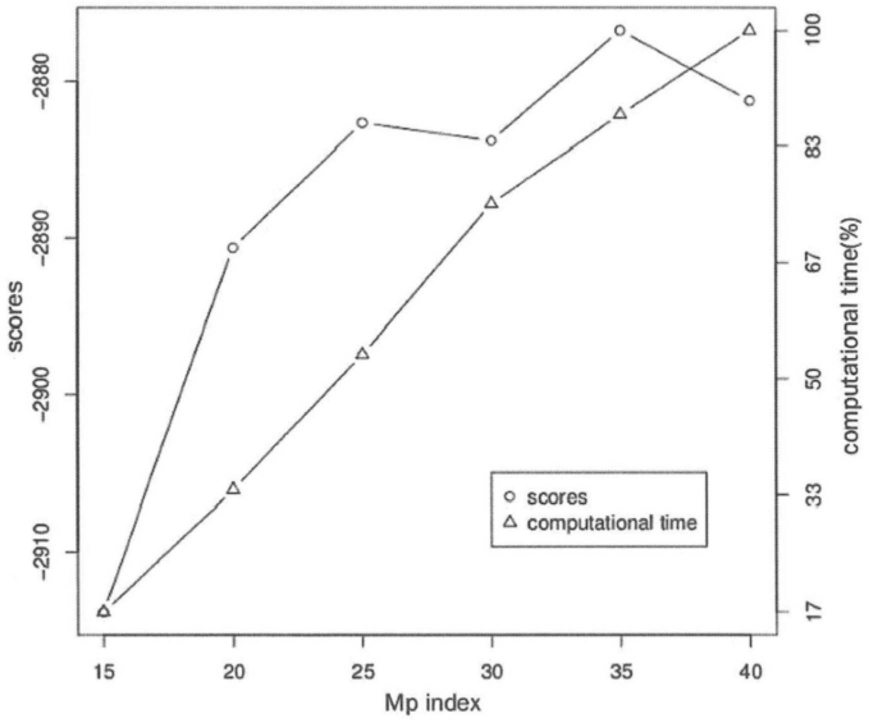


Figure 8.4: **Comparison of BN inference results using real datasets.** X-axis: M_p index representing the portion of parent genes included, e.g. the number 20 means that parent gene resides at the 1st to the 20th columns after the child gene in the M_p are included; Y-axis: BN score (left) and the percentage of computational time consumed.

Table 8.2: **Comparison of standard greedy search and sub-space greedy search using real dataset.**

M_p	Score	Absolute computational time(seconds)	Relative computational time	Precision	Sensitivity	Absolute efficiency(F)
15	2913.83	6658.1	0.4	0.08	0.15	0.11
20	2890.63	8780.16	0.52	0.09	0.16	0.12
25	2882.66	11161.03	0.66	0.11	0.2	0.14
30	2883.81	13811.01	0.82	0.12	0.21	0.15
35	2876.72	15367.59	0.91	0.13	0.25	0.17
40	2881.24	16854.16	1	0.12	0.21	0.15

8.5 Comparison to Pearson Correlation(COR) and mutual information(MI)

Mutual information(MI) has been used to narrow the parameter searching space to improve the efficiency of Bayesian network. In this section, we compared our method with other methods based on mutual information (MI) and Pearson Correlation (COR). A synthetic dataset generated from SynTReN was used as benchmarks. The inferred regulation pairs using these 3 methods were compared to the reference network. To ensure that the inferred regulatory network is independent on the initial regulation structure, we randomly assigned the initial gene pairs 100 times and calculated the number of times that any given gene pair is inferred (each pair has a score between 100 and 0). The PR-curves (Precision-Recall) under different Mp index were plotted by selecting different score values (Appendix B.1). To compare the performance of these 3 methods, the best predicted results (highest absolute efficiency) and the relative average time of each method were plotted under different Mp (Figure 8.5). Here the relative average time is the average time of 100 times divided by the maximum average time. From Figure 5, PCC and MI showed better performance and used less time than the classical method under most Mp. COR gave the worst result in terms of the absolute efficiency and the time spent. When Mp is 5, PCC achieved the highest absolute efficiency and consumed only 25% of the time compared to the classical method. It is reasonable because PCC can measure the dependence of two genes without the effect of the third gene. From this result, we can conclude that PCC is an efficient pre-processing method for limiting the search space in

Bayesian structure learning.

□ **End of chapter.**

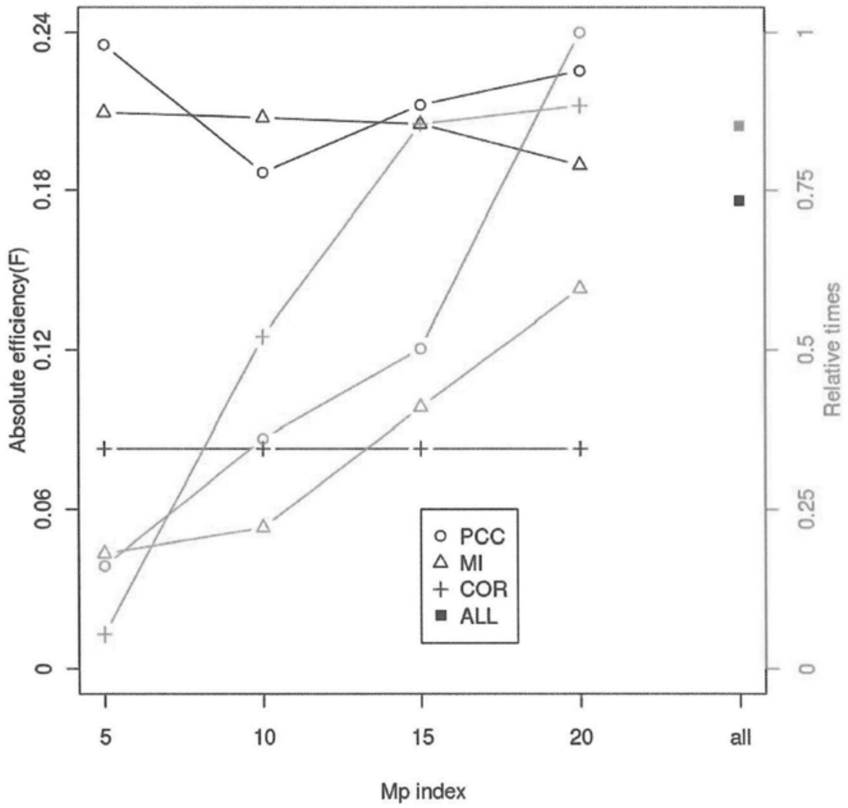


Figure 8.5: **Comparison among PCC, MI and COR.** The x-axis corresponds to the different M_p index. "All" means the all the gene pairs obtained using to the classic method. The y-axis corresponds to the absolute efficiency (black) and the relative times (red).

Chapter 9

Conclusion

Greedy search is an iteration process aiming to find a local optimizing state. During the iteration process, an added gene pair with low PCC value may affect other real pairs with higher PCCs In this part, we propose a sub-space search method to reduce the computational time while maximally retaining the BN inference accuracy. We showed that this method is feasible because BN tends to infer highly correlated gene pairs and a portion of high PCC gene pairs can be used instead of all the gene pairs. By comparing with classical greedy search algorithm using both synthetic dataset and real dataset, we demonstrated that sub-space search method can reduce nearly half of the computational time with minimum tradeoff in accuracy in BN inference. This method can be widely applied in efficient BN modeling for systems biology discovery.

□ **End of chapter.**

Appendix A

Appendix

A.1 The predictive regulatory interactions

TFs	Targets	RegulonDB	TFs	Targets	RegulonDB	TFs	Targets	RegulonDB
cspC_b1823_at	b1824_at	0	b2248_at	uidC_b1615_at	0	rhaR_b3906_at	rhaT_b3907_at	0
ymfL_b1147_at	ymfM_b1148_at	0	yagA_b0267_at	fhiA_b0229_at	0	fnr_b1334_at	b1008_at	0
ydaK_b1339_at	b1337_at	0	nac_b1988_at	narV_b1465_at	0	yneJ_b1526_at	b1592_at	0
b1747_at	b1746_at	0	purR_b1658_at	ydiJ_b1687_at	0	ydhB_b1659_at	b1410_at	0
ydaK_b1339_at	b1341_at	0	nac_b1988_at	yigP_b3834_at	0	yfeG_b2437_at	wzb_b2061_at	0
ydaK_b1339_at	ydaH_b1336_at	0	yneJ_b1526_at	dbpA_b1343_at	0	nac_b1988_at	hemD_b3804_at	0
fliA_b1922_at	fliZ_b1921_at	0	uidA_b1617_at	ydiD_b1701_at	0	b2248_at	b2246_at	0
ymfN_b1149_at	ymfM_b1148_at	0	purR_b1658_at	yciW_b1287_at	0	ydaS_b1357_at	ydaW_b1361_at	0
ydaK_b1339_at	ydaJ_b1338_at	0	csgD_b1040_at	purT_b1849_at	0	lrp_b0889_at	pntA_b1603_at	0
ydaK_b1339_at	b1342_at	0	ydaK_b1339_at	ydeZ_b1515_at	0	yneJ_b1526_at	ydjS_b1744_at	0
b1747_at	cstC_b1748_at	0	uidA_b1617_at	ydiR_b1698_at	0	ydaK_b1339_at	ydhU_b1670_at	0
fnr_b1334_at	ynaJ_b1332_at	0	ydaK_b1339_at	cheW_b1887_at	0	cbl_b1987_at	thiM_b2104_at	0
ymfL_b1147_at	ymfN_b1149_at	0	rcsB_b2217_at	yojN_b2216_at	0	uidA_b1617_at	fdnH_b1475_at	0
ymfN_b1149_at	ymfL_b1147_at	0	uidA_b1617_at	b1828_at	0	ydaS_b1357_at	narV_b1465_at	0
fliA_b1922_at	fliD_b1924_at	0	uidA_b1617_at	b1759_at	0	b1696_at	yddG_b1473_at	0
fliA_b1922_at	flgK_b1082_at	0	yneJ_b1526_at	ydhA_b1639_at	0	feaR_b1384_at	mcrA_b1159_at	0
ydaK_b1339_at	ydaL_b1340_at	0	ydaK_b1339_at	b1587_at	0	yeaM_b1790_at	yeaN_b1791_at	0
ydaK_b1339_at	dbpA_b1343_at	0	yneJ_b1526_at	b1443_at	0	yrbA_b3190_at	murA_b3189_at	0
fliA_b1922_at	flgE_b1076_at	0	b2248_at	ycjP_b1312_at	0	purR_b1658_at	aroD_b1693_at	0
ymfL_b1147_at	ymfJ_b1144_at	0	ydaK_b1339_at	b1541_at	0	nac_b1988_at	glyT_b3978_at	0
b1747_at	b1745_at	0	ydaK_b1339_at	cheY_b1882_at	0	dnaA_b3702_at	amiB_b4169_at	0
flhC_b1891_at	flhD_b1892_at	0	yheN_b3345_at	yrbI_b3198_at	0	uidA_b1617_at	ydaU_b1359_at	0
flhD_b1892_at	flhC_b1891_at	0	yfeG_b2437_at	yfdO_b2358_at	0	yheN_b3345_at	rfaD_b3619_at	0
fliA_b1922_at	flgC_b1074_at	0	fnr_b1334_at	b1593_at	0	iclR_b4018_at	spoU_b3651_at	0

fliA_b1922_at	flgN_b1070_at	0	b1978_at	ompG_b1319_at	0	celD_b1735_at	ydbD_b1407_at	0
fliA_b1922_at	flgB_b1073_at	0	purR_b1658_at	gdhA_b1761_at	0	b1978_at	b1462_at	0
yheN_b3345_at	yheM_b3344_at	0	nac_b1988_at	pphA_b1838_at	0	ydaK_b1339_at	b1592_at	0
fliA_b1922_at	fliS_b1925_at	0	uidA_b1617_at	ydjY_b1751_at	0	b1978_at	b1690_at	0
ydaK_b1339_at	ynaJ_b1332_at	0	purR_b1658_at	hisH_b2023_at	0	yneJ_b1526_at	fdnG_b1474_at	0
yneJ_b1526_at	b1640_at	0	yhjB_b3520_at	yihF_b3861_at	0	b2248_at	yebB_b1862_at	0
fnr_b1334_at	ydaA_b1333_at	0	ydaK_b1339_at	flgL_b1083_at	0	rpoN_b3202_at	yhbH_b3203_at	0
fliA_b1922_at	fliJ_b1942_at	0	uidA_b1617_at	tap_b1885_at	0	yjbK_b4046_at	ygfY_b2897_at	0
ydaK_b1339_at	ogt_b1335_at	0	ydaK_b1339_at	ynbD_b1411_at	0	ydaK_b1339_at	ydiR_b1698_at	0
fliA_b1922_at	fliK_b1943_at	0	uidA_b1617_at	rspA_b1581_at	0	ydaK_b1339_at	ydhA_b1639_at	0
ymfN_b1149_at	ymfO_b1151_at	0	ydaK_b1339_at	bisZ_b1872_at	0	b2248_at	pin_b1158_at	0
fnr_b1334_at	b1342_at	0	csgD_b1040_at	ydcF_b1414_at	0	ydaK_b1339_at	flgK_b1082_at	0
fliA_b1922_at	flgL_b1083_at	0	purR_b1658_at	thiM_b2104_at	0	ydaK_b1339_at	flgA_b1072_at	0
fliA_b1922_at	flgG_b1078_at	0	yjbK_b4046_at	rfaD_b3619_at	0	b1747_at	yneB_b1517_at	0
fliA_b1922_at	cheW_b1887_at	0	ydaK_b1339_at	flgG_b1078_at	0	uidA_b1617_at	b1202_at	0
fnr_b1334_at	ogt_b1335_at	0	csgD_b1040_at	yddG_b1473_at	0	nac_b1988_at	glnK_b0450_at	0
fliA_b1922_at	flgM_b1071_at	0	b1747_at	b1444_at	0	b1284_at	b1844_at	0
fliA_b1922_at	fliN_b1946_at	0	osmE_b1739_at	ybjP_b0865_at	0	purR_b1658_at	ynaJ_b1332_at	0
fliA_b1922_at	flgH_b1079_at	0	yneJ_b1526_at	b1828_at	0	nac_b1988_at	b1815_at	0
b1399_at	b1400_at	0	yneJ_b1526_at	narZ_b1468_at	0	ydaK_b1339_at	flgI_b1080_at	0
fliA_b1922_at	cheR_b1884_at	0	ydaK_b1339_at	ynfM_b1596_at	0	yneJ_b1526_at	maoC_b1387_at	0
b1747_at	b1488_at	0	lexA_b4043_at	rfaQ_b3632_at	0	fadR_b1187_at	btuC_b1711_at	0
ydaK_b1339_at	ydaA_b1333_at	0	ymfN_b1149_at	yeeT_b2003_at	0	dnaA_b3702_at	glmU_b3730_at	0
fliA_b1922_at	flxA_b1566_at	0	yneJ_b1526_at	narI_b1227_at	0	yneJ_b1526_at	b1746_at	0
nac_b1988_at	amtB_b0451_at	0	yheO_b3346_at	rfaD_b3619_at	0	ydaK_b1339_at	b1463_at	0
fliA_b1922_at	cheZ_b1881_at	0	yneJ_b1526_at	b1342_at	0	ydaK_b1339_at	yddA_b1496_at	0
fliA_b1922_at	flgD_b1075_at	0	b2248_at	b1314_at	0	b0373_s_at	yi22_1_b0361_s_at	0
fliA_b1922_at	cheB_b1883_at	0	yneJ_b1526_at	b1759_at	0	ydhB_b1659_at	ydiB_b1692_at	0
fliA_b1922_at	fliP_b1948_at	0	uidA_b1617_at	ydeF_b1534_at	0	purR_b1658_at	trpA_b1260_at	0
fliA_b1922_at	fliM_b1945_at	0	yneJ_b1526_at	b1487_at	0	yheN_b3345_at	yjeQ_b4161_at	0
yheN_b3345_at	yheO_b3346_at	0	nac_b1988_at	asr_b1597_at	0	ydaK_b1339_at	b1640_at	0
yheO_b3346_at	yheN_b3345_at	0	yfeG_b2437_at	b1010_at	0	yneJ_b1526_at	b1760_at	0
b1747_at	b1484_at	0	ycjC_b1299_at	b1297_at	0	uidA_b1617_at	b1746_at	0
fliA_b1922_at	fliF_b1938_at	0	yjbK_b4046_at	hemC_b3805_at	0	yneJ_b1526_at	ybiW_b0823_at	0
fnr_b1334_at	ydaK_b1339_at	0	yneJ_b1526_at	b1444_at	0	purR_b1658_at	yecS_b1918_at	0
ydaK_b1339_at	fnr_b1334_at	0	yneJ_b1526_at	ycgR_b1194_at	0	ydaK_b1339_at	b1601_at	0
uidA_b1617_at	b1433_at	0	b1747_at	b1337_at	0	yneJ_b1526_at	ynbD_b1411_at	0
fliA_b1922_at	flgJ_b1081_at	0	b1747_at	ydaK_b1339_at	0	ymfN_b1149_at	b1337_at	0
uidA_b1617_at	bisZ_b1872_at	0	ydaK_b1339_at	b1747_at	0	cbl_b1987_at	trpL_b1265_at	0
fnr_b1334_at	b1341_at	0	yneJ_b1526_at	bisZ_b1872_at	0	rhaR_b3906_at	yheI_b3331_at	0
ycjC_b1299_at	ycjL_b1298_at	0	uidA_b1617_at	b1834_at	0	b1978_at	lar_b1348_at	0
fliA_b1922_at	flgF_b1077_at	0	b1978_at	b0943_at	0	yneJ_b1526_at	b1742_at	0
fnr_b1334_at	ydaL_b1340_at	0	b1978_at	yeiC_b2166_at	0	b1978_at	yeaX_b1803_at	0
ydaK_b1339_at	b1484_at	0	nac_b1988_at	b1012_at	0	uidA_b1617_at	ybfM_b0681_at	0
fliA_b1922_at	flgA_b1072_at	0	b2248_at	b1012_at	0	ycjW_b1320_at	ynbD_b1411_at	0
b1747_at	ydjS_b1744_at	0	lexA_b4043_at	ynaJ_b1332_at	0	yneJ_b1526_at	tynA_b1386_at	0
ymfL_b1147_at	b1141_at	0	pspC_b1306_at	pspD_b1307_at	0	b1978_at	bisZ_b1872_at	0
b0373_s_at	tra5_1_b0372_s_at	0	uidA_b1617_at	yeeP_b1999_at	0	purR_b1658_at	ydeD_b1533_at	0
uidA_b1617_at	uidC_b1615_at	0	b1696_at	ydiR_b1698_at	0	b2248_at	b1759_at	0

uidA_b1617_at	uidB_b1616_at	0	yneJ_b1526_at	narJ_b1226_at	0	nac_b1988_at	eco_b2209_at	0
ymfL_b1147_at	ymfO_b1151_at	0	osmE_b1739_at	wrbA_b1004_at	0	ymfN_b1149_at	yeaW_b1802_at	0
fliA_b1922_at	motA_b1890_at	0	ydaK_b1339_at	ydcF_b1414_at	0	ydaK_b1339_at	ydiJ_b1687_at	0
fliA_b1922_at	flgI_b1080_at	0	yneJ_b1526_at	ydiR_b1698_at	0	b2248_at	ychG_b1239_at	0
fliA_b1922_at	fliL_b1944_at	0	uidA_b1617_at	b1314_at	0	yheN_b3345_at	yqiB_b3033_at	0
ymfN_b1149_at	b1141_at	0	ydaK_b1339_at	gdhA_b1761_at	0	nac_b1988_at	b1010_at	0
fnr_b1334_at	ydaJ_b1338_at	0	dnaA_b3702_at	holD_b4372_at	0	uidA_b1617_at	ydaL_b1340_at	0
fliA_b1922_at	cheA_b1888_at	0	yddM_b1477_at	chaC_b1218_at	0	purR_b1658_at	b1657_at	0
ttk_b3641_at	dfp_b3639_at	0	ydaK_b1339_at	b1488_at	0	nac_b1988_at	thrT_b3979_f_at	0
ymfN_b1149_at	ymfJ_b1144_at	0	ydaK_b1339_at	ppsA_b1702_at	0	tdcR_b3119_at	yiiG_b3896_at	0
uidA_b1617_at	lhr_b1653_at	0	fliA_b1922_at	b1760_at	0	yneJ_b1526_at	b1484_at	0
fliA_b1922_at	ycgR_b1194_at	0	ydaK_b1339_at	uidB_b1616_at	0	yneJ_b1526_at	b2387_at	0
fnr_b1334_at	ydaH_b1336_at	0	ydaK_b1339_at	leuU_b3174_at	0	yneJ_b1526_at	alkA_b2068_at	0
fliA_b1922_at	motB_b1889_at	0	ydaK_b1339_at	b1834_at	0	nac_b1988_at	b1695_at	0
b1284_at	aroD_b1693_at	0	ycjW_b1320_at	b1436_at	0	ydaK_b1339_at	ydiQ_b1697_at	0
fliA_b1922_at	fliG_b1939_at	0	uidA_b1617_at	ydeY_b1514_at	0	b2248_at	ydeF_b1534_at	0
fliA_b1922_at	cheY_b1882_at	0	b1284_at	ycjW_b1320_at	0	yheN_b3345_at	psd_b4160_at	0
fliA_b1922_at	tap_b1885_at	0	ycjW_b1320_at	b1284_at	0	ycjW_b1320_at	uidC_b1615_at	0
pspC_b1306_at	pspB_b1305_at	0	yneJ_b1526_at	b1360_at	0	ydaK_b1339_at	b1012_at	0
yhjB_b3520_at	yicK_b3659_at	0	yagA_b0267_at	ybgQ_b0718_at	0	tdcR_b3119_at	b3975_at	0
purR_b1658_at	b1686_at	0	yneJ_b1526_at	ydaL_b1340_at	0	yeiL_b2163_at	yeiC_b2166_at	0
yheO_b3346_at	yheM_b3344_at	0	yfeG_b2437_at	b1425_at	0	nac_b1988_at	yeaW_b1802_at	0
ymfL_b1147_at	ymfH_b1142_at	0	rhaR_b3906_at	yiaA_b3562_at	0	prpD_b0334_at	prpC_b0333_at	0
yneJ_b1526_at	yeiC_b2166_at	0	uidA_b1617_at	ydbS_b1393_at	0	purR_b1658_at	rnb_b1286_at	0
b0373_s_at	yi22_4_b2860_s_at	0	b1978_at	b1640_at	0	yneJ_b1526_at	yeaN_b1791_at	0
dnaA_b3702_at	rfaQ_b3632_at	0	fnr_b1334_at	bioC_b0777_at	0	b1747_at	b1011_at	0
fnr_b1334_at	dbpA_b1343_at	0	b1978_at	feaB_b1385_at	0	b1978_at	ycbE_b0933_at	0
yneJ_b1526_at	b1433_at	0	nac_b1988_at	rfaQ_b3632_at	0	ydaK_b1339_at	b1433_at	0
fliA_b1922_at	b1742_at	0	purR_b1658_at	yciH_b1282_at	0	yneJ_b1526_at	b1648_at	0
nac_b1988_at	yedL_b1932_at	0	uidA_b1617_at	b0943_at	0	yneJ_b1526_at	narV_b1465_at	0
rstA_b1608_at	rstB_b1609_at	0	ydaK_b1339_at	b1760_at	0	yneJ_b1526_at	flhA_b1879_at	0
fnr_b1334_at	b1337_at	0	iclR_b4018_at	murB_b3972_at	0	fnr_b1334_at	ydeZ_b1515_at	0
b1747_at	b1487_at	0	b1747_at	ydaJ_b1338_at	0	ycjW_b1320_at	yddG_b1473_at	0
ydaK_b1339_at	yeaW_b1802_at	0	yrbA_b3190_at	efp_b4147_at	0	slyA_b1642_at	ycgL_b1179_at	0
fliA_b1922_at	fliH_b1940_at	0	purR_b1658_at	trpD_b1263_at	0	b1696_at	pphA_b1838_at	0
uidA_b1617_at	ydbU_b1395_at	0	ycjW_b1320_at	b1731_at	0	b1747_at	b1513_at	0
uidA_b1617_at	yohG_b2138_at	0	b2248_at	b1433_at	0	celD_b1735_at	ydjE_b1769_at	0
ymfN_b1149_at	intE_b1140_at	0	uidA_b1617_at	b1640_at	0	b1696_at	celD_b1735_at	0
ymfL_b1147_at	intE_b1140_at	0	ydaK_b1339_at	yeeP_b1999_at	0	celD_b1735_at	b1696_at	0
purR_b1658_at	b1729_at	0	celD_b1735_at	ydiR_b1698_at	0	slyA_b1642_at	yciI_b1251_at	0
yidP_b3684_at	glvC_b3683_at	0	ycfX_b1119_at	cobB_b1120_at	0	uidA_b1617_at	b1410_at	0
fliA_b1922_at	fliT_b1926_at	0	uidA_b1617_at	narY_b1467_at	0	yhjB_b3520_at	yjgY_b4276_at	0
ymfN_b1149_at	ymfR_b1150_at	0	uidA_b1617_at	ynbF_b1389_at	0	ycaN_b0900_at	recE_b1350_at	0
ttk_b3641_at	dut_b3640_at	0	purR_b1658_at	thiD_b2103_at	0	tdcR_b3119_at	yhaB_b3120_at	0
b1747_at	b1008_at	0	yhcK_b3226_at	grxC_b3610_at	0	csgD_b1040_at	potG_b0855_at	0
b1747_at	b1486_at	0	ydhB_b1659_at	b1771_at	0	yfeG_b2437_at	wcaB_b2058_at	0
uidA_b1617_at	ydiF_b1694_at	0	fecI_b4293_at	fhuF_b4367_at	0	tdcR_b3119_at	yhiL_b3490_at	0
fliA_b1922_at	fliQ_b1949_at	0	uidA_b1617_at	yehP_b2121_at	0	celD_b1735_at	ydiD_b1701_at	0
yhiF_b3507_at	yhiD_b3508_at	0	yneJ_b1526_at	uidC_b1615_at	0	b1696_at	yeaW_b1802_at	0

purR_b1658_at	pyrF_b1281_at	0	nac_b1988_at	appA_b0980_at	0	feaR_b1384_at	ydaQ_b1346_at	0
fliA_b1922_at	fliO_b1947_at	0	nac_b1988_at	ydaK_b1339_at	0	yagA_b0267_at	speF_b0693_at	0
b1978_at	yneJ_b1526_at	0	ydaK_b1339_at	nac_b1988_at	0	b1978_at	feaR_b1384_at	0
yneJ_b1526_at	b1978_at	0	yfeG_b2437_at	yeaV_b1801_at	0	feaR_b1384_at	b1978_at	0
ycfX_b1119_at	ycfW_b1118_at	0	cbl_b1987_at	ydiT_b1700_at	0	yneJ_b1526_at	flhE_b1878_at	0
uidA_b1617_at	yneJ_b1526_at	0	uidA_b1617_at	cpsB_b2049_at	0	ydhB_b1659_at	b1624_at	0
yneJ_b1526_at	uidA_b1617_at	0	uidA_b1617_at	b1648_at	0	yfeG_b2437_at	b0872_at	0
fliA_b1922_at	b1904_at	0	csgD_b1040_at	b1668_at	0	ydaK_b1339_at	ydjS_b1744_at	0
yneJ_b1526_at	b1525_at	0	uidA_b1617_at	narI_b1227_at	0	b2248_at	b2100_at	0
uidA_b1617_at	yeaW_b1802_at	0	ydaK_b1339_at	wcaG_b2052_at	0	cbl_b1987_at	ydjC_b1733_at	0
yneJ_b1526_at	yeaL_b1789_at	0	dnaA_b3702_at	gmk_b3648_at	0	ydaK_b1339_at	asr_b1597_at	0
fliA_b1922_at	flhB_b1880_at	0	nac_b1988_at	b1484_at	0	uidA_b1617_at	ynbD_b1411_at	0
ydaK_b1339_at	b1485_at	0	yneJ_b1526_at	ydeZ_b1515_at	0	purR_b1658_at	yecP_b1871_at	0
uidA_b1617_at	b1397_at	0	fnr_b1334_at	flhA_b1879_at	0	csgD_b1040_at	gdhA_b1761_at	0
ymfN_b1149_at	ymfH_b1142_at	0	b1978_at	yeaW_b1802_at	0	uidA_b1617_at	cheB_b1883_at	0
b1747_at	b1483_at	0	yijC_b3963_at	yijD_b3964_at	0	lrp_b0889_at	pntB_b1602_at	0
fliA_b1922_at	flhA_b1879_at	0	yheO_b3346_at	dam_b3387_at	0	purR_b1658_at	b1180_at	0
yjbK_b4046_at	yiaF_b3554_at	0	yrbA_b3190_at	prfB_b2891_at	0	uidA_b1617_at	ydbP_b1390_at	0
b1284_at	b1688_at	0	csgD_b1040_at	trpL_b1265_at	0	nac_b1988_at	yddG_b1473_at	0
ydaK_b1339_at	b1487_at	0	yheO_b3346_at	yhbN_b3200_at	0	uidA_b1617_at	b1547_at	0
rhaR_b3906_at	yijF_b3944_at	0	uidA_b1617_at	b1443_at	0	nac_b1988_at	bioC_b0777_at	0
uidA_b1617_at	b1360_at	0	yneJ_b1526_at	b1690_at	0	b1696_at	ydaU_b1359_at	0
yhjB_b3520_at	yiiG_b3896_at	0	fliA_b1922_at	b1044_at	0	yjbK_b4046_at	yigZ_b3848_at	0
b2248_at	wzb_b2061_at	0	nac_b1988_at	trpE_b1264_at	0	ydaK_b1339_at	ycbF_b0944_at	0
purR_b1658_at	aroH_b1704_at	0	ydaK_b1339_at	bioA_b0774_at	0	cpxR_b3912_at	yihZ_b3887_at	0
yneJ_b1526_at	uidB_b1616_at	0	b1978_at	b1489_at	0	fnr_b1334_at	ppsA_b1702_at	0
uidA_b1617_at	maoC_b1387_at	0	nac_b1988_at	b1341_at	0	purR_b1658_at	b1624_at	0
b2248_at	yohG_b2138_at	0	sohA_b3129_at	yhaV_b3130_at	0	uidA_b1617_at	b1394_at	0
yneJ_b1526_at	b1486_at	0	ydaK_b1339_at	b1436_at	0	ydaK_b1339_at	cheZ_b1881_at	0
b1747_at	b1442_at	0	b1747_at	ycdG_b1006_at	0	yfeG_b2437_at	wza_b2062_at	0
uidA_b1617_at	ompG_b1319_at	0	uidA_b1617_at	b1543_at	0	uidA_b1617_at	ydaT_b1358_at	0
fliA_b1922_at	fliE_b1937_at	0	csgD_b1040_at	trpA_b1260_at	0	csgD_b1040_at	chaC_b1218_at	0
yneJ_b1526_at	ydjZ_b1752_at	0	b2248_at	b2099_at	0	b1284_at	b1624_at	0
b1747_at	b1441_at	0	uidA_b1617_at	narG_b1224_at	0	fnr_b1334_at	yecH_b1906_at	0
purR_b1658_at	purT_b1849_at	0	csgD_b1040_at	trpC_b1262_at	0	csgD_b1040_at	thiM_b2104_at	0
ydaK_b1339_at	b1442_at	0	ydaS_b1357_at	b1327_at	0	yneJ_b1526_at	b1543_at	0
ydaK_b1339_at	narV_b1465_at	0	purR_b1658_at	trpC_b1262_at	0	ydaK_b1339_at	narU_b1469_at	0
uidA_b1617_at	yecK_b1873_at	0	b2248_at	bisZ_b1872_at	0	uidA_b1617_at	ydiJ_b1687_at	0
yneJ_b1526_at	ydjY_b1751_at	0	ycjW_b1320_at	yeaW_b1802_at	0	yhjB_b3520_at	yiaA_b3562_at	0
ydaK_b1339_at	narH_b1225_at	0	yneJ_b1526_at	b1463_at	0	rhaS_b3905_at	yijF_b3944_at	0
yneJ_b1526_at	ompG_b1319_at	0	ycjW_b1320_at	b1771_at	0	b1284_at	gdhA_b1761_at	0
ydaK_b1339_at	b1483_at	0	feaR_b1384_at	b1345_at	0	ydaK_b1339_at	flhE_b1878_at	0
ydaK_b1339_at	wza_b2062_at	0	ydaK_b1339_at	rspA_b1581_at	0	ydaS_b1357_at	b1436_at	0
csgD_b1040_at	trpE_b1264_at	0	feaR_b1384_at	b1012_at	0	ymfN_b1149_at	ydaH_b1336_at	0
uidA_b1617_at	yeiC_b2166_at	0	frvR_b3897_at	yigE_b3815_at	0	nac_b1988_at	ycbF_b0944_at	0
b2248_at	yeaN_b1791_at	0	yneJ_b1526_at	b1012_at	0	uidA_b1617_at	ydiB_b1692_at	0
ydaK_b1339_at	ydiB_b1692_at	0	ycaN_b0900_at	ycbO_b0936_at	0	uidA_b1617_at	ydaK_b1339_at	0
fliA_b1922_at	flhE_b1878_at	0	b1978_at	uidA_b1617_at	0	ydaK_b1339_at	uidA_b1617_at	0
ydaK_b1339_at	narG_b1224_at	0	uidA_b1617_at	b1978_at	0	nac_b1988_at	yigR_b3835_at	0

yneJ_b1526_at	ydaH_b1336_at	0	yneJ_b1526_at	flgM_b1071_at	0	b1696_at	ydhU_b1670_at	0
yneJ_b1526_at	feaB_b1385_at	0	uidA_b1617_at	wcaJ_b2047_at	0	uidA_b1617_at	narW_b1466_at	0
yneJ_b1526_at	lhr_b1653_at	0	nac_b1988_at	b1488_at	0	uidA_b1617_at	ydjE_b1769_at	0
b2248_at	uidA_b1617_at	0	uidA_b1617_at	recT_b1349_at	0	yheN_b3345_at	dapF_b3809_at	0
uidA_b1617_at	b2248_at	0	uidA_b1617_at	yeaX_b1803_at	0	yneJ_b1526_at	gdhA_b1761_at	0
ydaK_b1339_at	b1008_at	0	b1978_at	maoC_b1387_at	0	uidA_b1617_at	b2387_at	0
nac_b1988_at	trpL_b1265_at	0	flhC_b1891_at	flgL_b1083_at	0	yheO_b3346_at	grxC_b3610_at	0
ydaK_b1339_at	b1746_at	0	b1284_at	yddG_b1473_at	0	yfeG_b2437_at	ycbF_b0944_at	0
fliA_b1922_at	flhC_b1891_at	0	uidA_b1617_at	trpE_b1264_at	0	ydaK_b1339_at	relF_b1562_at	0
uidA_b1617_at	b1398_at	0	b1978_at	uidB_b1616_at	0	yneJ_b1526_at	yddA_b1496_at	0
b1978_at	uidC_b1615_at	0	ydaK_b1339_at	motB_b1889_at	0	csgD_b1040_at	bioC_b0777_at	0
cbl_b1987_at	nac_b1988_at	0	ydaK_b1339_at	yeaV_b1801_at	0	fis_b3261_at	yhdG_b3260_at	1
nac_b1988_at	cbl_b1987_at	0	b2248_at	lhr_b1653_at	0	gutM_b2706_at	srlD_b2705_at	1
fliA_b1922_at	tar_b1886_at	0	uidA_b1617_at	b1513_at	0	betI_b0313_at	betB_b0312_at	1
gcvR_b2479_at	bcp_b2480_at	0	nac_b1988_at	ydaS_b1357_at	0	csgD_b1040_at	csgE_b1039_at	1
yneJ_b1526_at	ycbN_b0935_at	0	ydaS_b1357_at	nac_b1988_at	0	yhiE_b3512_at	hdeA_b3510_at	1
yrbA_b3190_at	aroK_b3390_at	0	yneJ_b1526_at	trpE_b1264_at	0	yhiE_b3512_at	hdeB_b3509_at	1
yneJ_b1526_at	b1442_at	0	fliA_b1922_at	fliC_b1923_at	0	csgD_b1040_at	csgF_b1038_at	1
b1747_at	b1009_at	0	uidA_b1617_at	b1444_at	0	yhiE_b3512_at	slp_b3506_at	1
purR_b1658_at	pntB_b1602_at	0	uidA_b1617_at	pin_b1158_at	0	gutM_b2706_at	srlE_b2703_at	1
uidA_b1617_at	feaB_b1385_at	0	yneJ_b1526_at	flgF_b1077_at	0	yhiE_b3512_at	hdeD_b3511_at	1
yneJ_b1526_at	yeaW_b1802_at	0	uidA_b1617_at	b1592_at	0	csgD_b1040_at	csgG_b1037_at	1
b1747_at	b1485_at	0	nac_b1988_at	ydiT_b1700_at	0	dnaA_b3702_at	dnaN_b3701_at	1
csgD_b1040_at	trpD_b1263_at	0	b2248_at	yfdN_b2357_at	0	gutM_b2706_at	srlB_b2704_at	1
ydaK_b1339_at	b1009_at	0	b1747_at	wcaI_b2050_at	0	lexA_b4043_at	dinF_b4044_at	1
b1747_at	b1443_at	0	b1696_at	gdhA_b1761_at	0	yhiW_b3515_at	yhiX_b3516_at	1
yneJ_b1526_at	b1009_at	0	b1747_at	trpE_b1264_at	0	yhiX_b3516_at	yhiW_b3515_at	1
ydaK_b1339_at	lhr_b1653_at	0	celD_b1735_at	b1509_at	0	gutM_b2706_at	srlA_b2702_at	1
yneJ_b1526_at	b1565_at	0	uidA_b1617_at	b1310_at	0	flhC_b1891_at	flgE_b1076_at	1
b1696_at	ydiQ_b1697_at	0	yneJ_b1526_at	b1008_at	0	gatR_2_b2090_f_at	gatD_b2091_at	1
yneJ_b1526_at	b1371_at	0	yneJ_b1526_at	flgA_b1072_at	0	flhC_b1891_at	flgC_b1074_at	1
b0373_s_at	yi21_5_b3044_s_at	0	ydhB_b1659_at	b1543_at	0	lexA_b4043_at	recA_b2699_at	1
b1696_at	b1543_at	0	rhaR_b3906_at	yiaB_b3563_at	0	b2531_at	yfhO_b2530_at	1
ydaK_b1339_at	b1191_at	0	ydaK_b1339_at	bioC_b0777_at	0	dnaA_b3702_at	recF_b3700_at	1
yneJ_b1526_at	ynjA_b1753_at	0	yneJ_b1526_at	yohG_b2138_at	0	flhC_b1891_at	flgG_b1078_at	1
yhiF_b3507_at	yhiU_b3513_at	0	yneJ_b1526_at	b1341_at	0	flhC_b1891_at	flgB_b1073_at	1
ydaK_b1339_at	flhA_b1879_at	0	nac_b1988_at	b1428_at	0	yhiE_b3512_at	gadA_b3517_at	1
b1284_at	sppA_b1766_at	0	ydaK_b1339_at	yedL_b1932_at	0	flhC_b1891_at	flgH_b1079_at	1
uidA_b1617_at	narZ_b1468_at	0	yheO_b3346_at	yrbH_b3197_at	0	flhC_b1891_at	flgI_b1080_at	1
nac_b1988_at	ydiQ_b1697_at	0	b1978_at	b1345_at	0	flhC_b1891_at	flgF_b1077_at	1
uidA_b1617_at	alkA_b2068_at	0	b1696_at	lhr_b1653_at	0	yhiE_b3512_at	gadB_b1493_at	1
b2248_at	b1310_at	0	b1696_at	b1771_at	0	lexA_b4043_at	yebG_b1848_at	1
uidA_b1617_at	fdnG_b1474_at	0	uidA_b1617_at	ychG_b1239_at	0	fnr_b1334_at	narG_b1224_at	1
ydaK_b1339_at	b1085_at	0	b1696_at	b1444_at	0	yhiE_b3512_at	yhiD_b3508_at	1
ydaK_b1339_at	ydaO_b1344_at	0	nac_b1988_at	potG_b0855_at	0	flhC_b1891_at	flgM_b1071_at	1
ydaK_b1339_at	flxA_b1566_at	0	ydaK_b1339_at	b1443_at	0	flhC_b1891_at	flgD_b1075_at	1
celD_b1735_at	b1543_at	0	csgD_b1040_at	aroD_b1693_at	0	lexA_b4043_at	dinD_b3645_at	1
uidA_b1617_at	b1371_at	0	b2248_at	ompG_b1319_at	0	lexA_b4043_at	recN_b2616_at	1
b1747_at	ynfM_b1596_at	0	yggD_b2929_at	yrbH_b3197_at	0	flhC_b1891_at	fliA_b1922_at	1

b1747_at	b1012_at	0	rpoE_b2573_at	rseA_b2572_at	0	lldR_b3604_at	lldP_b3603_at	1
yhiF_b3507_at	slp_b3506_at	0	b2248_at	feaB_b1385_at	0	lexA_b4043_at	sulA_b0958_at	1
nac_b1988_at	thiM_b2104_at	0	yneJ_b1526_at	b1485_at	0	tdcA_b3118_at	tdcB_b3117_at	1
yneJ_b1526_at	pspD_b1307_at	0	ycjW_b1320_at	b1624_at	0	flhC_b1891_at	flgA_b1072_at	1
yneJ_b1526_at	b1345_at	0	ydaK_b1339_at	b1759_at	0	rhaS_b3905_at	rhaB_b3904_at	1
b0373_s_at	yi22_2_b1402_s_at	0	ydaK_b1339_at	flgM_b1071_at	0	flhC_b1891_at	fliM_b1945_at	1
yagA_b0267_at	ylcD_b0574_at	0	uidA_b1617_at	flhE_b1878_at	0	flhC_b1891_at	fliN_b1946_at	1
flhC_b1891_at	flgK_b1082_at	0	b1696_at	b1360_at	0	rhaR_b3906_at	rhaS_b3905_at	1
uidA_b1617_at	b1392_at	0	b2248_at	uidB_b1616_at	0	rhaS_b3905_at	rhaR_b3906_at	1
uidA_b1617_at	ycjV_b1318_at	0	b0373_s_at	tra5_3_b1026_s_at	0	b2531_at	b2529_at	1
ydaK_b1339_at	cheR_b1884_at	0	uidA_b1617_at	tynA_b1386_at	0	fnr_b1334_at	narJ_b1226_at	1
b2248_at	b1828_at	0	uidA_b1617_at	ydaJ_b1338_at	0	lexA_b4043_at	dinI_b1061_at	1
ydaK_b1339_at	yohG_b2138_at	0	b2248_at	ynbF_b1389_at	0	flhC_b1891_at	flgN_b1070_at	1
ydaK_b1339_at	b1486_at	0	ycjW_b1320_at	ydhO_b1655_at	0	gatR_2_b2090_f_at	gatC_b2092_at	1
ydaK_b1339_at	b1444_at	0	uidA_b1617_at	yddA_b1496_at	0	lexA_b4043_at	oraA_b2698_at	1
osmE_b1739_at	elaB_b2266_at	0	b1747_at	ydeY_b1514_at	0	yhiX_b3516_at	gadA_b3517_at	1
fliA_b1922_at	fliI_b1941_at	0	ycaN_b0900_at	ycaK_b0901_at	0	flhC_b1891_at	b1904_at	1
nac_b1988_at	thiD_b2103_at	0	feaR_b1384_at	b1423_at	0	fnr_b1334_at	narI_b1227_at	1
yneJ_b1526_at	b1489_at	0	nac_b1988_at	ynaJ_b1332_at	0	flhC_b1891_at	flgJ_b1081_at	1
b2248_at	maoC_b1387_at	0	uidA_b1617_at	b0941_at	0	flhC_b1891_at	fliL_b1944_at	1
yneJ_b1526_at	flxA_b1566_at	0	uidA_b1617_at	narV_b1465_at	0	nac_b1988_at	gdhA_b1761_at	1
uidA_b1617_at	b1345_at	0	yneJ_b1526_at	narG_b1224_at	0	purR_b1658_at	pyrD_b0945_at	1
ydaK_b1339_at	b1686_at	0	ydaK_b1339_at	yneJ_b1526_at	0	flhD_b1892_at	flgB_b1073_at	1
uidA_b1617_at	ydjS_b1744_at	0	yneJ_b1526_at	ydaK_b1339_at	0	flhC_b1891_at	flhA_b1879_at	1
ydaK_b1339_at	trpE_b1264_at	0	nac_b1988_at	b1686_at	0	fnr_b1334_at	narH_b1225_at	1
nac_b1988_at	b1454_at	0	b1696_at	b1648_at	0	feaR_b1384_at	feaB_b1385_at	1
b1747_at	b1440_at	0	yneJ_b1526_at	ydaJ_b1338_at	0	fnr_b1334_at	fdnG_b1474_at	1

A.2 The 150bp promoter sequences

b0373_s_at:

>insEF-2E-2F-2

GATTTAACTAACGGTTCTTCGCTGAACATTGGTGAAGATGGCTACGTTGATACCGATCATCTGACTATTAACTCCTACAGTACTGTTGCGTTGACCGAATCTACTGGGTGGGGGGCTGATTGATCCTACCCACGTAATATGGACACAGGCC

>insCD-4C-4D-4-ygeONM

AATCCAAAGAATACATTGATGAAATAATAATGAAATATAATTAAAAATAAAATTTTTGCGTAAAAAAATACCACAGGCATTAAAAAATCATGAGATGATTAAAATATTACAATTAGATTATATTCAAATCATTAAACTTGAGCCAGGGAGC

>insC-5CD-5D-5-yqiGHI

ACATCATTAACGGATTAATGATAAGTGGATCAGATGTATAAAAAATTAAAATTAACCACAATAAGCGAATTGATTAAAAATATTTATTGTTCATTATCCGTTATTAGACTGGCCCCCTGAATCTCCAGACAACCAATATCACTTAAATAAG

>insCD-2C-2D-2

ACGACAATCAGCGGAGAATTCCCTCTGGGGTTTGCCGGGGTTATTCGGGTACAGGATAAAGCTTTGCTGGAAATTGGCAGTGGCGCTACGCTAACAATGCAGGATATTGACAGTTTTGAACATCATGGGACAAGAACCCTGGATTTGCCCC

>insEF-4E-4F-4

ACCCAAATGAGATCTATCGACTGGCTGATAACGCACTTTACGAGGCGAAAGAGACCGGGCGTAATAAGGTGGTTGTGAGGGATGTGGTGAATTTTTGTGAGTCACCATAAAGCGGCATTTTGATCCTACCCACGTAATATGGACACAGGCC

>insC-1CD-1D-1

AATCGTTGCTTCCGCTTCAATAACCGCATTTTTAATCTCGGTGGCAAAACCAATTTTTACGCCATTGCTGATTATTGTGCCAGGACGAATAAACGCATAATTACTAGACTGGCCCCCTGAATCTCCAGACAACCAATATCACTTAAATAAG

b1284_at:

>ydiNB-aroD

CAAAAACAAACTATGACATGCAATATTCCTGGAAACATAAACTTTATGCCATGTACCCAGGGAAAATCATCTTCAGTATAGTAATTATGTAAACCGTCGGAGAACAATACGTACGGTAACGAAATTATCTTTCAGCAAGGAGCTGTGAAAA

>ydiK

TGTACCACATTTTTTTCTAACACGCCCATCAGAATTAAGGGCAGAATCGGCCTGTTAAAAACCGCTGAAATTGCTCATCATTATGCAGGTGAGTTTCGCGTGTTCACGTCGCGTCGACGATTTGACGCACAAAAAAGGTGAAAAGTAGTTA

>sppA

CCTGTCGTGATATTTATTCACAAAATTAACACGAGAGTGGATTTTGTTACAGCACAGTCCGCAATTCCTGCTGACAAGTACCGGTTGGGTCATTACGATAACCACATCTATTGCGCCTGTGACAGGTGTGACCTTAAGTTGGGAGAATACA

>ycjW

ATTCCATTATGCAGGTGTCGGCGTAAATTACTCGTTCTGATAATGGGCTAAATTGCCGGATGCGGCGCGAGTACTTTATCCGATCTATAAATGTAGGCCGGATAAGATGCGCTAGCATCGCATCTGGCATTCAGGCAAGGTAGCTGGTATT

>yddG

TGAGAAAAATAGATGAAATAATATTATTTATCGATATGTGATCGAAGTCGAAATGAGATATAAGGTGAATTACTGGTATTTGAAATTTATTTTTTTAATATTGTCGGAATTTATCTGATTAACTACCGGGCCGTAGACCCGGCAGTTATTT

>yobB-exoX

GGCTTCGGTCAATCTGTCACGTTTACCTTACGAGCCCAAACTTAAATAAAACTTATACAGAGTTACACTTTCTTACATAACGCCTGCTAAATTATGAGTATTTTCTAAACCGCACTCATAATTTGCAGTCATTTTGAAAAGGAAGTCATTA

b1696_at:

>ydiQRST-fadK

TTGTTTTAGCGATCGCATTTTTTTTGCAAGATTGTTGGCAAGGAAAACAGCTTGCTCCGTCGAAAACCCCGCACCGCTATCGCACACTATTTTCAGGCCATTTTTACCTTCCATCGGAGATGGTTCCGTATGCGACTCACAGGAGAAATCA

>ydfJ

GCCAGGTTATTAAGATCATAAACAGGGAGTGTCGCTTTTGCTGATAACAAATTATTTCCCATAACAATTCCTTAAATATAAATATGGCAAGCTATATGTTTTGTTATATGAATAAAAATCCCCTCTCCGGTAAGAGAAGGGATTAAGGGTT

>gdhA

TATGAGATTACTCTCGTTATTAATTTGCTTTCCTGGGTCATTTTTTTCTTGCTTACCGTCACATTCTTGATGGTATAGTCGAAAACTGCAAAAGCACATGACATAAACAACATAAGCACAATCGTATTAATATATAAGGGTTTTATATCTA

>rnt-lhr

ATTGAGTTAATCGAAGAGAAAGACGCCGGTCGCGGTCTGGGCAACTAATCTCCTGCCGGGCGTGAACTCATCGCGCCCGCATCTTTACTGCATCGACAAGTAATATTTGTCATAATGCGCGCTGCAATTTATCCGTATTAAGAGAATCAGA

>ydjLKJIHG

TCTTTTGCCGCCACGATTACACCCCTGTATCTTTTTACATCACATTAGCGCGATTATCGCATAACCGATGTTTACTTTCAAAATAACCTGTTTGAATCACAGATTTTCATCACAGTTTTCACAGAAACAGAGGTGAATCGTGTTGAGTATT

>ydcW

CTCAATCAGCTTGGGCGACCGCGTGATGTACCGGTAACTAACGTGGTGGCACTGCTGGTTATGTTGGTAACAACCTTGCCGATCCTGGGGGCCTGGTGGCTAACCCGCGAAGGCGACAATGGTCAATAACCACTGATACAGGAATATGCTA

>ydaSTUVW-rzpR

TCGGCCTAAGTCTTTACCACTAAGCATTGCTTAATATTCTCCTATGCGCATTACATTAGGCAATCCCTACCCTTACTGCATTAGGCACAGCCTATTGACAATTGCGTTAGGCGTCGCCTAATATTTCTGTGTGTTTTTGGAGTTCATTCGA

>ydhL

CAGCTGGCGGGCGGACATATTCCAGTCCATCAATCGCCAGTAGCCCATCACAAAACGGGAAAACTCCGGGCCTTGCGGCGCAATAGTAATACGCTGAACCATAATCGCTTCCTCTTATCAGATATGAGAGGAGTATACGCAAGATTAGGTT

>yddG

TGAGAAAAATAGATGAAATAATATTATTTATCGATATGTGATCGAAGTCGAAATGAGATATAAGGTGAATTACTGGTATTTGAAATTTATTTTTTTAATATTGTCGGAATTTATCTGATTAACTACCGGGCCGTAGACCCGGCAGTTATTT

>pphA

TGTGAAGATATTGTTGTGGATGTCTACAACACGGAACAGCAGTGTCTTTATTCTATGAGCGATCAACGGATCCGCCAGGGCGGTTGTTTTCCGATTGAGGATTTTATAGATGGTTTCTGGCGACCTGCACAGGAGTACGGTGATTTTTAAT

>chbBCARFG

ATATATTGCAGGAAACACGTAGGCCTGATAAGCGAAGCGCATCAGGCAGTTTTGCGTTTGTCAGCAGTCTCAAGCGGCGCAGTTACGCCGCCTTTGTAGGAATTAATCGCCGGATGCAAGGTTCACGCCGCATCTGGCAAACATCCTCACT

>yeaVW

TTTTTATTCGTTCTTTGCAGTAAATAACCTGCGTCATTTCACCTTTTATTGTTTCCGTTTCGTGTTTTATGGCTTTCCGTATTCTTAATTGTTTAATTTATGTAACATGCAAATTTTGTTACGCGTACGTTAGGTTCCGCCGTACAGGTAT

b1747_at:

>astCADBE

GAGAGAATATTACAACACGATGATTTTGCAGAGATTATGAAGAACTATACCGGATGACTGGTGATAAATAAAGCAAATAACCAGGATTAATCTGTATTAATTTATAAGAAAGCAACTTAATACCCGCAGAATGATTTCTGCGGGTAAGTAT

>ddpXABCDF

CCTTTGACGGCATAATCCAGAAAGCAAAAGCAGGATGCCCGGTCTCTCAGGTACTGAAAGCGGAAATTACGCTGGATTACCAGTTGAAATCGTAAAGCATTGCCGGATGACGCGTCAGGCGCGTGAATGCCTGATGCGTTGTTAGCATCTC

>rutABCDEFG

GCGCGATGCTTATCAGGCCTACCAGAAGATTGCAATATATTGAATTTGCACTGTTTTGTAGGCCGGATAAGGCGTTTACGCCGCATCCGGCATGAACAATGCGTACGTTGTCAACAATCTGCACCGCCGGTAACCCCGGCGGTTTTCTGTT

>ydcSTUV

TTACAGACGGCATAATGCGCGGTAGCTCACAACCTGAATAAATTTTCTCAGGGGCGAAGGTGTGCCTGCAAGCCGCCGTCTATGGTTAAACAAGGAGATATTTTTACGGCACGGCGGCTGAACAATTAATTACGACAGGAGTAAGACCTTA

>ynfM

GCAACAGCAACAAAGTAACGCAGATGACGAAGTTCAATATTCATATTTAAAACATCTTATTTGAGATTATTAATATATTAGACAGAACAATTCGATTTTCCTACCCTATGTATAAGCCTGATCTACAGGCATATTTAGCAAGGATTTCAAG

>ydcW

CTCAATCAGCTTGGGCGACCGCGTGATGTACCGGTAACTAACGTGGTGGCACTGCTGGTTATGTTGGTAACAACCTTGCCGATCCTGGGGGCCTGGTGGCTAACCCGCGAAGGCGACAATGGTCAATAACCACTGATACAGGAATATGCTA

>abgABT-ogt

TTTTAGGGAGCAAGTAAGTCTAAGCAAACCTTAACAGCAGAGAATTCCGATATTAGATGTAAATATATGTCTATCTATTTGAAAACCCTTAAGTTGTTAAGGGTAACTTTACATAAAAGTGTGAACAAGCTGGCACAAATTGTTTAATGTT

>abgR

TGTGACCCGGTTCACGTAGCGATAGTTTTTACTTATCACTAACTGATTTTTCACAGTTTTAACCGTTCATAAATTACCCTGACACAATCATCTGCATTAAAGTAGATGCCAGTTTCTTTGGTCTGATAAATAACGGTTATCGGTGGCGTCA

>wcaCDEF-gmd-fcl-gmm-wcaI-cpsBG-wcaJ-wzxC

ATAAGGCGTTCACGCCGCATCCGGCATTCAGTGCCTGATGCGACGCTGACGCGTCTTATCAGGCCTACAGGTCCCGAGCACAGAACCGTAGGACGGATAAGGCGTTTTACGCCGCATCCGGCAACCGTTGTCGGAACCGAAAACAGCAACT

>trpLEDCBA

TATATTACTGTTGGGCGGAAAAATGACGTAAGTTGACGTTCGACCGGGGTAAGCGAAACGGTAAAAAGATAAATATTAAATGAATTTAGGATTTTTCCGGCTTCATTAAAGAAAGTTAAAATGCCGCCAGCGGAACTGGCGGCTGTGGGAT

>lsrACDBFG-tam

GCATGAACAAACGCAACCGTGAAAATCAAAATAGCATAAATTGTGATCTATTCGTCGGAAATATGTGCAATGTCCACCTAAGGTTATGAACAAATTAAAAGCAGAAATACATTTGTTCAAAACTCACCTGCAAAACTGAACGGGGGAAATA

b1978_at:

>yneJ

CCCGTGGCAGGATTTATCGAAATTGCATGAGTTGCCGGAGTAATGGTCATCGGGGTATCTCCTTTATGAGTCATGGTATGAAGATACGCAGATTTACTCTTGCTTTAAAATGAATAATATTAAGCCACTTATTCACGAATCGAGAATGCTA

>uidABC

TACGCTTATCAGGCCTACCTTTTCCTGCAACACTTTGAATTTATGAGTTTTTGTAGGCTGGATAAGGCGTTTACGCCGCATCCGGCATAAAAAACGCGCACTTTGTCAACAATCTGAAACGCCGGAGATTTTTCTCTCCGGCGTTATTTTT

>ompG

TGCTGGTGCGTTAAATGATTATCATGCAGGAGAAAATATCACTATTCATTTTGATATGACGAAATGTCATTTCTTTGATGCAGAAACGGAAATAGCAATTCGCTAAATACAGGGGGAAGGCATTCCCCCAGGATAATACAAGGAACAATAA

>ycbRSTUVF

TCGCTACGGGGAAAGCGACTTCAGGTAATGTGAATGCGGTAACAAATTTCCATATTAACTATTATTAATAGAACTCATTAATTGTTTTATTAATTAGTACCCCTCCAGTGTTCTGGAGGGGATATTCATATTTTTTAAGAGTGACTATTTA

>pscKG

CAGAAATGCTATTGCCAGTAACACCACCATCCCCACAACACTTCTCATTATATCCATAATGATTTTCCCTTCATGCCGGTAAACCCGGCGTCAGCGCCAGGTTTTGGTATGCTTGATGAGTACGGGCGACGGCTTTCTGCCCGTCAGAAAT

>anmK

CAGCCGGAGAGAAGGACCGGCAAAATGATAATTAACAGTTTTTCATAGTCATATCCCGAAGACTTTCCTGGTCTGGAGGGCAATACGCCCTCCCTAACGTTCCAAGTGTAACGGCAGACGCGGTAAGAAAAATTCAGTTAACTCTGATAT

>feaB

AAGTCTTGTCAGGCATAGAGACATAAGCGGTTATTGTCACGATTTGCGGAGCTTGTCACAGCTGACAAAGCGAATGTCACAGCGAAAAAAGTGACTTTTCTTGTCGCTGCGTACACTGAAATCACACTGGGTAAATAATAAGGAAAAGTGA

>yeaVW

TTTTTATTCGTTCTTTGCAGTAAATAACCTGCGTCATTTCACCTTTTATTGTTTCCGTTTCGTGTTTTATGGCTTTCCGTATTCTTAATTGTTTAATTTATGTAACATGCAAATTTTGTTACGCGTACGTTAGGTTCCGCCGTACAGGTAT

>dosCP

TTTTGATCAACATTTGTGCAGCGTAGTGCAGTTTTGGTGCAAGAGGGGAAGTTAAGGAAGGAATCTCCCGGAATCGTAGCTGAAATCACAGTATTTAAGTGACAGTGTCACGTTAAATGAAAACCCGCGAGTGCGGGCGAGAGGAATTTGT

>maoC

CTGCACCTGCCAGGTTGTTTGGCAGGTGTGCCAGCTTTTCATACAGTGGATGCCCTGAAAATAGATGTACACATCATGCATAATGTGACAACGTCACAAAACTTAGTGAAATAAAAGGGCAACTATTCGCCGTTGCCCTTCATTCACCGAT

>recET-lar-ydaCQ-intR

GAAGTCAGCAATGGCTTCGCCCACGTTACGACGCAGACGTTTTTGTAATTTGTTCAGGTTGTATTGTTCTTTCTTTGTAATTTGTTGATTTTCTTGCATTATTTCAGTTCTCTGGTACTAAATGGGGCAAATTGGGGGCAAACTTTGCAAC

>yddH

CGTTATTATTCGTCATCTGAACAGTAAAGACAGTTCGATAAGCATTGCTCTACAGCAGATTCAACCAGAATCTTGGCAAGCTATCTGGGATGCCGAAATCGCGCCCCAAATGGAGGCTTTGATAAAGAAACCTGGTTATAGCATGAATGCT

>ydiM

CAATCGGTAGCTGCTGAATTATATGCACACGATCTGGAGCGGCTTTGTTAATTTTTCCACAGAAAGGAATTGTCGTTGTTACAACAATAATGAACGGATGCTGACACAACATCGCTTCACTTTTTAAAGCACCTTTGCTAAGTAGAACCTA

>yeaX

AGGGGCGCATCATGGCCGACAGTAGCGGTAGTGGCATTTCCGAACATGGTATCGCCCATTTCCATAATCTGCTGGCGCAGGTGTTTAAGGACTAATGACATCGGCGGCGGTATTTTCCGCCGCTGGGCTGATTTTTGATGGAGTACAGCAA

>torYZ

TGCCGGATGCGGCGTGAACGCCTTATCCGGCCTACAAAACCTTGCTAATTCAATATATTGCAGGGACTATGTAGGCCTGATAAGCATAGCGCATCAGGCAGCTTTACGTTTGCATAACCTCAGCGCCCGTTTCCGGGCGCTATTCACGTCT

>ssuEADCB

TACGATGCCAAACGTCAGGAGAAAATGCGCGCGGCGCTGGAACAGTTGAAAGGGCTGGAAAATCTCTCTGGCGATCTGTACGAGAAGATAACTAAAGCACTGGCTTGATAAATAACCGAATGGCGGCAATAGCGCCGCCATTCGGGGAATT

>feaR

ATAACGCTACAGGAGAAGGCGATTGATCTATTTTCCTGAAACAAGGTGAATATTCAAAAACTCCTGTCAAATTGCCTTTTGCCCTGAAAAATGCATAGTCAAATATCTGTTTTAACTAATTGGCGTTGCAGTACATGCAACGCCAATTAGT

b2248_at:

>wza-wzb-wzc-wcaAB

AGTTCTGATGGCTGACGCTCTCTTTGCCGCCTTTGCCGTAACCGTAGACAAAATGTGACGCCAGCCCCTGTTGCAGCGCACGCTGGTGGAGATCTAACGCCACACCTGCTGCCCCGCCTTCCGCCAGTCGCACATTAAATTGCAAAATATT

>mdtQ

CAGGTAATACCAATATCAAACCCCTGCTGCGCCAGTAATAACGCGCACTCTTTGCCGATCCCCGAATCGGAGGCGGTAATAATCGCAACCTGTGCCATCGAGTTCTCCACTTAACGCTGAATAAACGTTAAGTATAGAAGGCGCATATCAT

>yeaN

GGCTTCATGGTGTCGGTCGGGTTCATAGCCATTGAGATTCAACCTGTGCATCATTTTGTCCGAACTTAGCGATAATTTGTCATTTTAGCTTGATTCAACATAACAATAAAAACGGTAAGGTACAGCCTCGTTTGTAACAATGAGAAGCATA

>uidABC

TACGCTTATCAGGCCTACCTTTTCCTGCAACACTTTGAATTTATGAGTTTTTGTAGGCTGGATAAGGCGTTTACGCCGCATCCGGCATAAAAAACGCGCACTTTGTCAACAATCTGAAACGCCGGAGATTTTTCTCTCCGGCGTTATTTTT

>ycjMNOPQRSTUV-ymjB

CCCATAGATTATTTGCGTCAGCTCACAAATACGCTTTTTCCCTGGTAAAAAATGATTTCCTGCGTGACTAAAACCCTTGTGCTCAATTGACAGTTTATTTTCTGCGGAGTAGTCTCTCGTTTCATGGGACCGCTACCACGGAAAGGCAACA

>yebQ

AAAAATCAGAACTGTTTTTTATTATAATTTCGCACCAGGGTGGTCGCAATCCATCTTTTGCCGGTTAGTTACAATTCTGCGACATCCACCGTGAATATCAGTGCTAGAATCATACCCCTGTTGATTATTCACCAAAGATATAAAATTCCTA

>maoC

CTGCACCTGCCAGGTTGTTTGGCAGGTGTGCCAGCTTTTCATACAGTGGATGCCCTGAAAATAGATGTACACATCATGCATAATGTGACAACGTCACAAAACTTAGTGAAATAAAAGGGCAACTATTCGCCGTTGCCCTTCATTCACCGAT

>rutABCDEFG

GCGCGATGCTTATCAGGCCTACCAGAAGATTGCAATATATTGAATTTGCACTGTTTTGTAGGCCGGATAAGGCGTTTACGCCGCATCCGGCATGAACAATGCGTACGTTGTCAACAATCTGCACCGCCGGTAACCCCGGCGGTTTTCTGTT

>ydcO

GCTCGCAACATTTTAGCGGCGGGGCACGCCGTTCTTGCCTGTGGAGAGATGGTGCAGTCAGGCCGCCCGTTGAAGCAGGAACCCACCGAAATGATTCAGGCGACAGCCTGAACGTAGCAGGGATCCACGTCCTTCAGGGCGTGGAGGATGT

>yegTUV

TTGGATGTTTTTTCAGCGTTTTCTGTTGGCTCGATTCATCAGAAAAAATGTTAGCGCGGTCAAGTTTTCACCGCAAAGGTATTTAAAAGGTATTATTAAGTGGTATTGTCATCGCGTACCTTACATTACCTGTCATGAAGGAATTAAAAGA

>torYZ

TGCCGGATGCGGCGTGAACGCCTTATCCGGCCTACAAAACCTTGCTAATTCAATATATTGCAGGGACTATGTAGGCCTGATAAGCATAGCGCATCAGGCAGCTTTACGTTTGCATAACCTCAGCGCCCGTTTCCGGGCGCTATTCACGTCT

>rnt-lhr

ATTGAGTTAATCGAAGAGAAAGACGCCGGTCGCGGTCTGGGCAACTAATCTCCTGCCGGGCGTGAACTCATCGCGCCCGCATCTTTACTGCATCGACAAGTAATATTTGTCATAATGCGCGCTGCAATTTATCCGTATTAAGAGAATCAGA

>yfdONMLK

TGATTGATACTGCAATGGCTTCCATTAGTCTGATTCAACTGAAATTACAGGCTGGGCGGAAGCTGATGCAGGCAGAGACCTCCCGACTTAACACTGTGCTGGATTACATTGACGCGGTGACGGCAACAGATACCAGCACCGCGCCGGATGT

>ompG

TGCTGGTGCGTTAAATGATTATCATGCAGGAGAAAATATCACTATTCATTTTGATATGACGAAATGTCATTTCTTTGATGCAGAAACGGAAATAGCAATTCGCTAAATACAGGGGGAAGGCATTCCCCCAGGATAATACAAGGAACAATAA

>feaB

AAGTCTTGTCAGGCATAGAGACATAAGCGGTTATTGTCACGATTTGCGGAGCTTGTCACAGCTGACAAAGCGAATGTCACAGCGAAAAAAGTGACTTTTCTTGTCGCTGCGTACACTGAAATCACACTGGGTAAATAATAAGGAAAAGTGA

>paaABCDEFGHIJK

ATAATGATTTATAAAAATAGGGTGCGAAATCCGTCACAGTTCAAACATACAAAATTTGTGATTTTACTTAACTATTGTGTAACTTTCATAAAACAATGTGATTCGTGTTTTTAATTAATTCACGAAAACTGGAATCGTAAAGGTGATGACG

>yfaXWVU

AACAAGCCATTAAAATTGCTTTGCGCATGCTGGAACAGGGCTTTGATCGTGACCAGGTGCTCGCGGCCACCCAGCTAAGCGAAGCCGATCTGGCAGCGAATAACCACTAATTAACACAGGCCCACAGCCGATCCCCATGGGCCTTTGATAT

>yebB

ATTTATTCTACAGGTTATATTGGAAGCAAATATTTTAATATTACATATTCAGTGAAGAAATGCGTAATAAAAATATACATTGCGCCTCCTGAAAAAATAAATTTTTTATGCTATTACGTATATTTGTATCTATTTCAATGGAATGACAACG

>pinE

ATTTCAGCCAGCCTGTTGGCGGAGTGGCTGAAGGCCACGGAACCGGGACACCAACAGGTAATGCAGAGCCTTCTCCCAAACCAACGTTTATGAAAATGAAGAAATAACAAGCAAATGGCATCATTCCTGCTTTTACCAGGGGGATTTAACA

>nudG

CAAGAACCCGCACGCACTGATGCAACAACGGTTTAATCCGGGGATGAAGATGGCGGTCTAGCACAGGCACTCCTTAAATATAAAGCCTTTCTGATTGAGCAACAGTGCGGATATTATGGCATTTTTCGCTTATCTGCCCGTGTGTAATTTA

>ydeE

TGAATTTACTTTTCTTTAACAGTTGATTCGTTAGTCGCCGGTTACGACGGCATTAATGCGCAAATAAGTCGCTATACTTCGGATTTTTGCCATGCTATTTCTTTACATCTCTAAAACAAAACATAACGAAACGCACTGCCGGACAGACAAA

cbl_b1987_at:

>nac

AATCCTGACGTGCAGCCTCGCGGATTATCTTTAGTTGTTGGAAATTCACGGTAAACTCCGGGCAGTTCAGATTTCCCGTTATTGTTAAAGTCTAATGCCCGGCATAACAAATAATAAAAACCCGCATCTTATTCCATCCCGATATAACACT

>ydiQRST-fadK

TTGTTTTAGCGATCGCATTTTTTTTGCAAGATTGTTGGCAAGGAAAACAGCTTGCTCCGTCGAAAACCCCGCACCGCTATCGCACACTATTTTCAGGCCATTTTTACCTTCCATCGGAGATGGTTCCGTATGCGACTCACAGGAGAAATCA

>thiMD

CGCGGATAAATTGCAACGATGGAAATTAGCCCAAGAGCGCATAATAGAGAGGTGAGTTTTTTTCGACTGGTAATTCTTAGCTGCATTGGTTTCATCCCTGAATGTCAGTGCCAAAGGCTGACAATAACCAAAGCAGCTATAGTACGGTGCT

>trpLEDCBA

TATATTACTGTTGGGCGGAAAAATGACGTAAGTTGACGTTCGACCGGGGTAAGCGAAACGGTAAAAAGATAAATATTAAATGAATTTAGGATTTTTCCGGCTTCATTAAAGAAAGTTAAAATGCCGCCAGCGGAACTGGCGGCTGTGGGAT

>chbBCARFG

ATATATTGCAGGAAACACGTAGGCCTGATAAGCGAAGCGCATCAGGCAGTTTTGCGTTTGTCAGCAGTCTCAAGCGGCGCAGTTACGCCGCCTTTGTAGGAATTAATCGCCGGATGCAAGGTTCACGCCGCATCTGGCAAACATCCTCACT

celD_b1735_at:

>ydfJ

GCCAGGTTATTAAGATCATAAACAGGGAGTGTCGCTTTTGCTGATAACAAATTATTTCCCATAACAATTCCTTAAATATAAATATGGCAAGCTATATGTTTTGTTATATGAATAAAAATCCCCTCTCCGGTAAGAGAAGGGATTAAGGGTT

>ydiQRST-fadK

TTGTTTTAGCGATCGCATTTTTTTTGCAAGATTGTTGGCAAGGAAAACAGCTTGCTCCGTCGAAAACCCCGCACCGCTATCGCACACTATTTTCAGGCCATTTTTACCTTCCATCGGAGATGGTTCCGTATGCGACTCACAGGAGAAATCA

>ydeU

ATCTGGAAACACATCTTATCCGCTCTACGGGATAAACGCTGCGATATCCGCGATCGCGGATAAAAAGGAGGATTTAAATAAGGCGTTGATGCGCCGTTTGAAGTCGGCAATATTAAGCCGCATGCCATCTCGACATGCGGCTTATACGGTT

>ydbD

TGCGCATTTACGGGAGAATATGGCTGCTGAAAAATTGCATCTTTCTGAGGAAGTGTTGTCTACGTTGGATGGTATTTCGCGAGAATAACGAATATACAAAAGGGAAAGATGCATTTCCCTTTTTTTCTTTTTTAATGGCATGGAGTGCATA

>ydjE

GTCAGCAGCTGGGGCAACGCTATATACGCTGGCAGACTGGGAAGAGACACAGGGGTAATTTTACGCTGGCCTACAATTCTGTACTGGCATTGTAGGCCAAATAAAACACGTCAGTGGCACATCTGGCAATTGATGCCATCAACGAAAGATT

>ydiP

TCCCGCTTCTGGCGTGATGTCCGTTGTGAACGTATCGGCGGCGGTACAGACGAAATTATGATTTACGTAGCAGGTCGGCAGATCCTGAAAGACTATCAGAACAAATAATCTGCAGGCGGCGCAGCTTCTTAACAAACTGCGCCGCCAGATT

csgD_b1040_at:

>csgDEFG

TACGCAATGAAAAACAGTGGCCTGTGGCTGGCGGCGCGTAGTGCAAAGACGGCGCACCGTGAGCAGGAAATCAAAAATAAAGCGTGAGGGGCACTCACGCTTTCGCTTAAACAGTAAAATGCCGGATGATAATTCCGGCTTTTTTATCTGT

>trpLEDCBA

TATATTACTGTTGGGCGGAAAAATGACGTAAGTTGACGTTCGACCGGGGTAAGCGAAACGGTAAAAAGATAAATATTAAATGAATTTAGGATTTTTCCGGCTTCATTAAAGAAAGTTAAAATGCCGCCAGCGGAACTGGCGGCTGTGGGAT

>purT

TTAACTTCAACCGCCATTTGCAGCCTCTCATAATAACTGTGATTTTATACAGTATATTTCTTTTCGGTTGAGAAATCAACATCAGCAATAAAGACACACGCAAACGTTTTCGTTTATACTGCGCGCGGAATTAATCAGGGGATATTCGTTA

>ydcF

GATGAATAATGTTTATCTACAGCATTTCCTTAAAAGATATGTCAGGCTTGCGGAGTGGCGGTTAAGGACATACGATTTCCTCCTTTCAGAGTGCTCCGCTTCTCACTATTATCTCACGCAGTATTCTTAAGGGAACGATAAGGAGGAACCA

>yddG

TGAGAAAAATAGATGAAATAATATTATTTATCGATATGTGATCGAAGTCGAAATGAGATATAAGGTGAATTACTGGTATTTGAAATTTATTTTTTTAATATTGTCGGAATTTATCTGATTAACTACCGGGCCGTAGACCCGGCAGTTATTT

>ydhS

AATCAATCAGGCAAAACTCGCCTGACAGAATTTAATCAAGGGCGGTTAGCGCCCTTTTCATCCCTGTCTGAAATTTCTCAAATTCTAAAAATCTCAACCAAACTTATCTGATAACACTAAATTCGAAAGAATGCGTACAGGTAAGTAACAA

>ydiNB-aroD

CAAAAACAAACTATGACATGCAATATTCCTGGAAACATAAACTTTATGCCATGTACCCAGGGAAAATCATCTTCAGTATAGTAATTATGTAAACCGTCGGAGAACAATACGTACGGTAACGAAATTATCTTTCAGCAAGGAGCTGTGAAAA

>potFGHI

TTGCGATGCTTTTATATAGCGAGCAGTGCTGGCCGGGAGAAAGTTCTCTTTTCTTACACCGCGCCGATAAAAAATATGCACGTTTATTGCATATCTTTCAGTGTGACAACTTTTGTTCGTTTGTTAACGAACTTTCAGAAGGAAAGAGATA

>gdhA

TATGAGATTACTCTCGTTATTAATTTGCTTTCCTGGGTCATTTTTTTCTTGCTTACCGTCACATTCTTGATGGTATAGTCGAAAACTGCAAAAGCACATGACATAAACAACATAAGCACAATCGTATTAATATATAAGGGTTTTATATCTA

dnaA_b3702_at:

>dnaAN-recF

AATACCATGTGGTGCAGACCGGTGCCGTCATCCGTGTCGCCGATATACATACCCGGGCGCTTACGCACCGCATCCAGCCCTTTCAGGACTTTGATACTGGAGGAGTCATAAGAATTCGACATCAACGTTTCTCGCTCATTTATACTTGGGT

>waaQGP-rfaS-waaBIJY-rfaZ-waaK

GCATTGTGGGGATTGGACTCAGTGATGTGATCATATGGGCACGCAGCATTCCAATTATCATTATATCCGCTATAGTCCTCTTACTCGTCATTAATAATCGTAACAATACAATTAATTAAGAATAAACAAGTTTAAGAAGTGAGTTAAAACT

>holD-rimI-yjjG

ATCACTGTGACGCAGCAAGACTTCACTTGCCGGGGTAAATGCAGACATGGAATGCTCCTCAATTGATACTGGCGGCGATTATAGCCATATGTTGGCGCGGTATCGACGAATTTGCTATATTTGCGCCCCTGACAACAGGAGCGATTCGCTA

>gmk

CAGTGAATGACAGGCAAATGCGGAAGCAGCTACGCAAAACGCAACAACTTTGCGCAAAAAGTGTGAGCAAGGGCTACGTCACATGGCCGCGCCGTGTATAATAAGCTCGTATGTAGGCTTTATTTCGCTAATCACATACGAAAGATACTCA

```
>yjeFE-amiB-mutL-miaA-hfq-hflXKC
TCGCCGTGGTATTGTTTGTCCAGCCATGCTTGCAGTTTGGGCTCGGACTCGCTGAGATCGGTATCGGTAATACCTACCTGCTGAAAGCCCAGTTCCAGCCCCCACTGTTTAATTTTTTGCGCTAACTGATTGAGATCGAGGGGCTCTGACA
>glmUS
ATTAATAACGAAGAGATGACAGAAAAATTTTCATTCTGTGACAGAGAAAAAGTAGCCGAAGATGACGGTTTGTCACATGGAGTTGGCAGGATGTTTGATTAAAAACATAACAGGAAGAAAAATGCCCCGCTTACGCAGGGCATCCATTTAT
```

feaR_b1384_at:

```
>recET-lar-ydaCQ-intR
GAAGTCAGCAATGGCTTCGCCCACGTTACGACGCAGACGTTTTTGTAATTTGTTCAGGTTGTATTGTTCTTTCTTTGTAATTTGTTGATTTTCTTGCATTATTTCAGTTCTCTGGTACTAAATGGGGCAAATTGGGGGCAAACTTTGCAAC
>rutABCDEFG
GCGCGATGCTTATCAGGCCTACCAGAAGATTGCAATATATTGAATTTGCACTGTTTTGTAGGCCGGATAAGGCGTTTACGCCGCATCCGGCATGAACAATGCGTACGTTGTCAACAATCTGCACCGCCGGTAACCCCGGCGGTTTTCTGTT
>ydcJ
CAAAATATTATTAACAATTAAAGCAATTATGTTACAGCAAAATGGATAATATTGATGTTTTCGCGGCGAGATCACAGTTTGTAAATTCTTCCCGCAAGAGTGAATGCGGTTACCTACACTCCAGATTACTGACCACTGGAGGCAGACACTA
>mcrA
GTGTGTCAACTTTGTATAAGAGGTTTCCTGCAGGGGATAAATAAAGTTAAAGACACTTTGTGTACAAAAGAAAGTAAAACAACAGCAACTTGTTGCAATTTTATCAATAAAAGTAGTATTGTCGTGAAAAATTGATTAAAGATTAATATTA
>feaB
AAGTCTTGTCAGGCATAGAGACATAAGCGGTTATTGTCACGATTTGCGGAGCTTGTCACAGCTGACAAAGCGAATGTCACAGCGAAAAAAGTGACTTTTCTTGTCGCTGCGTACACTGAAATCACACTGGGTAAATAATAAGGAAAAGTGA
>yeeJ
CCATTTTCCCCTCGATTATAAAACTTGAGTTATTCAGTAGTCTCCCCTCTTGCAACTCACACCCAAAACTGCCTAACGAAAAGTTATTAATTTTCAATCATATTGCTATCAGTATTTACATTTTTTCGCTGTGCTAGAAAGGGCGCATTTA
```

flhC_b1891_at:

```
>flhDC
CAGGTAACCTAATAAGATAAGCACGACATCATCCTTCCACTGTTGACCATGACAGGATGTTCAGTCGTCAGGCGTTAACGCGCGATTGGGGCAAAAAAAAGCAGCGGTACGTCGTTACCGCTGCTGGAATGTTGCGCCTCACCGTATCAGT
>flgBCDEFGHIJ
GCCCCCAGCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAACGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATTTAAGCTGATGGCAGAATTTTGATACCTGCGGAGGAGATA
>flgAMN
TGGCGTTTACTGCGTTTAATGGCGGTCGTGCTGGCGGGGATTGCCGCGTACTTCGCTGCACTGGCGGTACTGGGCTTCAAAGTTAAAGAATTTGCCCGCCGGACGGTGTAACAATGCATTCCGGCCTGCAGTGCAGGCCGGAGATAATCTT
>fliAZY
CGATAAACTCCAGCCGTGGAAAACGGGTTAAATTATGCAGTGGCATAACAGCCTCCGATGTGTGTTGTTGTGATTTTCTTATTATGCACGCTGAAAACGCGTAAATAAAAAAGGCGCTAGTGAAAGCGCCCTTTTTTGTCATTATGCTGAT
>flgKL
ACTCACCAACATGATTCAGCAGATGAAATCGATAAGCGACAAGGTGAGCAAAACCTACAGTATGAACATTGATAATCTGTTCTGAATAACTCAAGTCCGGCGGGTCGCTGCCGATAATACTCTGTAATTGAAGGCTTATAAGGAACCTCCA
>fliLMNOPQR
ACAAGGGCGTGTAACAGGCAACAGCGGCGTTGATATTTTCGCCTAACGTCAGAGGTAGCACCGTAATCCGCGTCTTTTCCCCGCTTTGTTGCGCTCAAGACGCAGGATAATTAGCCGATAAGCAGTAGCGACACAGGAAGACCGCAACACA
>yecR
ATGTTATAAAAATGATAATCAAAAAACAGCCCCCCTATTTCTGACACCTACAGATGGCAAGAAATAGCGCCTGCCAGGCGTCTTTTCCGGCCATTGTCGCAGCACTGTAACGCGTAAAATAGTGCTTTCTCTTACTCTTCTGGCTGGACCA
```

>flhBAE

TTTTTATCCAAGCCCTTTGACAAGAGGATAATTCACATCTTTTTGGCATGTTTTGTTGCAAGCTATTCCTGATAAATAATTGCAACAAGACATCGAGCCTTTTTCACTGAGTTATTAAACATACTCGCGAGCGCGTAATTTTTTTGTCCTT

fliA_b1922_at:

>fliAZY

CGATAAACTCCAGCCGTGGAAAACGGGTTAAATTATGCAGTGGCATAACAGCCTCCGATGTGTGTTGTTGTGATTTTCTTATTATGCACGCTGAAAACGCGTAAATAAAAAAGGCGCTAGTGAAAGCGCCCTTTTTTGTCATTATGCTGAT

>fliDST

AAAACAGCCATTTTTTGTTAGTCGCCGAAATACTCTTTTCTCTGCCCCTTATTCCCGCTATTAAAAAAAACAATTAAACGTAAACTTTGCGCAATTCAGACCGATAACCCCGGTATTCGTTTTACGTGTCGAAAGATAAAAGGAAATCGCA

>flgKL

ACTCACCAACATGATTCAGCAGATGAAATCGATAAGCGACAAGGTGAGCAAAACCTACAGTATGAACATTGATAATCTGTTCTGAATAACTCAAGTCCGGCGGGTCGCTGCCGATAATACTCTGTAATTGAAGGCTTATAAGGAACCTCCA

>flgBCDEFGHIJ

GCCCCCAGCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAACGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATTTAAGCTGATGGCAGAATTTTGATACCTGCGGAGGAGATA

>flgAMN

TGGCGTTTACTGCGTTTAATGGCGGTCGTGCTGGCGGGGATTGCCGCGTACTTCGCTGCACTGGCGGTACTGGGCTTCAAAGTTAAAGAATTTGCCCGCCGGACGGTGTAACAATGCATTCCGGCCTGCAGTGCAGGCCGGAGATAATCTT

>fliFGHIJK

AATAGCAACAAAAAAACGGGTTTATTGGCGGATAGAAAAAAACGAAAGCACAAATAATGGGAGCGTCAATTTTTCGAGTTTGCTGACCCGGGAGTGAGTCTTGTTCCACTTTGCCAATAACGCCGTCCATAATCAGCCACGAGGTGCGCGA

>motAB-cheAW

AATCATAAGGCACCTTCCTGAAAACAAGTTGATCTCGTTATCGGCAAGGAGGGGGGAAACTTTATTGCTGATGCCACCCGCCGCGAAATTGAAATAAAAAACCCGATGCGCAGATCATCGGGTTCATTTCAATTGAGGAAATCGGGAGAAT

>fliLMNOPQR

ACAAGGGCGTGTAACAGGCAACAGCGGCGTTGATATTTTCGCCTAACGTCAGAGGTAGCACCGTAATCCGCGTCTTTTCCCCGCTTTGTTGCGCTCAAGACGCAGGATAATTAGCCGATAAGCAGTAGCGACACAGGAAGACCGCAACACA

>tar-tap-cheRBYZ

CGTTCAATGGTTTGAGTAAGGGGCAAAACAGGCGGGATTTAGGGCTTTTGCTGCCACACATCAAGCATAGTGTGCGTTTGTCGGATGCGGCGTCATCGCCTTATCAGACCGCCTGATATGACGTGGTCACGCCACATCAGGCAATACAAAT

>flxA

GAACGTTTACCCCTGTCAACAAGCTTTACTTTCTGAGGCGCGCCAGCCCGCGAGGAAAACAATCTGAACATCAAACAATTAATGACACAAGAAATACGATTAAAGATTTTTTTGTGCATGCCGATAGTGCTTTTTTAAAAGGAGAAATCTA

>ycgR

GGACTTTCTCGTCAGATCGGAAAAAGGCGATCAGCAAAATCAACGATTTGGATGCTGACGAGTTCCTCGAACACGTAGCGCGAAATCACCCTGCGCCGCAGGCTCCGCGCTATATCTACAAACTTGAGCAGGCACTGGACGCGATGTAAAT

>ves

TGACTCAATATCATATTATTCAAAACTGGCTGTGGCTGGGGGCGGTTAATTCGCTGGAAGAAGCGACAACGTTAATTCGGACACCCGCCGGGTTTGATCACGACGGTTATAAAATTCTTTGTAAGCCGCTGCTTTCCGGTAACTATGAAAT

>yecR

ATGTTATAAAAATGATAATCAAAAAACAGCCCCCCTATTTCTGACACCTACAGATGGCAAGAAATAGCGCCTGCCAGGCGTCTTTTCCGGCCATTGTCGCAGCACTGTAACGCGTAAAATAGTGCTTTCTCTTACTCTTCTGGCTGGACCA

>flhBAE

TTTTTATCCAAGCCCTTTGACAAGAGGATAATTCACATCTTTTTGGCATGTTTTGTTGCAAGCTATTCCTGATAAATAATTGCAACAAGACATCGAGCCTTTTTCACTGAGTTATTAAACATACTCGCGAGCGCGTAATTTTTTTGTCCTT

>fliE

GATGTTCAGGGGGGTTATCGCGTATACGGATCGTGGATGGCAGAAAACGTTCAGGATCAGGTATCCATTTTAAACCAGAAATTGAGTGAGTTTGCCCCATCCATGCCCCACGCGGTGAGATCGGATGTAATTAACAACAGGTTACAAAACC

>flhDC

CAGGTAACCTAATAAGATAAGCACGACATCATCCTTCCACTGTTGACCATGACAGGATGTTCAGTCGTCAGGCGTTAACGCGCGATTGGGGCAAAAAAAAGCAGCGGTACGTCGTTACCGCTGCTGGAATGTTGCGCCTCACCGTATCAGT

>ynjH

CGGGGCGGATTATCCATCTTCATGCCTGGCACGTACCCGACTTCCACGGGACGTTACAGGCACATGAACATCAGGCGCTGGTCTGGTGCTCACCTGAAGAGGCGCTGCAATATCCGCTGGCCCCTGCTGACATTCCATTATTAGAGGCGTT

>ymdA

CAATGGCCGCCCTCTTCAGAAAAGTCTTAATTTGTTGAAATATCGAGCATAAGATGAATCTGGAGAGAATGGTCTGCTGCGAATCAGCCAACCTGAAAGTATGGATAACACAACCCTCAAGGATGACTAATCATTGAGGAAATAGAATAAA

>fliC

TTTATGTTGCCGGATGCGGCGTAAACGCCTTATCCGGCCTACAAAAATGTGCAAATTCAATAAATTGCAATTCAACTTGTAGGCCTGATAAGCGCAGCGCATCAGGCAATTTGGCGTTGCCGTCAGTCTCAGTTAATCAGGTTACAACGAT

fnr_b1334_at:

>ynaJ

TTTCCCCGCTTATTCGCACCTTCCCTAACTAATCAATGCGTTGATTGTAAATCCAGCTAAGAGGTGAGGTTTTCAGAGCAGACAACGGTGAAATGTCATGGTATTGTTACGTTTAGGTAACAAGAAATTTGTCTGCACAAGGATTACATCA

>uspE

ATGTGATTCAGTCATCGATGATTTTTGTTTATAGCCTGGTCTTTACTGTGTTGCTGGCAATCCCGTTGGGAATTTATTTCCTTGGCGGCGAAGAGCAGTAAGTAAAAAATAGGCCCGATAACTCGGGCCTTGTCAGTTATTGAAGAGTCGT

>ydaN

ACACGGTTTGTATACGGAAAAGCATTTTGCTTTTTGTATTCAATTTAGACAGAATTTTATTAATCATTTCAGGGTAATGGGGTGATGAGATGTTGCGTAACAGGGCCAGAAGGCTAGACTACAAAATAATGCGTTGATGATGGAGGCACTG

>abgABT-ogt

TTTTAGGGAGCAAGTAAGTCTAAGCAAACCTTAACAGCAGAGAATTCCGATATTAGATGTAAATATATGTCTATCTATTTGAAAACCCTTAAGTTGTTAAGGGTAACTTTACATAAAAGTGTGAACAAGCTGGCACAAATTGTTTAATGTT

>abgR

TGTGACCCGGTTCACGTAGCGATAGTTTTTACTTATCACTAACTGATTTTTCACAGTTTTAACCGTTCATAAATTACCCTGACACAATCATCTGCATTAAAGTAGATGCCAGTTTCTTTGGTCTGATAAATAACGGTTATCGGTGGCGTCA

>ydaM

TGATGATGTTCAGGCATATTGCACCGCGCTACCGCATCATGGCGGCAGCGGGGCGTGTTACGTCGCACTACGTAAAACGGCGCAGGCGAAGCAAGAAAACTGGGAGCGCCACGCTAAGCGCAGTCGTTGATCTCGAGACGCATCCGCGGCT

>ydaL

ATTAATATCGTGAATGAATAATCATGCATAAGTATTTTGCTTAAAATATCGGCAATATTTGGAACTTATTACTGGAAATTTGGGTAATACGTTGTTGGACCGACCCGGTCTGGTTATCATATCGCGCTCTTAATTGCGGGAGGATGTAACA

>C0343-dbpA

TTCGTACGACGCCGACTTTGATGAGTCGGCTTTTTTTTGCCTGTTATTTATCAGCGTCTACCCTTTAAGAGTCCACCCAATGACCAGAGGGAAATATGACGACACTTATTTATTTGCAAATTCCTGTCCCTGAACCGATTCCTGGCGATCC

>narGHJI

AAGGGGTATCTTAGGAATTTACTTTATTTTTCATCCCCATCACTCTTGATCGTTATCAATTCCCACGCTGTTTCAGAGCGTTACCTTGCCCTTAAACATTAGCAATGTCGATTTATCAGAGGGCCGACAGGCTCCCACAGGAGAAAACCGA

>dgsA-ynfK

CTCGGATTGACCGGGATGGCGACACTGTTGGCGGCGACCACGCACGCGCCGATTATGTCGACGTTGATGATATGTGAAATGACCGGGGAGTATCAGCTACTCCCCGGTTTATTGATTGCCTGCGTAATTGCGTCGGTAATTTCGCGGACGT

>bioBFCD

AGGTCATGGATGTGTATGGGTGCCAGATATGGCGTTGGTCAAAGGCAAGATCGTCCGTTGTCATAATCGACTTGTAAACCAAATTGAAAAGATTTAGGTTTACAAGTCTACACCGAATTAACAACAAAAAACACGTTTTGGAGAAGCCCCA

>flhBAE

TTTTTATCCAAGCCCTTTGACAAGAGGATAATTCACATCTTTTTGGCATGTTTTGTTGCAAGCTATTCCTGATAAATAATTGCAACAAGACATCGAGCCTTTTTCACTGAGTTATTAAACATACTCGCGAGCGCGTAATTTTTTTGTCCTT

>rutABCDEFG

GCGCGATGCTTATCAGGCCTACCAGAAGATTGCAATATATTGAATTTGCACTGTTTTGTAGGCCGGATAAGGCGTTTACGCCGCATCCGGCATGAACAATGCGTACGTTGTCAACAATCTGCACCGCCGGTAACCCCGGCGGTTTTCTGTT

>lsrACDBFG-tam

GCATGAACAAACGCAACCGTGAAAATCAAAATAGCATAAATTGTGATCTATTCGTCGGAAATATGTGCAATGTCCACCTAAGGTTATGAACAAATTAAAAGCAGAAATACATTTGTTCAAAACTCACCTGCAAAACTGAACGGGGGAAATA

lexA_b4043_at:

>lexA-dinF

TGATTACCTGGTGCATTCTGTTATGGTCGCATTTTGGATAACCCTTCCAGAATTCGATAAATCTCTGGTTTATTGTGCAGTTTATGGTTCCAAAATCGCCTTTTGCTGTATATACTCACAGCATAACTGTATATACACCCAGGGGGCGGAA

>recAX

ATCTCAGCGGTGCTCTTGCTCATAATTATCCTGAAATCAAGCTAACGAAATATCGCCACCAGCTCCAGCGTGTCTTAACCGCCGGGCTGGTAACTGAAAAGTGGGAATAAGATAAGTTTTCTTGACTGGGAAGTAAAATACCGTATGCGTT

>yebG

CTTGCTGGTTTCATTATTGGCAGCGAAAACTGATGCGCAGGCAGAAACCAACAACAGCCCTAAAAACGCCCCTCTTTTTTTCATGTTTTTCTCCATAGCACAATGATTCAGGAGAAAGCATGGTACAAATTGTCAGGAGCGCAAGTTGCTT

>dinD

CCCGCAACATTCAATTCTGTTTTGCGTGCCTGCTCCAGATTTTGCGATGTTTTTTTGCCCAGCACACTGAGAACGTGAGATACTCACAACTGTATATAAATACAGTTACAGATTTACTTTCTTTGCAATTGATATCACATGGAGTGGGCAA

>recN

CGAAAGATTACAGTTATTTCAACACATTAAGCACCAAGCTCGGCTGGTCAAAAAAATTATTCTAATTTTACGCCAGCCTCTTTACTGTATATAAAACCAGTTTATACTGTACACAATAACAGTAATGGTTTTTCATACAGGAAAACGACTA

>sulA

ATAAGGTATGTTTAATCTTTTTTGTCAGCGACAATTTACAGAAGAGAATCGCGGAAACCGCTTCAGACAAGCCTCCGCAAGGAAAATTAGTCACGACTGAAAGCATTGGCTGGGCGACAAAAAAAGTTCCAGGATTAATCCTAAATTTACT

>dinI

TAAGATTTATATGAACAATAAAACAGCATGTCATTCATATTTTTTTAGCATATTGTGCAATTATTTTGAGGAAGTGTAGAAATTTTGTACTCAAAATTCGTAAGTAAAATAAAAAAGCCGGGGCGACCCGGCAAAAAAAATCACTGCATAT

>waaQGP-rfaS-waaBIJY-rfaZ-waaK

GCATTGTGGGGATTGGACTCAGTGATGTGATCATATGGGCACGCAGCATTCCAATTATCATTATATCCGCTATAGTCCTCTTACTCGTCATTAATAATCGTAACAATACAATTAATTAAGAATAAACAAGTTTAAGAAGTGAGTTAAAACT

>ynaJ

TTTCCCCGCTTATTCGCACCTTCCCTAACTAATCAATGCGTTGATTGTAAATCCAGCTAAGAGGTGAGGTTTTCAGAGCAGACAACGGTGAAATGTCATGGTATTGTTACGTTTAGGTAACAAGAAATTTGTCTGCACAAGGATTACATCA

nac_b1988_at:

>glnK-amtB

ATGTAACGCACTGTGCACTGTCATAGTGCGTTTTCATTTTCAAACTTCTTAACTTCCTGCTCTCTTTCTCGTTTTTCATTTCTGGCACACCGCTTGCAATACCTTCTTCGTGTAGCAGAACCATTACCGAATTCTGACCGGAGGGGATCTA

>yedL

TGTTAAAAACCAGGGCGCGGAGTTAATTCAACCTGTTTGATAAGCCTGGAGATTATTGATCAACAATACTGCGTCATAAGAAATCTCTATTAGACAAAGATTTCATTACCTGTTGGCATATTGCAAAAATAACACCAATACGGAATCGTCA

>trpLEDCBA

TATATTACTGTTGGGCGGAAAAATGACGTAAGTTGACGTTCGACCGGGGTAAGCGAAACGGTAAAAAGATAAATATTAAATGAATTTAGGATTTTTCCGGCTTCATTAAAGAAAGTTAAAATGCCGCCAGCGGAACTGGCGGCTGTGGGAT

```
>cbl
AAGATTTTTCGCCGTTAAGATGTGCCTCAACAACGATTCCTCTGTAGTTCAGTCGGTAGAACGGCGGACTGTTAATCCGTATGTCACTGGTTCGAGTCCAGTCAGAGGAGCCAAATTCCTGAAAAGCCCGCTTTTATAGCGGGATTTTTGC
>ydiQRST-fadK
TTGTTTTAGCGATCGCATTTTTTTTGCAAGATTGTTGGCAAGGAAAACAGCTTGCTCCGTCGAAAACCCCGCACCGCTATCGCACACTATTTTCAGGCCATTTTTACCTTCCATCGGAGATGGTTCCGTATGCGACTCACAGGAGAAATCA
>thiMD
CGCGGATAAATTGCAACGATGGAAATTAGCCCAAGAGCGCATAATAGAGAGGTGAGTTTTTTTCGACTGGTAATTCTTAGCTGCATTGGTTTCATCCCTGAATGTCAGTGCCAAAGGCTGACAATAACCAAAGCAGCTATAGTACGGTGCT
>yncG
AGCATCCTACCCGTTATCGATAAACGATGCAAAACATCCCCTTACAATCCTGAAGGGGATTAATACAACTGACGAAAAAATGACAAATCCTTTTGCTGGTTAACCTGTGTACTGTCCTACACTTAATCTTTAAAAGATTGTGAGGGGCATA
>narZYWV
AACAACACCGGCAGAATTGCCGCGAAACGGTTGTGAGGTAAAAGCATCGACGTGGTACACCTGCGGTTTCATTAACGTTCTCCTGTGACTGGAGAACTATCATAGCCTGCAAGTGGCCGGAGAGCGAAGGGCTATCCGGCCAGGGTGAAAT
>ubiE-yigP-ubiB
AGCCGAATGATGAAGCTTATCAACGCGATGATGAATATAATCAGCAGTCGCGCTAGCCCATTGGGAGTAGTTAAGCCGGGTAGAAATCTAGGGCATCGACGCCCAATCTGTTACACTTCTGGAACAATTTTTTGATGAGCAGGCATTGAGA
>pphA
TGTGAAGATATTGTTGTGGATGTCTACAACACGGAACAGCAGTGTCTTTATTCTATGAGCGATCAACGGATCCGCCAGGGCGGTTGTTTTCCGATTGAGGATTTTATAGATGGTTTCTGGCGACCTGCACAGGAGTACGGTGATTTTTAAT
>asr
TTATTATCAACGCTGTAATTTATTCAGCGTTTGTACATATCGTTACACGCTGAAACCAACCACTCACGGAAGTCTGCCATTCCCAGGGATATAGTTATTTCAACGGCCCCGCAGTGGGGTTAAATGAAAAAACAAATTGAGGGTATGACAA
>rutABCDEFG
GCGCGATGCTTATCAGGCCTACCAGAAGATTGCAATATATTGAATTTGCACTGTTTTGTAGGCCGGATAAGGCGTTTACGCCGCATCCGGCATGAACAATGCGTACGTTGTCAACAATCTGCACCGCCGGTAACCCCGGCGGTTTTCTGTT
>waaQGP-rfaS-waaBIJY-rfaZ-waaK
GCATTGTGGGGATTGGACTCAGTGATGTGATCATATGGGCACGCAGCATTCCAATTATCATTATATCCGCTATAGTCCTCTTACTCGTCATTAATAATCGTAACAATACAATTAATTAAGAATAAACAAGTTTAAGAAGTGAGTTAAAACT
>appCB-yccB-appA
AAGCTGCACCGACGTAAAAAGACGGTAAGTATCGCTTTCAGTCTTATGAATATCGCAATCGGCGAATACCTCTGGTCGTAGAGTTTCAGGATAAAGAGGGAGATCTACCATTATCGGGTTATTTTTCTCTCTTCGCCTACAGGAGTGCGCA
>abgR
TGTGACCCGGTTCACGTAGCGATAGTTTTTACTTATCACTAACTGATTTTTCACAGTTTTAACCGTTCATAAATTACCCTGACACAATCATCTGCATTAAAGTAGATGCCAGTTTCTTTGGTCTGATAAATAACGGTTATCGGTGGCGTCA
>ddpXABCDF
CCTTTGACGGCATAATCCAGAAAGCAAAAGCAGGATGCCCGGTCTCTCAGGTACTGAAAGCGGAAATTACGCTGGATTACCAGTTGAAATCGTAAAGCATTGCCGGATGACGCGTCAGGCGCGTGAATGCCTGATGCGTTGTTAGCATCTC
>ydaM
TGATGATGTTCAGGCATATTGCACCGCGCTACCGCATCATGGCGGCAGCGGGGCGTGTTACGTCGCACTACGTAAAACGGCGCAGGCGAAGCAAGAAAACTGGGAGCGCCACGCTAAGCGCAGTCGTTGATCTCGAGACGCATCCGCGGCT
>gdhA
TATGAGATTACTCTCGTTATTAATTTGCTTTCCTGGGTCATTTTTTTCTTGCTTACCGTCACATTCTTGATGGTATAGTCGAAAACTGCAAAAGCACATGACATAAACAACATAAGCACAATCGTATTAATATATAAGGGTTTTATATCTA
>ydaSTUVW-rzpR
TCGGCCTAAGTCTTTACCACTAAGCATTGCTTAATATTCTCCTATGCGCATTACATTAGGCAATCCCTACCCTTACTGCATTAGGCACAGCCTATTGACAATTGCGTTAGGCGTCGCCTAATATTTCTGTGTGTTTTTGGAGTTCATTCGA
>ydcK
```

AGACGCTTCGTGATCAAATGTCGGGTGGACAATCCGCAAAGCAACCAGGTCGCTTTGCGCAATGGTTTTATCCTTGAAGGTTGCCTGAAACAGGCTGAGTTCCTGAATGATGCCTATGATGATGTGAACTTATACGCGCGTATTATCGATT

>potFGHI

TTGCGATGCTTTTATATAGCGAGCAGTGCTGGCCGGGAGAAAGTTCTCTTTTCTTACACCGCGCCGATAAAAAATATGCACGTTTATTGCATATCTTTCAGTGTGACAACTTTTGTTCGTTTGTTAACGAACTTTCAGAAGGAAAGAGATA

>ynaJ

TTTCCCCGCTTATTCGCACCTTCCCTAACTAATCAATGCGTTGATTGTAAATCCAGCTAAGAGGTGAGGTTTTCAGAGCAGACAACGGTGAAATGTCATGGTATTGTTACGTTTAGGTAACAAGAAATTTGTCTGCACAAGGATTACATCA

>ydiJIH

ATAAAGCTAACCCGCCGTTTTAACACAAACTGCGATTAGTATTATTTTTGAACAATATCAGGCGGTAGATAAGCAGTATTAAGAAGGTCATCGAACCTGGACGGAGGTTAATCCAGGTCGATTTGGCGAACTTGCGGCATTAAGTCAGGAT

>hemCDXY

TGCCTCATCAGACACCATGGACACAACGTTGAGTGAAGCACCCACTTGTTGTCATACAGACCTGTTTTAACGCCTGCTCCGTAATAAGAGCAGGCGTTTTTTTATGTATCAGGAAGGCCCCGGAGGTGCTTGCCTCCGGGTGAGAAGGAAC

>thrU-tyrU-glyT-thrT-tufB

TGAGCACAATGATGTTGAAAAAGTGTGCTAATCTGCCCTCCGTTCGGCTGTTTCTTCATCGTGTCGCATAAAATGTGACCAATAAAACAAATTATGCAATTTTTTAGTTGCATGAACTCGCATGTCTCCATAGAATGCGCGCTACTTGATG

>yoaD

TCAAAGTCCAGTGTGACTAATACTTCTTACTCGCCCATCTGCAACGGATGGGCGAATTTATACCCGCTTTCTCGTCTGCTGTAATATTCCCCACTACACTTCCACTGTTGCGTCAGGCGTTTGTCGCCATACGCTTACAGGGTGGCCCGCA

>eco

TCTCATTAGAATTTAACACTAAAAGAGCAGGTAAAATTGTCTGAATGTTCTTTAAGTTATTCATAAAGCAAATTAATAAATCTGATGAATATGTTAACCTTCAGCGACATCATCGGTGAAAACCTATAAATGAAGAAGGAAAGCAAAAAAA

>ydiFO

AGGCATAAGCAATAATATTTCGGCGGGAACACCCTCCCCGCCGAACTAAAAAATATATTCAATCGTATTTAATAAAAATATTTCGTGAGTCTCTGTGCGCTAATTCTCCATTTGGCGTAGGGAAAATCACATCTGAATCAGGAATTAACAA

>yeaVW

TTTTTATTCGTTCTTTGCAGTAAATAACCTGCGTCATTTCACCTTTTATTGTTTCCGTTTCGTGTTTTATGGCTTTCCGTATTCTTAATTGTTTAATTTATGTAACATGCAAATTTTGTTACGCGTACGTTAGGTTCCGCCGTACAGGTAT

purR_b1658_at:

>ydiJIH

ATAAAGCTAACCCGCCGTTTTAACACAAACTGCGATTAGTATTATTTTTGAACAATATCAGGCGGTAGATAAGCAGTATTAAGAAGGTCATCGAACCTGGACGGAGGTTAATCCAGGTCGATTTGGCGAACTTGCGGCATTAAGTCAGGAT

>ydjN

CATCAGGTTGAATACTAAAAAGGCAAAAATCACCTTTCTGGAATAAGCAATTCCATTTGAATATAAGAGCCAGCTCACAGTTCTGTTAATCTTGCGCCAACACTATGACTGCTACGCAGTGATAGAAATAATAAGATCAGGAGAACGGGGA

>yciSM-pyrF-yciH

AACGCTGGTTGATTTTCCGAATTTAGCCCTTAAATCATCAACAATGCGTGTGGATGCCATTTCGCAGACGGCGCGAAAATGGTACTTTAAAGGGCTATTGCGGTAAGTTGACCATAATTTATTCGCTCTAACCACATAACGGGAAGTAATG

>aroH

AGTGCATTAGCTTATTTTTTTGTTATCATGCTAACCACCCGGCGAGGTGTGACACACCTCGCACTTGAAATCAGCAGCGATTGGTTTATCGTGATGCGCATCACTTCCCGGCAGTCCTGCCGTAGAAGCAACAAATTTCTGAGACTTGTAA

>purT

TTAACTTCAACCGCCATTTGCAGCCTCTCATAATAACTGTGATTTTATACAGTATATTTCTTTTCGGTTGAGAAATCAACATCAGCAATAAAGACACACGCAAACGTTTTCGTTTATACTGCGCGCGGAATTAATCAGGGGATATTCGTTA

>pntAB

TGGTTGTTAGGACCGGTGGGTATGCTGCTTTCCGTGCCGTTGACAATTATTGTCAAAATTGCGCTTGAACAAACAGCGGGAGGTCAAAGCATCGCCGTTCTGTTAAGCGATCTCAATAAAGAGTGACGGCCTCAGCAGAGGCCGTCAGGGT

```
>yciW
GCTTTTACCACCCCTTCAGCGCGTGGCGTCTGGGAATGCAGTTGCTGTTTAAGCTGCGCTAGCAGCGGGTTGTCCTGAAACATAATTGT
CTATTTTGGTGGCCATTAGAGCGGCTGACAGTTTTACGCGAATCTGTCTGACGCGGCAAGGT
>gdhA
TATGAGATTACTCTCGTTATTAATTTGCTTTCCTGGGTCATTTTTTTCTTGCTTACCGTCACATTCTTGATGGTATAGTCGAAAACTGCA
AAAGCACATGACATAAACAACATAAGCACAATCGTATTAATATATAAGGGTTTTATATCTA
>hisLGDCBHAFI
AAGAGGACAGTCCTCTTAGCCCCCTCCTTTCCCCGCTCATTCATTAAACAAATCCATTGCCATAAAATATATAAAAAAGCCCTTGCTTTC
TAACGTGAAAGTGGTTTAGGTTAAAAGACATCAGTTGAATAAACATTCACAGAGACTTTTA
>thiMD
CGCGGATAAATTGCAACGATGGAAATTAGCCCAAGAGCGCATAATAGAGAGGTGAGTTTTTTTCGACTGGTAATTCTTAGCTGCATTGG
TTTCATCCCTGAATGTCAGTGCCAAAGGCTGACAATAACCAAAGCAGCTATAGTACGGTGCT
>trpLEDCBA
TATATTACTGTTGGGCGGAAAAATGACGTAAGTTGACGTTCGACCGGGGTAAGCGAAACGGTAAAAAGATAAATATTAAATGAATTTAG
GATTTTTCCGGCTTCATTAAAGAAAGTTAAAATGCCGCCAGCGGAACTGGCGGCTGTGGGAT
>pyrD
CGGCGGGTTGAGTGCAAAGAAGGAGCAAAATCTGCCCTGAAACAGGTTCGGAAAACGTTTGCGTTTTTTTTGCCGCAGGTCAATTCCCT
TTTGGTCCGAACTCGCACATAATACGCCCCCGGTTTGCACACCGGGAATCCAGGAGAGTTCA
>ydiNB-aroD
CAAAAACAAACTATGACATGCAATATTCCTGGAAACATAAACTTTATGCCATGTACCCAGGGAAAATCATCTTCAGTATAGTAATTATGT
AAACCGTCGGAGAACAATACGTACGGTAACGAAATTATCTTTCAGCAAGGAGCTGTGAAAA
>ynaJ
TTTCCCCGCTTATTCGCACCTTCCCTAACTAATCAATGCGTTGATTGTAAATCCAGCTAAGAGGTGAGGTTTTCAGAGCAGACAACGGT
GAAATGTCATGGTATTGTTACGTTTAGGTAACAAGAAATTTGTCTGCACAAGGATTACATCA
>dcyD-yecSC
ATTTATATTATTCAGGCAATGAATTACTTTTGCAAGCCATCGCATTCTCTTATGTTATTAATGAGTTATGCTGATTTGTTAAGCAGTTTT
ATCAGGCTTGAAATGGCGTCCAGCCCCGACAGGTGAATCGTCGGGGCTGATTTTTTCTTAT
>eamA
AATGTCGTGGTTGTTCCCCCATCGCAGGAACACCCACCGTTTGATTTAAATCACATGGGTACTGGCAGTGATAAGTCGGATGCGCTCGG
CGTGCCCTATTATAATCAACACGCTATGTAGTTTGTTCTGGCCCCGACATCTCGGGGCTTAT
>ydhP
AAATGCAGCATTGCCTGAAGCGCTACGCTTATCAGGCCTACGCGGATCATCGATGTAGGTCGGATAAGGCACTCGCCGCATCCGGCAAG
ATAAATCGCACGTTGTCAGCAACTGTAACGCAGAAGGTTATCCTTCTGCGTTTTTGTTTAAT
>rnb
CATTGTTTTTTTATATATTTATTTGTAATCCAGTTTTGGAAAAACGCCAGTTTTCAAACGAAAGTCAGTTAAAAAATCTGCCTGGATAT
AACGAAGGTAGAGCGGGGAAATAAACGGCCCATCCATGAGGAATGGGCCGTGAAAGGAGAT
>yecN-cmoAB
CACCGGAACTGGTACATACCCTGTGGGAAGACTTACGTAAAACGCTTCCAGAGATCGGAGCAGTTCCGGCTATTCCACAGCAATCTTCT
CTTTTCTGAATTTGCCACCTATCATAGACAGGTGCCATCGGCCATTTTAAAGGGAGTTTGTA
>ycgM
GTTACCGCCACCACCCGAAGATTTGCTGAAGCAACATCTTTCCGTCATGGGGCAGAAAACAGACGACACTAACAAATAACCGATATCCG
GCGGTGGCATTATCTTTGTCGGCGCGGGTTTTCATATCCACGATAAGGTGAGGGGAACGTTA
```

rhaR_b3906_at:

```
>yijF
TTCCTGGTGGCTGCTCGGCGAGAATCCGGGGGCCGTTGAAGGTAGCGGTATTGTGCTGATTGTGCTGGCACTGGCGCTGGTGAGCCGT
AAGAAAAAAGAAGCCGTCAGTGTAAAAAGGATCTGAATTTTTTCTTCATGTGGGGCGATCTCT
>rhaSR
```

TGGTGATGTGATGCTCACCGCATTTCCTGAAAATTCACGCTGTATCTTGAAAAATCGACGTTTTTACGTGGTTTTCCGTCGAAAATTTAAGGTAAGAACCTGACCTCGTGATTACTATTTCGCCGTGTTGACGACATCAGGAGGCCAGTA

>yiaA

AAATTGGCCAATACTGGATATTATTTGGATCTTTTGCGCGACGTTGGCAGCGAGTTTGTTACTTTCTATGCTGGTACAACGAATCGACAGAAACAGATTAGTGAGTTAAGTAAAAGCCCGGTCACATTGGACTGACCGGGCTTACGTGAGT

>yiaB

CTCTGCATTCCATAGCCCTAACAGATAGGTAACGATACCACCAACGAGAGCTATCCATGACACAATACTAAAGGCCGGTGAATAGGTTGATATTTTGTTGTCCATCACAGTATTCCTTTCATTCCTGAATATGTAAGAGCTTTATGTTGCT

>rhaT

TGCGCAATATCTTCTCCAGCATAGCCGCCTGTTAATCAGTGATATTTCGACCGAATGTGGCTTTGAAGATAGTAACTATTTTTCGGTGGTGTTTACCCGGGAAACCGGGATGACGCCCAGCCAGTGGCGTCATCTCAATTCGCAGAAAGAT

>gspCDEFGHIJKLMO

CTTAATCATAAAAATAATGTATGTATTTATTAATTAAATACATATTACTAATATAAATAAATTTTGCTTGATTCATGCAAGCGGCATTAATACTATTTATACTAACGTCAATATACAACCCACCCAATCTATTTTATTAGAACATACATCG

uidA_b1617_at:

>ydcO

GCTCGCAACATTTTAGCGGCGGGGCACGCCGTTCTTGCCTGTGGAGAGATGGTGCAGTCAGGCCGCCCGTTGAAGCAGGAACCCACCGAAATGATTCAGGCGACAGCCTGAACGTAGCAGGGATCCACGTCCTTCAGGGCGTGGAGGATGT

>torYZ

TGCCGGATGCGGCGTGAACGCCTTATCCGGCCTACAAAACCTTGCTAATTCAATATATTGCAGGGACTATGTAGGCCTGATAAGCATAGCGCATCAGGCAGCTTTACGTTTGCATAACCTCAGCGCCCGTTTCCGGGCGCTATTCACGTCT

>uidABC

TACGCTTATCAGGCCTACCTTTTCCTGCAACACTTTGAATTTATGAGTTTTTGTAGGCTGGATAAGGCGTTTACGCCGCATCCGGCATAAAAAACGCGCACTTTGTCAACAATCTGAAACGCCGGAGATTTTTCTCTCCGGCGTTATTTTT

>rnt-lhr

ATTGAGTTAATCGAAGAGAAAGACGCCGGTCGCGGTCTGGGCAACTAATCTCCTGCCGGGCGTGAACTCATCGCGCCCGCATCTTTACTGCATCGACAAGTAATATTTGTCATAATGCGCGCTGCAATTTATCCGTATTAAGAGAATCAGA

>paaABCDEFGHIJK

ATAATGATTTATAAAAATAGGGTGCGAAATCCGTCACAGTTCAAACATACAAAATTTGTGATTTTACTTAACTATTGTGTAACTTTCATAAAACAATGTGATTCGTGTTTTTAATTAATTCACGAAAACTGGAATCGTAAAGGTGATGACG

>mdtQ

CAGGTAATACCAATATCAAACCCCTGCTGCGCCAGTAATAACGCGCACTCTTTGCCGATCCCCGAATCGGAGGCGGTAATAATCGCAACCTGTGCCATCGAGTTCTCCACTTAACGCTGAATAAACGTTAAGTATAGAAGGCGCATATCAT

>ydiFO

AGGCATAAGCAATAATATTTCGGCGGGAACACCCTCCCCGCCGAACTAAAAAATATATTCAATCGTATTTAATAAAAATATTTCGTGAGTCTCTGTGCGCTAATTCTCCATTTGGCGTAGGGAAAATCACATCTGAATCAGGAATTAACAA

>yneJ

CCCGTGGCAGGATTTATCGAAATTGCATGAGTTGCCGGAGTAATGGTCATCGGGGTATCTCCTTTATGAGTCATGGTATGAAGATACGCAGATTTACTCTTGCTTTAAAATGAATAATATTAAGCCACTTATTCACGAATCGAGAATGCTA

>yeaVW

TTTTTATTCGTTCTTTGCAGTAAATAACCTGCGTCATTTCACCTTTTATTGTTTCCGTTTCGTGTTTTATGGCTTTCCGTATTCTTAATTGTTTAATTTATGTAACATGCAAATTTTGTTACGCGTACGTTAGGTTCCGCCGTACAGGTAT

>ydaSTUVW-rzpR

TCGGCCTAAGTCTTTACCACTAAGCATTGCTTAATATTCTCCTATGCGCATTACATTAGGCAATCCCTACCCTTACTGCATTAGGCACAGCCTATTGACAATTGCGTTAGGCGTCGCCTAATATTTCTGTGTGTTTTTGGAGTTCATTCGA

>maoC

CTGCACCTGCCAGGTTGTTTGGCAGGTGTGCCAGCTTTTCATACAGTGGATGCCCTGAAAATAGATGTACACATCATGCATAATGTGACAACGTCACAAAACTTAGTGAAATAAAAGGGCAACTATTCGCCGTTGCCCTTCATTCACCGAT

```
>ompG
TGCTGGTGCGTTAAATGATTATCATGCAGGAGAAAATATCACTATTCATTTTGATATGACGAAATGTCATTTCTTTGATGCAGAAACGG
AAATAGCAATTCGCTAAATACAGGGGGAAGGCATTCCCCCAGGATAATACAAGGAACAATAA
>pscKG
CAGAAATGCTATTGCCAGTAACACCACCATCCCCACAACACTTCTCATTATATCCATAATGATTTTCCCTTCATGCCGGTAAACCCGGCG
TCAGCGCCAGGTTTTGGTATGCTTGATGAGTACGGGCGACGGCTTTCTGCCCGTCAGAAAT
>yfaXWVU
AACAAGCCATTAAAATTGCTTTGCGCATGCTGGAACAGGGCTTTGATCGTGACCAGGTGCTCGCGGCCACCCAGCTAAGCGAAGCCGAT
CTGGCAGCGAATAACCACTAATTAACACAGGCCCACAGCCGATCCCCATGGGCCTTTGATAT
>feaB
AAGTCTTGTCAGGCATAGAGACATAAGCGGTTATTGTCACGATTTGCGGAGCTTGTCACAGCTGACAAAGCGAATGTCACAGCGAAAAA
AGTGACTTTTCTTGTCGCTGCGTACACTGAAATCACACTGGGTAAATAATAAGGAAAAGTGA
>narZYWV
AACAACACCGGCAGAATTGCCGCGAAACGGTTGTGAGGTAAAAGCATCGACGTGGTACACCTGCGGTTTCATTAACGTTCTCCTGTGAC
TGGAGAACTATCATAGCCTGCAAGTGGCCGGAGAGCGAAGGGCTATCCGGCCAGGGTGAAAT
>alkA
CTGGATTTGCTGGTGAATAGTAGTTATTTCGCGATTAACTGAGGCGTGCTTCCCCATCGCCTGATGCGACGCTAACGCGTCTTATCATG
CCTACAAATCGCTCATTCCCCAGGCCGGATAAGGCGCTCGCACCGCATCCGGCGACCAACGT
>fdnGHI
GCGTTTTTCTACCGCTATTGAGGTAGGTCAATTTGCGAAGGCGGATTATTTTGTGGCAAACAGATGTTCTTTTTGATTTCGCGCAAAAA
GATTCAGAATTTTACTGTTAGTTTCCTCGCGCAGTAATACCCCTGAAAAAAGAGGAAAGCAA
>lomR_2-stfR-tfaR
TTTTCGGGTCTGACGGCGCTTAGTGCTGAATTCACTATCGGCGAAGGTGAGTTGATGGCTCATGATGTCCCTCTGGGATGCGCTCCGGA
TGAATATGATGATCTCATATCAGGAACTTGTTCGCACCTTCCCAAGGGGAAAACGCACGACG
>ycjMNOPQRSTUV-ymjB
CCCATAGATTATTTGCGTCAGCTCACAAATACGCTTTTTCCCTGGTAAAAAATGATTTCCTGCGTGACTAAAACCCTTGTGCTCAATTGA
CAGTTTATTTTCTGCGGAGTAGTCTCTCGTTTCATGGGACCGCTACCACGGAAAGGCAACA
>recET-lar-ydaCQ-intR
GAAGTCAGCAATGGCTTCGCCCACGTTACGACGCAGACGTTTTTGTAATTTGTTCAGGTTGTATTGTTCTTTCTTTGTAATTTGTTGAT
TTTCTTGCATTATTTCAGTTCTCTGGTACTAAATGGGGCAAATTGGGGGCAAACTTTGCAAC
>astCADBE
GAGAGAATATTACAACACGATGATTTTGCAGAGATTATGAAGAACTATACCGGATGACTGGTGATAAATAAAGCAAATAACCAGGATTA
ATCTGTATTAATTTATAAGAAAGCAACTTAATACCCGCAGAATGATTTCTGCGGGTAAGTAT
>ydiQRST-fadK
TTGTTTTAGCGATCGCATTTTTTTTGCAAGATTGTTGGCAAGGAAAACAGCTTGCTCCGTCGAAAACCCCGCACCGCTATCGCACACTA
TTTTCAGGCCATTTTTACCTTCCATCGGAGATGGTTCCGTATGCGACTCACAGGAGAAATCA
>yebQ
AAAAATCAGAACTGTTTTTTATTATAATTTCGCACCAGGGTGGTCGCAATCCATCTTTTGCCGGTTAGTTACAATTCTGCGACATCCACC
GTGAATATCAGTGCTAGAATCATACCCCTGTTGATTATTCACCAAAGATATAAAATTCCTA
>nudG
CAAGAACCCGCACGCACTGATGCAACAACGGTTTAATCCGGGGATGAAGATGGCGGTCTAGCACAGGCACTCCTTAAATATAAAGCCTT
TCTGATTGAGCAACAGTGCGGATATTATGGCATTTTTCGCTTATCTGCCCGTGTGTAATTTA
>ydjXYZ-ynjABCD
GCTCAAACTGGGTCAGGAGAATTAACCTTGAGAAAAATCAACAAACTGTCAGTAATGATTTGTTGCCTGCCGTCCTTTGTTATACCGTC
TCTGCGTTTTTAGTTGTCTGACCACTTCTCTATTATCAAGTTTGATATAGGAAACTCCACGA
>tar-tap-cheRBYZ
CGTTCAATGGTTTGAGTAAGGGGCAAAACAGGCGGGATTTAGGGCTTTTGCTGCCACACATCAAGCATAGTGTGCGTTTGTCGGATGCG
GCGTCATCGCCTTATCAGACCGCCTGATATGACGTGGTCACGCCACATCAGGCAATACAAAT
>rspAB
```

GAAATCCATAAATTCAAGCGCAGTGCCCAGCCATCCCGATACTGCTGCTTTCACCAAATCCTTAGTGCTTCTTTCGTGTTTTTCTATTGTCATAATGGTTATCTCTAAAAAAGAGGTAAGATGCGTACTACTTACTCGCCGTTATTGGTAT

>ydeE

TGAATTTACTTTTCTTTAACAGTTGATTCGTTAGTCGCCGGTTACGACGGCATTAATGCGCAAATAAGTCGCTATACTTCGGATTTTTGCCATGCTATTTCTTTACATCTCTAAAACAAAACATAACGAAACGCACTGCCGGACAGACAAA

>yebST

GTAAAATTCTGTTTTGTTCATTATATAATCACTTGGTTGTCTTACCTGGATCTGCCAGCCTATTAAAATAAGCATTAAATGCGTTAATGCTCAAGATCATTCCCATCATGGGTTAAGATTAATGTTAATTCTTATTACATTTGGCACGTCA

>yeeP

ACTTATTGCTCTCGCGTAAGCGGGTACCGTGACATTCTGCCTGAACTTGACCTGGTACTGTGGCTGATTAAAGCCGATGACCGTGCCCTGTCTGTGGATGAGTATTTCTGGCGACACATCCTGCAGTGCGGACATCAGCAGGTGCTGTTTG

>lsrACDBFG-tam

GCATGAACAAACGCAACCGTGAAAATCAAAATAGCATAAATTGTGATCTATTCGTCGGAAATATGTGCAATGTCCACCTAAGGTTATGAACAAATTAAAAGCAGAAATACATTTGTTCAAAACTCACCTGCAAAACTGAACGGGGGAAATA

>ycbRSTUVF

TCGCTACGGGGAAAGCGACTTCAGGTAATGTGAATGCGGTAACAAATTTCCATATTAACTATTATTAATAGAACTCATTAATTGTTTTATTAATTAGTACCCCTCCAGTGTTCTGGAGGGGATATTCATATTTTTTAAGAGTGACTATTTA

>anmK

CAGCCGGAGAGAAGGACCGGCAAAATGATAATTAACAGTTTTTTCATAGTCATATCCCGAAGACTTTCCTGGTCTGGAGGGCAATACGCCCTCCCTAACGTTCCAAGTGTAACGGCAGACGCGGTAAGAAAAAATTCAGTTAACTCTGATAT

>yehLMPQ

TAGTAAATAAAAGAACCTGCCTCACCAGCAGGTTTTTTTATTTACTGTGATCTGCTTTCCAGATATTTTTCGCTCAAACAACTAATGCGCCAAACATTTATTGCGCGTAAAATATCGTTTATTTCATTAATACATTTCAGGGATGAATATA

>wcaCDEF-gmd-fcl-gmm-wcaI-cpsBG-wcaJ-wzxC

ATAAGGCGTTCACGCCGCATCCGGCATTCAGTGCCTGATGCGACGCTGACGCGTCTTATCAGGCCTACAGGTCCCGAGCACAGAACCGTAGGACGGATAAGGCGTTTTACGCCGCATCCGGCAACCGTTGTCGGAACCGAAAACAGCAACT

>ydhL

CAGCTGGCGGGCGGACATATTCCAGTCCATCAATCGCCAGTAGCCCATCACAAAACGGGAAAACTCCGGGCCTTGCGGCGCAATAGTAATACGCTGAACCATAATCGCTTCCTCTTATCAGATATGAGAGGAGTATACGCAAGATTAGGTT

>narGHJI

AAGGGGTATCTTAGGAATTTACTTTATTTTTCATCCCCATCACTCTTGATCGTTATCAATTCCCACGCTGTTTCAGAGCGTTACCTTGCCCTTAAACATTAGCAATGTCGATTTATCAGAGGGCCGACAGGCTCCCACAGGAGAAAACCGA

>ydcSTUV

TTACAGACGGCATAATGCGCGGTAGCTCACAACCTGAATAAATTTTCTCAGGGGCGAAGGTGTGCCTGCAAGCCGCCGTCTATGGTTAAACAAGGAGATATTTTTACGGCACGGCGGCTGAACAATTAATTACGACAGGAGTAAGACCTTA

>ydfJ

GCCAGGTTATTAAGATCATAAACAGGGAGTGTCGCTTTTGCTGATAACAAATTATTTCCCATAACAATTCCTTAAATATAAATATGGCAAGCTATATGTTTTGTTATATGAATAAAAATCCCCTCTCCGGTAAGAGAAGGGATTAAGGGTT

>yeeJ

CCATTTTCCCCTCGATTATAAAACTTGAGTTATTCAGTAGTCTCCCCTCTTGCAACTCACACCCAAAACTGCCTAACGAAAAGTTATTAATTTTCAATCATATTGCTATCAGTATTTACATTTTTTCGCTGTGCTAGAAAGGGCGCATTTA

>yeaX

AGGGGCGCATCATGGCCGACAGTAGCGGTAGTGGCATTTCCGAACATGGTATCGCCCATTTCCATAATCTGCTGGCGCAGGTGTTTAAGGACTAATGACATCGGCGGCGGTATTTTCCGCCGCTGGGCTGATTTTTGATGGAGTACAGCAA

>trpLEDCBA

TATATTACTGTTGGGCGGAAAAATGACGTAAGTTGACGTTCGACCGGGGTAAGCGAAACGGTAAAAAGATAAATATTAAATGAATTTAGGATTTTTCCGGCTTCATTAAAGAAAGTTAAAATGCCGCCAGCGGAACTGGCGGCTGTGGGAT

>ydcW

CTCAATCAGCTTGGGCGACCGCGTGATGTACCGGTAACTAACGTGGTGGCACTGCTGGTTATGTTGGTAACAACCTTGCCGATCCTGGGGGCCTGGTGGCTAACCCGCGAAGGCGACAATGGTCAATAACCACTGATACAGGAATATGCTA

>pinE

ATTTCAGCCAGCCTGTTGGCGGAGTGGCTGAAGGCCACGGAACCGGGACACCAACAGGTAATGCAGAGCCTTCTCCCAAACCAACGTTTATGAAAATGAAGAAATAACAAGCAAATGGCATCATTCCTGCTTTTACCAGGGGGATTTAACA

>clcB

GCAATAATAGGTTACAGTGTCACGTTTTTTTATCTCTTAAAGCACGCACTGCTTTTGCGGCTGGCCTCTTTTGCCGCAAAATAGTCGCCCGTGTTTCATTGCCCATTTCTGCTCATGCATCATCTACACATCTATCCGGATCTGCGCACTA

>flhBAE

TTTTTATCCAAGCCCTTTGACAAGAGGATAATTCACATCTTTTTGGCATGTTTTGTTGCAAGCTATTCCTGATAAATAATTGCAACAAGACATCGAGCCTTTTTCACTGAGTTATTAAACATACTCGCGAGCGCGTAATTTTTTTGTCCTT

>tynA

GGTGGGATGAAGCAGTCAGGAACGGGCCGTGATTTTGGCCCCGACTGGCTGGACGGTTGGTGTGAAACTAAGTCGGTGTGTGTACGGTATTAATCTGGTTCGCTCATAAGTAAAAAACGGCACCTGGTGCCGTTTTTTTGTCTGAAACAAT

>abgABT-ogt

TTTTAGGGAGCAAGTAAGTCTAAGCAAACCTTAACAGCAGAGAATTCCGATATTAGATGTAAATATATGTCTATCTATTTGAAAACCCTTAAGTTGTTAAGGGTAACTTTACATAAAAGTGTGAACAAGCTGGCACAAATTGTTTAATGTT

>yddAB

TTAACTTTTCATCCTGCGGTAAGGCGGCGGCAATCAGCCGCCCGGGGAGCAACAGAGTTGCCACTAACGTCAGTAAGAAACAGAGGTTTCTCATAATTATCTCCATGCGAAAACCGGGCGAATTTACCCGGTTAAGTAAAATCCGAACTAT

>ycgV

AAAATGAAGCAACATGGCATTGATGCAGGGCAGTTTAAGCGCCGGGTATGAAAGACAGAAACGATTTCTGATACATCAGAGTGATCTGTATTTCATTCCGGCGCACGCTAACAATTTTCAGCATCGTTTAAGGGCTTGTCTATCCCGCACT

>chiP-ybfN

CTGGTTACAGAAAGAGATTGATAATTCGCGTCGCGAAAAATAGTCTGTTCCTGTAGTCAGCGAGACTTTTCTCAACGCTACTTTTTTAATTTTTATTTTTTCGCTGTTCACCTTTGGTGCAGCAATTTATACGTCAAAGAGGATTAACCCA

>ydaL

ATTAATATCGTGAATGAATAATCATGCATAAGTATTTTGCTTAAAATATCGGCAATATTTGGAACTTATTACTGGAAATTTGGGTAATACGTTGTTGGACCGACCCGGTCTGGTTATCATATCGCGCTCTTAATTGCGGGAGGATGTAACA

>ynbABCD

AACTATCTGATTAATTGGGGATAATCATTCCTGACAGTGAGTCCCCAATACCTTGATATATTCTGAATTTTTAATGAAACGGCGTGTTGCGATATCTCCGTCAGGGGAATTGATGCACCATAGCGCAAACCGAATTATCAAGGATTGATAA

ycjW_b1320_at:

>yncJ

TAGTGGCATCCGTAAGTCCGGATGAGCTGCTGAAAACATTGCCGAAGCGAAAAGGGTAAAACGCCAGTTTTCTGGTTACTCACAACTTATTGAATCTGCATGATATTGCCTGCCGGGTAAGGCGTTACGCCGCATCCGGCATCAAATGACT

>yciZT

AGGTGATAATGACTTCCTGTTATATCGCTGATAATAATTTTATATCTTGAGAGTGTTAATAACAGGTAAATAGTCTTAATTATCAACCAGGAATCATCTTAGAGCGGATGATTTGCCAAACTGCAAATCATCCGTAGAGAAGGGAAATGGT

>cedA

AAATAAACAATTAAAGAAAAAAGATCACTTATTTATAGCAATAGATCGTCAAAGGCAGCTTTTTGTTACAGGTGGTTTGAATGAATGTAGCAACGAAATACAGAATTTCAGGTCATGTAACTCCCGGCAAAACCGGGAGGTATGTAATCCT

>yeaVW

TTTTTATTCGTTCTTTGCAGTAAATAACCTGCGTCATTTCACCTTTTATTGTTTCCGTTTCGTGTTTTATGGCTTTCCGTATTCTTAATTGTTTAATTTATGTAACATGCAAATTTTGTTACGCGTACGTTAGGTTCCGCCGTACAGGTAT

>ydjLKJIHG

TCTTTTGCCGCCACGATTACACCCCTGTATCTTTTTACATCACATTAGCGCGATTATCGCATAACCGATGTTTACTTTCAAAATAACCTGTTTGAATCACAGATTTTCATCACAGTTTTCACAGAAACAGAGGTGAATCGTGTTGAGTATT

>ydgJ

CCAGCTGCTGGGTTATCCCGCGAGCAAATCCGCCAGGCACAGATTAATGGTCTGGAAATGGCTTTCCTCAGCGCTGAGGAAAAACGCGCACTGCGAGAAAAAGTCGCCGCGAAGTAACAAAATGGATGGTGCAAATGCACCATCCATTTTT

>ydhO

TTTCACTGCTGCATTTTTTTGCGCTCGCCAACGAAACGTATTTTTTAACAATAAAAGCTATTAACTTTCTCTTCTTCTATGCATTAGAATCATCAAGTTTTGTAAATCAGACGCAGGCATGATAGACCTGCCTTTACAGAGGGACGCTCAG

>ynbABCD

AACTATCTGATTAATTGGGGATAATCATTCCTGACAGTGAGTCCCCAATACCTTGATATATTCTGAATTTTTAATGAAACGGCGTGTTGCGATATCTCCGTCAGGGGAATTGATGCACCATAGCGCAAACCGAATTATCAAGGATTGATAA

>uidABC

TACGCTTATCAGGCCTACCTTTTCCTGCAACACTTTGAATTTATGAGTTTTTGTAGGCTGGATAAGGCGTTTACGCCGCATCCGGCATAAAAAACGCGCACTTTGTCAACAATCTGAAACGCCGGAGATTTTTCTCTCCGGCGTTATTTTT

>yddG

TGAGAAAAATAGATGAAATAATATTATTTATCGATATGTGATCGAAGTCGAAATGAGATATAAGGTGAATTACTGGTATTTGAAATTTATTTTTTTAATATTGTCGGAATTTATCTGATTAACTACCGGGCCGTAGACCCGGCAGTTATTT

ydaK_b1339_at:

>abgABT-ogt

TTTTAGGGAGCAAGTAAGTCTAAGCAAACCTTAACAGCAGAGAATTCCGATATTAGATGTAAATATATGTCTATCTATTTGAAAACCCTTAAGTTGTTAAGGGTAACTTTACATAAAAGTGTGAACAAGCTGGCACAAATTGTTTAATGTT

>ydaM

TGATGATGTTCAGGCATATTGCACCGCGCTACCGCATCATGGCGGCAGCGGGGCGTGTTACGTCGCACTACGTAAAACGGCGCAGGCGAAGCAAGAAAACTGGGAGCGCCACGCTAAGCGCAGTCGTTGATCTCGAGACGCATCCGCGGCT

>ydaN

ACACGGTTTGTATACGGAAAAGCATTTTGCTTTTTGTATTCAATTTAGACAGAATTTTATTAATCATTTCAGGGTAATGGGGTGATGAGATGTTGCGTAACAGGGCCAGAAGGCTAGACTACAAAATAATGCGTTGATGATGGAGGCACTG

>ydaL

ATTAATATCGTGAATGAATAATCATGCATAAGTATTTTGCTTAAAATATCGGCAATATTTGGAACTTATTACTGGAAATTTGGGTAATACGTTGTTGGACCGACCCGGTCTGGTTATCATATCGCGCTCTTAATTGCGGGAGGATGTAACA

>C0343-dbpA

TTCGTACGACGCCGACTTTGATGAGTCGGCTTTTTTTTGCCTGTTATTTATCAGCGTCTACCCTTTAAGAGTCCACCCAATGACCAGAGGGAAATATGACGACACTTATTTATTTGCAAATTCCTGTCCCTGAACCGATTCCTGGCGATCC

>ynaJ

TTTCCCCGCTTATTCGCACCTTCCCTAACTAATCAATGCGTTGATTGTAAATCCAGCTAAGAGGTGAGGTTTTCAGAGCAGACAACGGTGAAATGTCATGGTATTGTTACGTTTAGGTAACAAGAAATTTGTCTGCACAAGGATTACATCA

>uspE

ATGTGATTCAGTCATCGATGATTTTTGTTTATAGCCTGGTCTTTACTGTGTTGCTGGCAATCCCGTTGGGAATTTATTTCCTTGGCGGCGAAGAGCAGTAAGTAAAAAATAGGCCCGATAACTCGGGCCTTGTCAGTTATTGAAGAGTCGT

>fnr

CAGGGTCTCCTTACAACAACTGTCAACGCAGTTTGTAATTAAAAGATTAACCCATATCTGGTGAATGAAACAGTGATGAACCTTCTGCCAGATCAATAAATCAGAAAAATTTAATGATATGACAGAAGGATAGTGAGTTATGCGGAAAAAT

>ddpXABCDF

CCTTTGACGGCATAATCCAGAAAGCAAAAGCAGGATGCCCGGTCTCTCAGGTACTGAAAGCGGAAATTACGCTGGATTACCAGTTGAAATCGTAAAGCATTGCCGGATGACGCGTCAGGCGCGTGAATGCCTGATGCGTTGTTAGCATCTC

>yeaVW

TTTTTATTCGTTCTTTGCAGTAAATAACCTGCGTCATTTCACCTTTTATTGTTTCCGTTTCGTGTTTTATGGCTTTCCGTATTCTTAATTGTTTAATTTATGTAACATGCAAATTTTGTTACGCGTACGTTAGGTTCCGCCGTACAGGTAT

>ydcSTUV

TTACAGACGGCATAATGCGCGGTAGCTCACAACCTGAATAAATTTTCTCAGGGGCGAAGGTGTGCCTGCAAGCCGCCGTCTATGGTTAAACAAGGAGATATTTTTACGGCACGGCGGCTGAACAATTAATTACGACAGGAGTAAGACCTTA

>narZYWV

AACAACACCGGCAGAATTGCCGCGAAACGGTTGTGAGGTAAAAGCATCGACGTGGTACACCTGCGGTTTCATTAACGTTCTCCTGTGACTGGAGAACTATCATAGCCTGCAAGTGGCCGGAGAGCGAAGGGCTATCCGGCCAGGGTGAAAT

>narGHJI

AAGGGGTATCTTAGGAATTTACTTTATTTTTCATCCCCATCACTCTTGATCGTTATCAATTCCCACGCTGTTTCAGAGCGTTACCTTGCCCTTAAACATTAGCAATGTCGATTTATCAGAGGGCCGACAGGCTCCCACAGGAGAAAACCGA

>wza-wzb-wzc-wcaAB

AGTTCTGATGGCTGACGCTCTCTTTGCCGCCTTTGCCGTAACCGTAGACAAAATGTGACGCCAGCCCCTGTTGCAGCGCACGCTGGTGGAGATCTAACGCCACACCTGCTGCCCCGCCTTCCGCCAGTCGCACATTAAATTGCAAAATATT

>ydiNB-aroD

CAAAAACAAACTATGACATGCAATATTCCTGGAAACATAAACTTTATGCCATGTACCCAGGGAAAATCATCTTCAGTATAGTAATTATGTAAACCGTCGGAGAACAATACGTACGGTAACGAAATTATCTTTCAGCAAGGAGCTGTGAAAA

>rutABCDEFG

GCGCGATGCTTATCAGGCCTACCAGAAGATTGCAATATATTGAATTTGCACTGTTTTGTAGGCCGGATAAGGCGTTTACGCCGCATCCGGCATGAACAATGCGTACGTTGTCAACAATCTGCACCGCCGGTAACCCCGGCGGTTTTCTGTT

>astCADBE

GAGAGAATATTACAACACGATGATTTTGCAGAGATTATGAAGAACTATACCGGATGACTGGTGATAAATAAAGCAAATAACCAGGATTAATCTGTATTAATTTATAAGAAAGCAACTTAATACCCGCAGAATGATTTCTGCGGGTAAGTAT

>rnt-lhr

ATTGAGTTAATCGAAGAGAAAGACGCCGGTCGCGGTCTGGGCAACTAATCTCCTGCCGGGCGTGAACTCATCGCGCCCGCATCTTTACTGCATCGACAAGTAATATTTGTCATAATGCGCGCTGCAATTTATCCGTATTAAGAGAATCAGA

>cvrA

ATGGAGCCTGTACATAGATTTGTGTAATTGCCTGATTTTGATATGTTCAATTCAACATCAAATGAAGGTTAAATTATGGACGACAAACAATTGCAGGCTCAGGCTGCGTTCAGCAAAGCATCGCAACCGGCGATAGATGCTTCATTAAATT

>flhBAE

TTTTTATCCAAGCCCTTTGACAAGAGGATAATTCACATCTTTTTGGCATGTTTTGTTGCAAGCTATTCCTGATAAATAATTGCAACAAGACATCGAGCCTTTTTCACTGAGTTATTAAACATACTCGCGAGCGCGTAATTTTTTTGTCCTT

>yceQ

TTAACATTCTTTTCATCGTAACTTACTCATTATTCTTACATTGACGACTAAGCTGCGGGCAAAGTAACGCCTTTCCGGGTGTGAACCGATGGCCTCGTGTCTAGTCGCGTCGCCAACCTCACGGTTATCGTCAGCTCAAAGAGGCGCAGAG

>ttcA

CCGGGTGCGGTTATTAAAATAATGAAATGTTGAATTGCCGGGTGCAAGAGTAAACATCTTATTCGGGATTGCCGGATGCGACGCTGGCCGCGTCTTATCCGGCCTCCATAAGAGTAGCCCGATACGCTTGCGCATCGGGCGCTATCCTGGT

>flxA

GAACGTTTACCCCTGTCAACAAGCTTTACTTTCTGAGGCGCGCCAGCCCGCGAGGAAAACAATCTGAACATCAAACAATTAATGACACAAGAAATACGATTAAAGATTTTTTTGTGCATGCCGATAGTGCTTTTTTAAAAGGAGAAATCTA

>tar-tap-cheRBYZ

CGTTCAATGGTTTGAGTAAGGGGCAAAACAGGCGGGATTTAGGGCTTTTGCTGCCACACATCAAGCATAGTGTGCGTTTGTCGGATGCGGCGTCATCGCCTTATCAGACCGCCTGATATGACGTGGTCACGCCACATCAGGCAATACAAAT

>mdtQ

CAGGTAATACCAATATCAAACCCCTGCTGCGCCAGTAATAACGCGCACTCTTTGCCGATCCCCGAATCGGAGGCGGTAATAATCGCAACCTGTGCCATCGAGTTCTCCACTTAACGCTGAATAAACGTTAAGTATAGAAGGCGCATATCAT

>ydcW

CTCAATCAGCTTGGGCGACCGCGTGATGTACCGGTAACTAACGTGGTGGCACTGCTGGTTATGTTGGTAACAACCTTGCCGATCCTGGGGGCCTGGTGGCTAACCCGCGAAGGCGACAATGGTCAATAACCACTGATACAGGAATATGCTA

>ydiJIH

ATAAAGCTAACCCGCCGTTTTAACACAAACTGCGATTAGTATTATTTTTGAACAATATCAGGCGGTAGATAAGCAGTATTAAGAAGGTCATCGAACCTGGACGGAGGTTAATCCAGGTCGATTTGGCGAACTTGCGGCATTAAGTCAGGAT

>trpLEDCBA

TATATTACTGTTGGGCGGAAAAATGACGTAAGTTGACGTTCGACCGGGGTAAGCGAAACGGTAAAAAGATAAATATTAAATGAATTTAGGATTTTTCCGGCTTCATTAAAGAAAGTTAAAATGCCGCCAGCGGAACTGGCGGCTGTGGGAT

>lsrACDBFG-tam

GCATGAACAAACGCAACCGTGAAAATCAAAATAGCATAAATTGTGATCTATTCGTCGGAAATATGTGCAATGTCCACCTAAGGTTATGAACAAATTAAAAGCAGAAATACATTTGTTCAAAACTCACCTGCAAAACTGAACGGGGGAAATA

>motAB-cheAW

AATCATAAGGCACCTTCCTGAAAACAAGTTGATCTCGTTATCGGCAAGGAGGGGGGAAACTTTATTGCTGATGCCACCCGCCGCGAAATTGAAATAAAAAACCCGATGCGCAGATCATCGGGTTCATTTCAATTGAGGAAATCGGGAGAAT

>ynfEFGH-dmsD

TACAACATAATCCCACCTTATTACTCATACCCTTCTATTGATATGGATTAATAATTCTTAACCCAAAATGGGTAGACTCCCTCTATTGTTAGCGCGCTAAATATTCAATATATAAACTTTTATATAACGATAAAGAACAGGGAGTGAGTTA

>ydfZ

TGAGAGTCTTTTCCCCCGCCTTATGGCTCATGCATGCATCAAAAAAGATGTGAGCTTGATCAAAAACAAAAAATATTTCACTCGACAGGAGTATTTATATTGCGCCCGTTACGTGGGCTTCGACTGTAAATCAGAAAGGAGAAAACACCTA

>flgKL

ACTCACCAACATGATTCAGCAGATGAAATCGATAAGCGACAAGGTGAGCAAAACCTACAGTATGAACATTGATAATCTGTTCTGAATAACTCAAGTCCGGCGGGTCGCTGCCGATAATACTCTGTAATTGAAGGCTTATAAGGAACCTCCA

>ynbABCD

AACTATCTGATTAATTGGGGATAATCATTCCTGACAGTGAGTCCCCAATACCTTGATATATTCTGAATTTTTAATGAAACGGCGTGTTGCGATATCTCCGTCAGGGGAATTGATGCACCATAGCGCAAACCGAATTATCAAGGATTGATAA

>torYZ

TGCCGGATGCGGCGTGAACGCCTTATCCGGCCTACAAAACCTTGCTAATTCAATATATTGCAGGGACTATGTAGGCCTGATAAGCATAGCGCATCAGGCAGCTTTACGTTTGCATAACCTCAGCGCCCGTTTCCGGGCGCTATTCACGTCT

>flgBCDEFGHIJ

GCCCCCAGCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAACGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATTTAAGCTGATGGCAGAATTTTGATACCTGCGGAGGAGATA

>ynfM

GCAACAGCAACAAAGTAACGCAGATGACGAAGTTCAATATTCATATTTAAAACATCTTATTTGAGATTATTAATATATTAGACAGAACAATTCGATTTTCCTACCCTATGTATAAGCCTGATCTACAGGCATATTTAGCAAGGATTTCAAG

>ydcF

GATGAATAATGTTTATCTACAGCATTTCCTTAAAAGATATGTCAGGCTTGCGGAGTGGCGGTTAAGGACATACGATTTCCTCCTTTCAGAGTGCTCCGCTTCTCACTATTATCTCACGCAGTATTCTTAAGGGAACGATAAGGAGGAACCA

>gdhA

TATGAGATTACTCTCGTTATTAATTTGCTTTCCTGGGTCATTTTTTTCTTGCTTACCGTCACATTCTTGATGGTATAGTCGAAAACTGCAAAAGCACATGACATAAACAACATAAGCACAATCGTATTAATATATAAGGGTTTTATATCTA

>ppsA

ACCGCGAACTACCTCAGGTAAAATACAAAAGTTTTTGTTAAGAAAAGATATTATGCGGCGTTTAACGCAGGATGTCTGTGAAGAGATTGAATAAGTTTCATCTTCGGGGATCACATAACCCCGGCGACTAAACGCCGCCGGGGATTTATTT

>uidABC

TACGCTTATCAGGCCTACCTTTTCCTGCAACACTTTGAATTTATGAGTTTTTGTAGGCTGGATAAGGCGTTTACGCCGCATCCGGCATAAAAAACGCGCACTTTGTCAACAATCTGAAACGCCGGAGATTTTTCTCTCCGGCGTTATTTTT

>secG-leuU

AAAATTTATAACTAAGGCGTACCGGCACCATCGTTTCAAGGTACCAGCTACGAGTAAAGCAACTGGACGAGATACAGATACCTGACAACCATTCCTCAGACCAGGACCAAAACGAAAAAAGACGCTTTTCAGCGTCTCTTTTCTGGAATTT

>yebST

```
GTAAAATTCTGTTTTGTTCATTATATAATCACTTGGTTGTCTTACCTGGATCTGCCAGCCTATTAAAATAAGCATTAAATGCGTTAATGC
TCAAGATCATTCCCATCATGGGTTAAGATTAATGTTAATTCTTATTACATTTGGCACGTCA
>ynjH
CGGGGCGGATTATCCATCTTCATGCCTGGCACGTACCCGACTTCCACGGGACGTTACAGGCACATGAACATCAGGCGCTGGTCTGGTGC
TCACCTGAAGAGGCGCTGCAATATCCGCTGGCCCCTGCTGACATTCCATTATTAGAGGCGTT
>yeeP
ACTTATTGCTCTCGCGTAAGCGGGTACCGTGACATTCTGCCTGAACTTGACCTGGTACTGTGGCTGATTAAAGCCGATGACCGTGCCCT
GTCTGTGGATGAGTATTTCTGGCGACACATCCTGCAGTGCGGACATCAGCAGGTGCTGTTTG
>nac
AATCCTGACGTGCAGCCTCGCGGATTATCTTTAGTTGTTGGAAATTCACGGTAAACTCCGGGCAGTTCAGATTTCCCGTTATTGTTAAA
GTCTAATGCCCGGCATAACAAATAATAAAAACCCGCATCTTATTCCATCCCGATATAACACT
>wcaCDEF-gmd-fcl-gmm-wcaI-cpsBG-wcaJ-wzxC
ATAAGGCGTTCACGCCGCATCCGGCATTCAGTGCCTGATGCGACGCTGACGCGTCTTATCAGGCCTACAGGTCCCGAGCACAGAACCGT
AGGACGGATAAGGCGTTTTACGCCGCATCCGGCAACCGTTGTCGGAACCGAAAACAGCAACT
>bioA
GAAATATTATCGCCATCGTAACCCATGCCGTTAAAGACATGACGATGCGGCAATTTATCGCCATCGCGCAGATCGTTACTGATGAGTTT
CATGAACCCTCCTTTCTTGTTTGCAGAAAGTGTAGCCAGAAACCCTCACGCGGACTTCTCGT
>yncJ
TAGTGGCATCCGTAAGTCCGGATGAGCTGCTGAAAACATTGCCGAAGCGAAAAGGGTAAAACGCCAGTTTTCTGGTTACTCACAACTTA
TTGAATCTGCATGATATTGCCTGCCGGGTAAGGCGTTACGCCGCATCCGGCATCAAATGACT
>rspAB
GAAATCCATAAATTCAAGCGCAGTGCCCAGCCATCCCGATACTGCTGCTTTCACCAAATCCTTAGTGCTTCTTTCGTGTTTTTCTATTGT
CATAATGGTTATCTCTAAAAAAGAGGTAAGATGCGTACTACTTACTCGCCGTTATTGGTAT
>bioBFCD
AGGTCATGGATGTGTATGGGTGCCAGATATGGCGTTGGTCAAAGGCAAGATCGTCCGTTGTCATAATCGACTTGTAAACCAAATTGAAA
AGATTTAGGTTTACAAGTCTACACCGAATTAACAACAAAAAACACGTTTTGGAGAAGCCCCA
>yedL
TGTTAAAAACCAGGGCGCGGAGTTAATTCAACCTGTTTGATAAGCCTGGAGATTATTGATCAACAATACTGCGTCATAAGAAATCTCTA
TTAGACAAAGATTTCATTACCTGTTGGCATATTGCAAAAATAACACCAATACGGAATCGTCA
>nudG
CAAGAACCCGCACGCACTGATGCAACAACGGTTTAATCCGGGGATGAAGATGGCGGTCTAGCACAGGCACTCCTTAAATATAAAGCCTT
TCTGATTGAGCAACAGTGCGGATATTATGGCATTTTTCGCTTATCTGCCCGTGTGTAATTTA
>flgAMN
TGGCGTTTACTGCGTTTAATGGCGGTCGTGCTGGCGGGGATTGCCGCGTACTTCGCTGCACTGGCGGTACTGGGCTTCAAAGTTAAAGA
ATTTGCCCGCCGGACGGTGTAACAATGCATTCCGGCCTGCAGTGCAGGCCGGAGATAATCTT
>yneJ
CCCGTGGCAGGATTTATCGAAATTGCATGAGTTGCCGGAGTAATGGTCATCGGGGTATCTCCTTTATGAGTCATGGTATGAAGATACGC
AGATTTACTCTTGCTTTAAAATGAATAATATTAAGCCACTTATTCACGAATCGAGAATGCTA
>ydhYVWXUT
GTTCGGCGCGTTGCCCTGGTTGATGCACGACCAGCCTTTCGTTCAGGGACTTACGGCATGTGCAGAAATTGGTGAGGCGATGGCTCGGG
CGGTCGTAAAGCCTGCATCCCGTGCTCGTCGGCGTCTTTCGTTTGATTAAAGGTAAAGCTAT
>clcB
GCAATAATAGGTTACAGTGTCACGTTTTTTTATCTCTTAAAGCACGCACTGCTTTTGCGGCTGGCCTCTTTTGCCGCAAAATAGTCGCCC
GTGTTTCATTGCCCATTTCTGCTCATGCATCATCTACACATCTATCCGGATCTGCGCACTA
>ydiQRST-fadK
TTGTTTTAGCGATCGCATTTTTTTTGCAAGATTGTTGGCAAGGAAAACAGCTTGCTCCGTCGAAAACCCCGCACCGCTATCGCACACTA
TTTTCAGGCCATTTTTACCTTCCATCGGAGATGGTTCCGTATGCGACTCACAGGAGAAATCA
>mliC
```

GGATCGGCGGGAAGATCGCGGCGGCGTAACCCGCCTTTGGTGTATTCACGGCGCAGATGCGCGATTTGCTGCAATTCGTCGTTATCAGACATGGTTTTCTTTACGGATTGTCAGTGGGTGACGCTATTGTGCGCCGCCCCTGGAAAAATCT

>nhoA

TTTTTTAAGCGTAGTCCGTAACGCAATAAGTAACGAAATTAACGGGATTGGCGATTTGCGAACGTGATGCATGTCCGCGATCGCACAAAATAGCCGGTGCGGCGTCTATTCCAGGTTATAAGTTGAGAAAACCACTAAGGGAAACGCCTGA

>yddAB

TTAACTTTTCATCCTGCGGTAAGGCGGCGGCAATCAGCCGCCCGGGGAGCAACAGAGTTGCCACTAACGTCAGTAAGAAACAGAGGTTTCTCATAATTATCTCCATGCGAAAACCGGGCGAATTTACCCGGTTAAGTAAAATCCGAACTAT

>anmK

CAGCCGGAGAGAAGGACCGGCAAAATGATAATTAACAGTTTTTTCATAGTCATATCCCGAAGACTTTCCTGGTCTGGAGGGCAATACGCCCTCCCTAACGTTCCAAGTGTAACGGCAGACGCGGTAAGAAAAATTCAGTTAACTCTGATAT

>tqsA

CTCGGTAGTTGAACACGCCTGATTTGTATCATAGCTTAAGAATTAACTCAAAATATTTTCACTTCTTTACCTGAGCGGTTTGATTTTCGTTATGATGACGGAGCGAAAAAGACATTATTATTAGCAAAGGAAGAAAAAAACGGGGACAAGCA

>ydcO

GCTCGCAACATTTTAGCGGCGGGGCACGCCGTTCTTGCCTGTGGAGAGATGGTGCAGTCAGGCCGCCCGTTGAAGCAGGAACCCACCGAAATGATTCAGGCGACAGCCTGAACGTAGCAGGGATCCACGTCCTTCAGGGCGTGGAGGATGT

>asr

TTATTATCAACGCTGTAATTTATTCAGCGTTTGTACATATCGTTACACGCTGAAACCAACCACTCACGGAAGTCTGCCATTCCCAGGGATATAGTTATTTCAACGGCCCCGCAGTGGGGTTAAATGAAAAAACAAATTGAGGGTATGACAA

ydhB_b1659_at:

>ydjLKJIHG

TCTTTTGCCGCCACGATTACACCCCTGTATCTTTTTACATCACATTAGCGCGATTATCGCATAACCGATGTTTACTTTCAAAATAACCTGTTTGAATCACAGATTTTCATCACAGTTTTCACAGAAACAGAGGTGAATCGTGTTGAGTATT

>ydfJ

GCCAGGTTATTAAGATCATAAACAGGGAGTGTCGCTTTTGCTGATAACAAATTATTTCCCATAACAATTCCTTAAATATAAATATGGCAAGCTATATGTTTTGTTATATGAATAAAAATCCCCTCTCCGGTAAGAGAAGGGATTAAGGGTT

>ynbABCD

AACTATCTGATTAATTGGGGATAATCATTCCTGACAGTGAGTCCCCAATACCTTGATATATTCTGAATTTTTAATGAAACGGCGTGTTGCGATATCTCCGTCAGGGGAATTGATGCACCATAGCGCAAACCGAATTATCAAGGATTGATAA

>ydiNB-aroD

CAAAAACAAACTATGACATGCAATATTCCTGGAAACATAAACTTTATGCCATGTACCCAGGGAAAATCATCTTCAGTATAGTAATTATGTAAACCGTCGGAGAACAATACGTACGGTAACGAAATTATCTTTCAGCAAGGAGCTGTGAAAA

>ydgJ

CCAGCTGCTGGGTTATCCCGCGAGCAAATCCGCCAGGCACAGATTAATGGTCTGGAAATGGCTTTCCTCAGCGCTGAGGAAAAACGCGCACTGCGAGAAAAAGTCGCCGCGAAGTAACAAAATGGATGGTGCAAATGCACCATCCATTTTT

yfeG_b2437_at:

>yfdONMLK

TGATTGATACTGCAATGGCTTCCATTAGTCTGATTCAACTGAAATTACAGGCTGGGCGGAAGCTGATGCAGGCAGAGACCTCCCGACTTAACACTGTGCTGGATTACATTGACGCGGTGACGGCAACAGATACCAGCACCGCGCCGGATGT

>rutABCDEFG

GCGCGATGCTTATCAGGCCTACCAGAAGATTGCAATATATTGAATTTGCACTGTTTTGTAGGCCGGATAAGGCGTTTACGCCGCATCCGGCATGAACAATGCGTACGTTGTCAACAATCTGCACCGCCGGTAACCCCGGCGGTTTTCTGTT

>yeaVW

TTTTTATTCGTTCTTTGCAGTAAATAACCTGCGTCATTTCACCTTTTATTGTTTCCGTTTCGTGTTTTATGGCTTTCCGTATTCTTAATTGTTTAATTTATGTAACATGCAAATTTTGTTACGCGTACGTTAGGTTCCGCCGTACAGGTAT

>wza-wzb-wzc-wcaAB

AGTTCTGATGGCTGACGCTCTCTTTGCCGCCTTTGCCGTAACCGTAGACAAAATGTGACGCCAGCCCCTGTTGCAGCGCACGCTGGTGGAGATCTAACGCCACACCTGCTGCCCCGCCTTCCGCCAGTCGCACATTAAATTGCAAAATATT

>hcp-hcr

TGCAACCGTTTGTTTCATGGTTCTCCATCTCCTGAATGTGATAACGGTAACAAGTTTAGTTCATCTGACGGAGGGGGAAGGGATGGGAGAGAAAGGAGGCACTAACGGTTAAATAGCCCGATGAAAGGAATATCATCGGGCATAAGGCGAT

yheN_b3345_at:

>yheONML

GTACCAGCTGGTTAACTGTTGCCATTAAATAGCTCCTGGTTTTAGCTTTTGCTTCGTAAACACGTAATAAAACGTCCTCACACAATATGAGGACGCCGAATTTTAGGGCGATGCCGAAAAGGTGTCAAGAAATATACAACGATCCCGCCAT

>kdsC-lptC

TGGCATTCTGGCCGTTGAGGCACTGAACTTAATGCAGTCCCGCCATATCACCTCCGTGATGGTTGCCGATGGCGACCATTTACTCGGTGTGTTACATATGCATGATTTACTGCGTGCAGGCGTAGTGTAAAGATTCAAGGATAAACAACAA

>rfaD-waaFC-rfaL

TGAAGGACTAGCTAAAACCCAAACTAGTTTGTTGCAATTAGCATCCTTGCACCTCTATGTAAAGGGCTGAAGGGATTCGGATGTGATGGTATGATTACAGACATTCGTGTCTGAGATTGTCTCTGACTCCATAATTCGAAGGTTACAGTTA

>rsgA

AAGCCATAGTTTCGGCAGAATGTACTGTAGCGAAAGTTTAAATGAATTTAACAAGGTAGCCTCCAGGCCATTGTTTTGTCGTTCCTGATCCGGCCTACATGCCGGATCCTGAAAAAAAGGGGACGATTCTAACGACGGTTAGCTTAATTGT

>nudF-yqiB-cpdA-yqiA-parE

TCTTGAAGCGCATCTGCAAACCCCGCACATGAAGGCGTATAGCGAAGCCGTAAAAGGTGACGTGCTGGAGATGAATATCCGTATTCTGCAGCCAGGGATTTAATCCTGCCTTGTTTGCCCGGCCATCCTGACCGGGCAATGTTCTTTCCTT

>psd-yjeP

TCCCTTTTCTCTGGTTCGCTTTGTGGCGAAAAAAACGCGGCTGGTTTATGTACGCCACCGCGCTGGCTATTTTCGGCTACTGGCTGTGGCAGTTTTTTCTGCGCTATCAGTTTTGTTTGTGAGCCGGATTGGTTCATCCGGCACACAAACT

yheO_b3346_at:

>yheONML

GTACCAGCTGGTTAACTGTTGCCATTAAATAGCTCCTGGTTTTAGCTTTTGCTTCGTAAACACGTAATAAAACGTCCTCACACAATATGAGGACGCCGAATTTTAGGGCGATGCCGAAAAGGTGTCAAGAAATATACAACGATCCCGCCAT

>rfaD-waaFC-rfaL

TGAAGGACTAGCTAAAACCCAAACTAGTTTGTTGCAATTAGCATCCTTGCACCTCTATGTAAAGGGCTGAAGGGATTCGGATGTGATGGTATGATTACAGACATTCGTGTCTGAGATTGTCTCTGACTCCATAATTCGAAGGTTACAGTTA

>aroKB-damX-dam-rpe-gph-trpS

TTTCAGAGTGATGATAAAAGCAAAATTGCCTGATGCGCTACGCTTATCAGGCCTACATTTCCTTGCAATATGTGCATTACTTTGTAGGCCGGATAAGGCGTTCACGCCGCATCCGGCATGAACAAAGCGCAATTTGCCAGCAATAGTGAAT

>lptAB-rpoN-hpf-ptsN-yhbJ-npr

ATAACGCGCAGATCAATCTGGTGACGCAGGATGTTACCTCTGAAGACCTCGTCACGTTATACGGAACAACATTTAACTCCAGCGGTCTGAAAATGCGCGGCAACTTACGCAGCAAGAACGCCGAGCTGATTGAAAAGGTTAGAACATCCTA

>yrbG-kdsD

CAACATATTTACAGAATATTACCCGCCGTGGTTAGCGAAAGCTGGCATTTGTTTTACTTTTTAGCCGCATAAAGTCAAAATTAAGCATCCGTTACGGCTTTCTGAAAATCTTCAGCGGACCGGCGAGTATACCTGAAGAAAGGACGTTAGA

yhiE_b3512_at:

>hdeAB-yhiD

AATATCGCTAAAAACGTTCTACTGCCATAAACACAATATCATGATGATCCTCAATCTTAAGCGGATCAATGAGCTGGTACGCCATCAGCATATTGATTATCTGGTGTGAATTTCAGGCTTACGGTGAGTCTGGCTACGCTGCCACACAGAT

>slp-dctR

CGCCCCAAAATTAACTGAGTTCACCTAAACAGAAAGGATATAAACATCAGACAGGTTTACGTTACTATCAGGCATATCACCTCAGAATCAGATGAAAACTATAAAGAAATATCTATTATGGTTTTAATATTTGTTGATAAGGATAGTAACA

>hdeD

AATCAGATATTTTTATTTCAATACAATGAGTTACAGATGCATCAGATACTGCAATTAGGAAATTTTTATTAAATCGACTGCATTCTTAGACGCGTTTTTGGCATAGATTGATAGCAGGGGATTTTCTTCTTAATTTTATAGGGTGGTTCTA

>gadAXW

AAACATGCCGAGTCTTTTCCTCCTGAAAATACCGATACCGCCTGGCAAACTGCGAAAGAGAACTAGCCGGTAGCCCGGCAGAAATCATCAGGGAAGAGTTTCACATGAAGCAGGTGTGAGATCCTGACCAATATTCAAATGCGAAATATGT

>gadBC

TTTCGGGAGCAGATATCCATATGTGCTGGTTTCCGGTAAACAGATGTGCGCTCATTCTCGTACTTATCCCCGTTAAGTCAATACGACAGCAAGCACGAAAAAGGGAGCGATGAATTATCGCTCCCTTGTCTTATAACCATTCAGACATGGT

ymfN_b1149_at:

>ymfTLMNROPQ-ycfK-ymfS

GCCAACACCATCTCTCAGTTTTCTGGCGTTAGACCGCCGGATGTCATGGATTGTTTTCATAACGAAATTAAAACCCTTGTACCGTTAAGGTACAAGTATCTTGAAGGTTCATTTCAATCATGTAATATGTACACCGGAGGTACATATTGTA

>ymfH-xisE-intE

TCCTTTTCATCGATTTTAGGTGATCTTCTCTGTGATATTTCACGTCTTACCAGTAACTGATATGGCTGTCCGCCGCTCGCTTAAAGTGGACTTTTTAGTTTTTATCATGTGCGGTGAGAAATTCAATGTGGCGTTGAGATGCTTAAAGGTT

>ymfJ

CTGATTCCAGAGGGATATATCCTCAAAAGTACAATGAACTGCCAACAAATCCTTGATCAAACATTTTCAATTGCTAACAGTGCCGGTGTTGACGCAAATATATTTGTCTGTAAATTTGAACAAAGCGCATGCTTACTTCCGTCTGCTTCCT

>yeeRSTUVW

GGTCAGGCCACACTGAATGTGACCTTCTGACAGAACCATCGCCTCTCTGTGGTCCCGGTCATCATGACCGGGACCCGGACCGGCGCAACGGATCTTCAACGCCACATTCGCTGGCATTAACAATAACATGATATTCATCACGGAGTGACTA

>abgABT-ogt

TTTTAGGGAGCAAGTAAGTCTAAGCAAACCTTAACAGCAGAGAATTCCGATATTAGATGTAAATATATGTCTATCTATTTGAAAACCCTTAAGTTGTTAAGGGTAACTTTACATAAAAGTGTGAACAAGCTGGCACAAATTGTTTAATGTT

>yeaVW

TTTTTATTCGTTCTTTGCAGTAAATAACCTGCGTCATTTCACCTTTTATTGTTTCCGTTTCGTGTTTTATGGCTTTCCGTATTCTTAATTGTTTAATTTATGTAACATGCAAATTTTGTTACGCGTACGTTAGGTTCCGCCGTACAGGTAT

yneJ_b1526_at:

>anmK

CAGCCGGAGAGAAGGACCGGCAAAATGATAATTAACAGTTTTTTCATAGTCATATCCCGAAGACTTTCCTGGTCTGGAGGGCAATACGCCCTCCCTAACGTTCCAAGTGTAACGGCAGACGCGGTAAGAAAAATTCAGTTAACTCTGATAT

>pscKG

CAGAAATGCTATTGCCAGTAACACCACCATCCCCACAACACTTCTCATTATATCCATAATGATTTTCCCTTCATGCCGGTAAACCCGGCGTCAGCGCCAGGTTTTGGTATGCTTGATGAGTACGGGCGACGGCTTTCTGCCCGTCAGAAAT

>ydcO

GCTCGCAACATTTTAGCGGCGGGGCACGCCGTTCTTGCCTGTGGAGAGATGGTGCAGTCAGGCCGCCCGTTGAAGCAGGAACCCACCGAAATGATTCAGGCGACAGCCTGAACGTAGCAGGGATCCACGTCCTTCAGGGCGTGGAGGATGT

>yeeJ

CCATTTTCCCCTCGATTATAAAACTTGAGTTATTCAGTAGTCTCCCCTCTTGCAACTCACACCCAAAACTGCCTAACGAAAAGTTATTAATTTTCAATCATATTGCTATCAGTATTTACATTTTTTCGCTGTGCTAGAAAGGGCGCATTTA

>uidABC

TACGCTTATCAGGCCTACCTTTTCCTGCAACACTTTGAATTTATGAGTTTTTGTAGGCTGGATAAGGCGTTTACGCCGCATCCGGCATAAAAAACGCGCACTTTGTCAACAATCTGAAACGCCGGAGATTTTTCTCTCCGGCGTTATTTTT

>sad

ATAATCCGCGACTTTACCCTGACCAATGAGCGGCCGCACTTGCCGCAAGATGTTTTCTAAAATTGCATTATCCATGGCGACTGCCACTTTCTACTCCTGGACCGCAGGTCTGAAAAGACCTGCGAGTATATCAGAGCTGAATATGTCGCGT

>yeaL

GCTGTTTTCACGTTATTGTTAGCGTGCACAAATGGCAGTTTGATGACAGTTCGCCTGATTTTTTATGCAAAAAACGTAAAGATTTTCTACTTCCTTCCTGCAGCAAGCGTAAAGTAAGCAGGCTTATTATTTTTTGGCAAGGAAACCACGA

>ddpXABCDF

CCTTTGACGGCATAATCCAGAAAGCAAAAGCAGGATGCCCGGTCTCTCAGGTACTGAAAGCGGAAATTACGCTGGATTACCAGTTGAAATCGTAAAGCATTGCCGGATGACGCGTCAGGCGCGTGAATGCCTGATGCGTTGTTAGCATCTC

>ydjXYZ-ynjABCD

GCTCAAACTGGGTCAGGAGAATTAACCTTGAGAAAAATCAACAAACTGTCAGTAATGATTTGTTGCCTGCCGTCCTTTGTTATACCGTCTCTGCGTTTTTAGTTGTCTGACCACTTCTCTATTATCAAGTTTGATATAGGAAACTCCACGA

>ompG

TGCTGGTGCGTTAAATGATTATCATGCAGGAGAAAATATCACTATTCATTTTGATATGACGAAATGTCATTTCTTTGATGCAGAAACGGAAATAGCAATTCGCTAAATACAGGGGGAAGGCATTCCCCCAGGATAATACAAGGAACAATAA

>abgABT-ogt

TTTTAGGGAGCAAGTAAGTCTAAGCAAACCTTAACAGCAGAGAATTCCGATATTAGATGTAAATATATGTCTATCTATTTGAAAACCCTTAAGTTGTTAAGGGTAACTTTACATAAAAGTGTGAACAAGCTGGCACAAATTGTTTAATGTT

>feaB

AAGTCTTGTCAGGCATAGAGACATAAGCGGTTATTGTCACGATTTGCGGAGCTTGTCACAGCTGACAAAGCGAATGTCACAGCGAAAAAAGTGACTTTTCTTGTCGCTGCGTACACTGAAATCACACTGGGTAAATAATAAGGAAAAGTGA

>rnt-lhr

ATTGAGTTAATCGAAGAGAAAGACGCCGGTCGCGGTCTGGGCAACTAATCTCCTGCCGGGCGTGAACTCATCGCGCCCGCATCTTTACTGCATCGACAAGTAATATTTGTCATAATGCGCGCTGCAATTTATCCGTATTAAGAGAATCAGA

>ssuEADCB

TACGATGCCAAACGTCAGGAGAAAATGCGCGCGGCGCTGGAACAGTTGAAAGGGCTGGAAAATCTCTCTGGCGATCTGTACGAGAAGATAACTAAAGCACTGGCTTGATAAATAACCGAATGGCGGCAATAGCGCCGCCATTCGGGGAATT

>ydcSTUV

TTACAGACGGCATAATGCGCGGTAGCTCACAACCTGAATAAATTTTCTCAGGGGCGAAGGTGTGCCTGCAAGCCGCCGTCTATGGTTAAACAAGGAGATATTTTTACGGCACGGCGGCTGAACAATTAATTACGACAGGAGTAAGACCTTA

>yeaVW

TTTTTATTCGTTCTTTGCAGTAAATAACCTGCGTCATTTCACCTTTTATTGTTTCCGTTTCGTGTTTTATGGCTTTCCGTATTCTTAATTGTTTAATTTATGTAACATGCAAATTTTGTTACGCGTACGTTAGGTTCCGCCGTACAGGTAT

>rutABCDEFG

GCGCGATGCTTATCAGGCCTACCAGAAGATTGCAATATATTGAATTTGCACTGTTTTGTAGGCCGGATAAGGCGTTTACGCCGCATCCGGCATGAACAATGCGTACGTTGTCAACAATCTGCACCGCCGGTAACCCCGGCGGTTTTCTGTT

>ydfV

TTCATTGTCAGCGATATACTCGAGCATGAGACGAAGCGCTTCAGAAGGAGTTACACCCATTTTTTCAAGCGCGGCGTAAGAACGCGCTTTAAGTTCATCGTCAATACGCAGGTTAATGCTACCCATGTCTTACACCTCTTGTAATTACAAA

>lomR_2-stfR-tfaR

TTTTCGGGTCTGACGGCGCTTAGTGCTGAATTCACTATCGGCGAAGGTGAGTTGATGGCTCATGATGTCCCTCTGGGATGCGCTCCGGATGAATATGATGATCTCATATCAGGAACTTGTTCGCACCTTCCCAAGGGGAAAACGCACGACG

>pspABCDE

TGTTAGTGTAATTCGCTAACTCATCCTGGCATGTTGCTGTTGATTCTTCAATCAGATCTTTATAAATCAAAAAGATAAAAAATTGGCACGCAAATTGTATTAACAGTTCAGCAGGACAATCCTGAACGCAGAAATCAAGAGGACAACATTA

>recET-lar-ydaCQ-intR

GAAGTCAGCAATGGCTTCGCCCACGTTACGACGCAGACGTTTTTGTAATTTGTTCAGGTTGTATTGTTCTTTCTTTGTAATTTGTTGATTTTCTTGCATTATTTCAGTTCTCTGGTACTAAATGGGGCAAATTGGGGGCAAACTTTGCAAC

>dosCP

TTTTGATCAACATTTGTGCAGCGTAGTGCAGTTTTGGTGCAAGAGGGGAAGTTAAGGAAGGAATCTCCCGGAATCGTAGCTGAAATCACAGTATTTAAGTGACAGTGTCACGTTAAATGAAAACCCGCGAGTGCGGGCGAGAGGAATTTGT

>flxA

GAACGTTTACCCCTGTCAACAAGCTTTACTTTCTGAGGCGCGCCAGCCCGCGAGGAAAACAATCTGAACATCAAACAATTAATGACACAAGAAATACGATTAAAGATTTTTTTGTGCATGCCGATAGTGCTTTTTTAAAAGGAGAAATCTA

>C0343-dbpA

TTCGTACGACGCCGACTTTGATGAGTCGGCTTTTTTTTGCCTGTTATTTATCAGCGTCTACCCTTTAAGAGTCCACCCAATGACCAGAGGGAAATATGACGACACTTATTTATTTGCAAATTCCTGTCCCTGAACCGATTCCTGGCGATCC

>mliC

GGATCGGCGGGAAGATCGCGGCGGCGTAACCCGCCTTTGGTGTATTCACGGCGCAGATGCGCGATTTGCTGCAATTCGTCGTTATCAGACATGGTTTTCTTTACGGATTGTCAGTGGGTGACGCTATTGTGCGCCGCCCCTGGAAAAATCT

>yebQ

AAAAATCAGAACTGTTTTTTATTATAATTTCGCACCAGGGTGGTCGCAATCCATCTTTTGCCGGTTAGTTACAATTCTGCGACATCCACCGTGAATATCAGTGCTAGAATCATACCCCTGTTGATTATTCACCAAAGATATAAAATTCCTA

>narZYWV

AACAACACCGGCAGAATTGCCGCGAAACGGTTGTGAGGTAAAAGCATCGACGTGGTACACCTGCGGTTTCATTAACGTTCTCCTGTGACTGGAGAACTATCATAGCCTGCAAGTGGCCGGAGAGCGAAGGGCTATCCGGCCAGGGTGAAAT

>narGHJI

AAGGGGTATCTTAGGAATTTACTTTATTTTTCATCCCCATCACTCTTGATCGTTATCAATTCCCACGCTGTTTCAGAGCGTTACCTTGCCCTTAAACATTAGCAATGTCGATTTATCAGAGGGCCGACAGGCTCCCACAGGAGAAAACCGA

>ydaN

ACACGGTTTGTATACGGAAAAGCATTTTGCTTTTTGTATTCAATTTAGACAGAATTTTATTAATCATTTCAGGGTAATGGGGTGATGAGATGTTGCGTAACAGGGCCAGAAGGCTAGACTACAAAATAATGCGTTGATGATGGAGGCACTG

>nudG

CAAGAACCCGCACGCACTGATGCAACAACGGTTTAATCCGGGGATGAAGATGGCGGTCTAGCACAGGCACTCCTTAAATATAAAGCCTTTCTGATTGAGCAACAGTGCGGATATTATGGCATTTTTCGCTTATCTGCCCGTGTGTAATTTA

>ydcW

CTCAATCAGCTTGGGCGACCGCGTGATGTACCGGTAACTAACGTGGTGGCACTGCTGGTTATGTTGGTAACAACCTTGCCGATCCTGGGGGCCTGGTGGCTAACCCGCGAAGGCGACAATGGTCAATAACCACTGATACAGGAATATGCTA

>ycgR

GGACTTTCTCGTCAGATCGGAAAAAGGCGATCAGCAAAATCAACGATTTGGATGCTGACGAGTTCCTCGAACACGTAGCGCGAAATCACCCTGCGCCGCAGGCTCCGCGCTATATCTACAAACTTGAGCAGGCACTGGACGCGATGTAAAT

>torYZ

TGCCGGATGCGGCGTGAACGCCTTATCCGGCCTACAAAACCTTGCTAATTCAATATATTGCAGGGACTATGTAGGCCTGATAAGCATAGCGCATCAGGCAGCTTTACGTTTGCATAACCTCAGCGCCCGTTTCCGGGCGCTATTCACGTCT

>ydiQRST-fadK

TTGTTTTAGCGATCGCATTTTTTTTGCAAGATTGTTGGCAAGGAAAACAGCTTGCTCCGTCGAAAACCCCGCACCGCTATCGCACACTATTTTCAGGCCATTTTTACCTTCCATCGGAGATGGTTCCGTATGCGACTCACAGGAGAAATCA

>ydaSTUVW-rzpR

TCGGCCTAAGTCTTTACCACTAAGCATTGCTTAATATTCTCCTATGCGCATTACATTAGGCAATCCCTACCCTTACTGCATTAGGCACAGCCTATTGACAATTGCGTTAGGCGTCGCCTAATATTTCTGTGTGTTTTTGGAGTTCATTCGA

>ydaL

ATTAATATCGTGAATGAATAATCATGCATAAGTATTTTGCTTAAAATATCGGCAATATTTGGAACTTATTACTGGAAATTTGGGTAATACGTTGTTGGACCGACCCGGTCTGGTTATCATATCGCGCTCTTAATTGCGGGAGGATGTAACA

>lsrACDBFG-tam

GCATGAACAAACGCAACCGTGAAAATCAAAATAGCATAAATTGTGATCTATTCGTCGGAAATATGTGCAATGTCCACCTAAGGTTATGAACAAATTAAAAGCAGAAATACATTTGTTCAAAACTCACCTGCAAAACTGAACGGGGGAAATA

>ydiM

CAATCGGTAGCTGCTGAATTATATGCACACGATCTGGAGCGGCTTTGTTAATTTTTCCACAGAAAGGAATTGTCGTTGTTACAACAATAATGAACGGATGCTGACACAACATCGCTTCACTTTTTAAAGCACCTTTGCTAAGTAGAACCTA

>nhoA

TTTTTTAAGCGTAGTCCGTAACGCAATAAGTAACGAAATTAACGGGATTGGCGATTTGCGAACGTGATGCATGTCCGCGATCGCACAAAATAGCCGGTGCGGCGTCTATTCCAGGTTATAAGTTGAGAAAACCACTAAGGGAAACGCCTGA

>flgAMN

TGGCGTTTACTGCGTTTAATGGCGGTCGTGCTGGCGGGGATTGCCGCGTACTTCGCTGCACTGGCGGTACTGGGCTTCAAAGTTAAAGAATTTGCCCGCCGGACGGTGTAACAATGCATTCCGGCCTGCAGTGCAGGCCGGAGATAATCTT

>trpLEDCBA

TATATTACTGTTGGGCGGAAAAATGACGTAAGTTGACGTTCGACCGGGGTAAGCGAAACGGTAAAAAGATAAATATTAAATGAATTTAGGATTTTTCCGGCTTCATTAAAGAAAGTTAAAATGCCGCCAGCGGAACTGGCGGCTGTGGGAT

>flgBCDEFGHIJ

GCCCCCAGCCATTTCTACAACGTGAATTGTACCTGTCCGCAATGACCATCAACGGCATAAATAGCGACCCATTTTGCGTTTATTCCGCCGATAACGCGCGCGTAAAGGCATTTAAGCTGATGGCAGAATTTTGATACCTGCGGAGGAGATA

>mdtQ

CAGGTAATACCAATATCAAACCCCTGCTGCGCCAGTAATAACGCGCACTCTTTGCCGATCCCCGAATCGGAGGCGGTAATAATCGCAACCTGTGCCATCGAGTTCTCCACTTAACGCTGAATAAACGTTAAGTATAGAAGGCGCATATCAT

>ydaM

TGATGATGTTCAGGCATATTGCACCGCGCTACCGCATCATGGCGGCAGCGGGGCGTGTTACGTCGCACTACGTAAAACGGCGCAGGCGAAGCAAGAAAACTGGGAGCGCCACGCTAAGCGCAGTCGTTGATCTCGAGACGCATCCGCGGCT

>abgR

TGTGACCCGGTTCACGTAGCGATAGTTTTTACTTATCACTAACTGATTTTTCACAGTTTTAACCGTTCATAAATTACCCTGACACAATCATCTGCATTAAAGTAGATGCCAGTTTCTTTGGTCTGATAAATAACGGTTATCGGTGGCGTCA

>clcB

GCAATAATAGGTTACAGTGTCACGTTTTTTTATCTCTTAAAGCACGCACTGCTTTTGCGGCTGGCCTCTTTTGCCGCAAAATAGTCGCCCGTGTTTCATTGCCCATTTCTGCTCATGCATCATCTACACATCTATCCGGATCTGCGCACTA

>astCADBE

GAGAGAATATTACAACACGATGATTTTGCAGAGATTATGAAGAACTATACCGGATGACTGGTGATAAATAAAGCAAATAACCAGGATTAATCTGTATTAATTTATAAGAAAGCAACTTAATACCCGCAGAATGATTTCTGCGGGTAAGTAT

>fdnGHI

GCGTTTTTCTACCGCTATTGAGGTAGGTCAATTTGCGAAGGCGGATTATTTTGTGGCAAACAGATGTTCTTTTTGATTTCGCGCAAAAAGATTCAGAATTTTACTGTTAGTTTCCTCGCGCAGTAATACCCCTGAAAAAAGAGGAAAGCAA

>maoC

CTGCACCTGCCAGGTTGTTTGGCAGGTGTGCCAGCTTTTCATACAGTGGATGCCCTGAAAATAGATGTACACATCATGCATAATGTGACAACGTCACAAAACTTAGTGAAATAAAAGGGCAACTATTCGCCGTTGCCCTTCATTCACCGAT

>ynjH

CGGGGCGGATTATCCATCTTCATGCCTGGCACGTACCCGACTTCCACGGGACGTTACAGGCACATGAACATCAGGCGCTGGTCTGGTGCTCACCTGAAGAGGCGCTGCAATATCCGCTGGCCCCTGCTGACATTCCATTATTAGAGGCGTT

>ybiYW

CTCATAGGTGTGCTCCTGGCTCGAAAATGAAACCGTAACAGTGTAATAACAATGTGACGCAGAGCACAAATTATATTTCGAATGAAAGTAAGGATGAAATTGATGATGTGAATGATTTAGCCCGGCGACGACGCCGCCGGGCCGAGGAGAT

>ynbABCD

AACTATCTGATTAATTGGGGATAATCATTCCTGACAGTGAGTCCCCAATACCTTGATATATTCTGAATTTTTAATGAAACGGCGTGTTGCGATATCTCCGTCAGGGGAATTGATGCACCATAGCGCAAACCGAATTATCAAGGATTGATAA

>ves

TGACTCAATATCATATTATTCAAAACTGGCTGTGGCTGGGGGCGGTTAATTCGCTGGAAGAAGCGACAACGTTAATTCGGACACCCGCCGGGTTTGATCACGACGGTTATAAAATTCTTTGTAAGCCGCTGCTTTCCGGTAACTATGAAAT

>tynA

GGTGGGATGAAGCAGTCAGGAACGGGCCGTGATTTTGGCCCCGACTGGCTGGACGGTTGGTGTGAAACTAAGTCGGTGTGTGTACGGTATTAATCTGGTTCGCTCATAAGTAAAAAACGGCACCTGGTGCCGTTTTTTTGTCTGAAACAAT

>fryBC-ypdFE-fryA

GGCGAGGATGATTTTGCAGAAATATCATCTGTCAATTCATGAAGTGGCACAGCGTTGCGGTTTTCCGGATAGCGACTATTTTTGTCGCGTTTTCCGGCGTCAGTTTGGTCTGACGCCGGGAGAGTACAGCGCCCGTTTTCAGGGCTAACGT

>alkA

CTGGATTTGCTGGTGAATAGTAGTTATTTCGCGATTAACTGAGGCGTGCTTCCCCATCGCCTGATGCGACGCTAACGCGTCTTATCATGCCTACAAATCGCTCATTCCCCAGGCCGGATAAGGCGCTCGCACCGCATCCGGCGACCAACGT

>yeaN

GGCTTCATGGTGTCGGTCGGGTTCATAGCCATTGAGATTCAACCTGTGCATCATTTTGTCCGAACTTAGCGATAATTTGTCATTTTAGCTTGATTCAACATAACAATAAAAACGGTAAGGTACAGCCTCGTTTGTAACAATGAGAAGCATA

>ydhL

CAGCTGGCGGGCGGACATATTCCAGTCCATCAATCGCCAGTAGCCCATCACAAAACGGGAAAACTCCGGGCCTTGCGGCGCAATAGTAATACGCTGAACCATAATCGCTTCCTCTTATCAGATATGAGAGGAGTATACGCAAGATTAGGTT

>flhBAE

TTTTTATCCAAGCCCTTTGACAAGAGGATAATTCACATCTTTTTGGCATGTTTTGTTGCAAGCTATTCCTGATAAATAATTGCAACAAGACATCGAGCCTTTTTCACTGAGTTATTAAACATACTCGCGAGCGCGTAATTTTTTTGTCCTT

A.3 Manual of conditional distribution inferring functions

function: **construct_edges(net,regnet)**

Description: Construct a net based on the gene pairs.

Arguments:

net, graph object from “graph” package.

regnet, gene pairs(from transcription factor to target gene).

function: **modules(pairs,type=”single”)**

Description: Transform the gene pairs into module structure.

Arguments:

pairs, gene pairs.

type, if this value is “single”, each module will include just one target gene, otherwize, each module will include all the target genes with same regulators.

function: **mix_nor_distri_fit(regnet,exprs,prior=FALSE)**

Description: Calculate the parameters for each module.

Arguments:

regnet, a module structure from function “module“.

exprs, the expression profiles of genes.

prior, prior knoweledge of the parameters(please refer to the Mclust package for the detail).

function: **mod_con_dis_infer(mod,dat)**

Description: Infer the expression profiles based on known genes' expression in each module.

Arguments:

mod, the parameterized module.

dat, the expression profiles of genes in the module.

function: **pre_specify_gene(regnetp,exprs,spe.genes)** Description: Given specific gene, inferring the expression profiles of genes regulated by this gene.

Arguments:

regnetp, parameterized modules list.

exprs, the expression profiles of genes.

exprs.est, the specific genes.

An example of procedure:

```
require(graph)

require(mclust)

require(MASS)

exprs<-read.csv(file=”expression_profiles”)

genepairs<-read.csv(file=”regulation_relations”)

colnames(genepairs)<-c(”TFs”,”Targets”)

reg.net<-modules(genepairs) # construt the modules based on regulation relations

reg.net.paras<-mix_nor_distri_fit(reg.net,exprs.train) # train each module

pre.results<-mod_con_dis_infer(mod,dat) # infer target genes' expression profiles based on the regulator in each module
```

Appendix B

Appendix

B.1 The PR-curves (Precision-Recall) under different M_p index by selecting different score values

□ End of chapter.

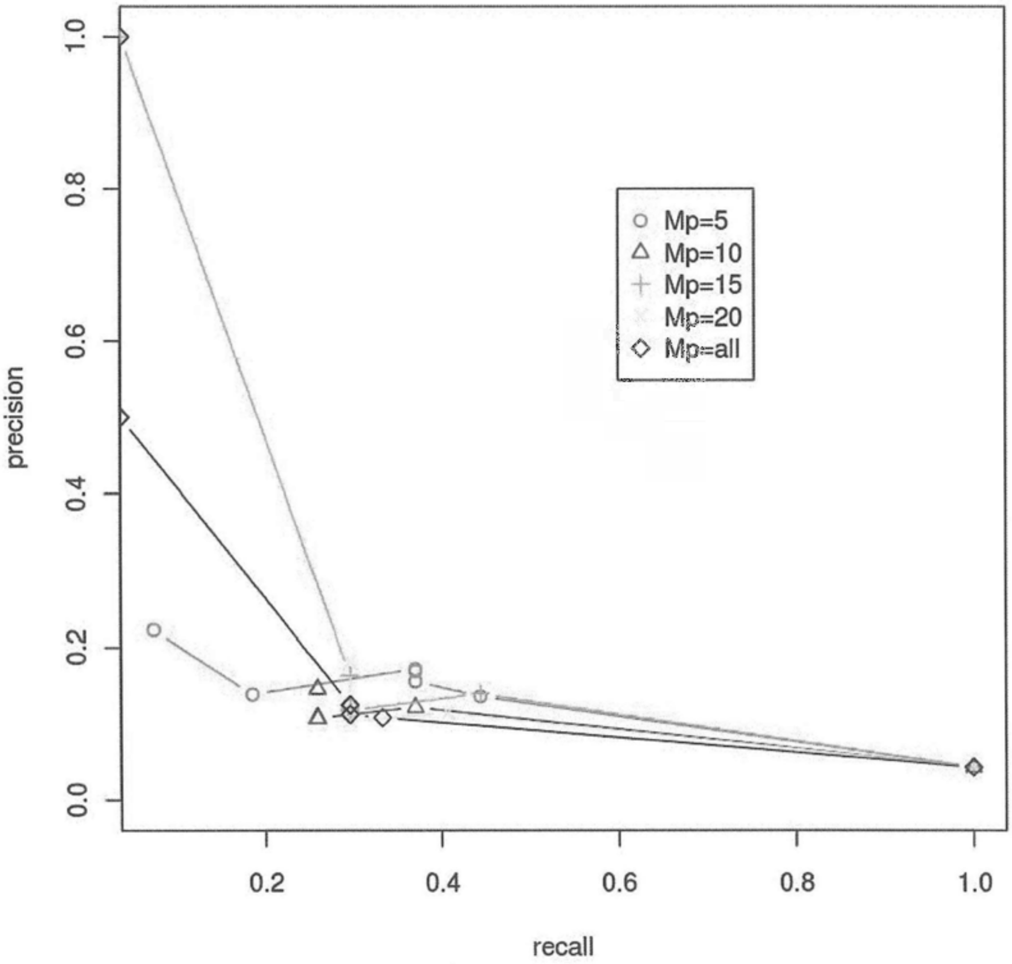


Figure B.1: **The PR-curve of the prediction result based on partial correlation.**

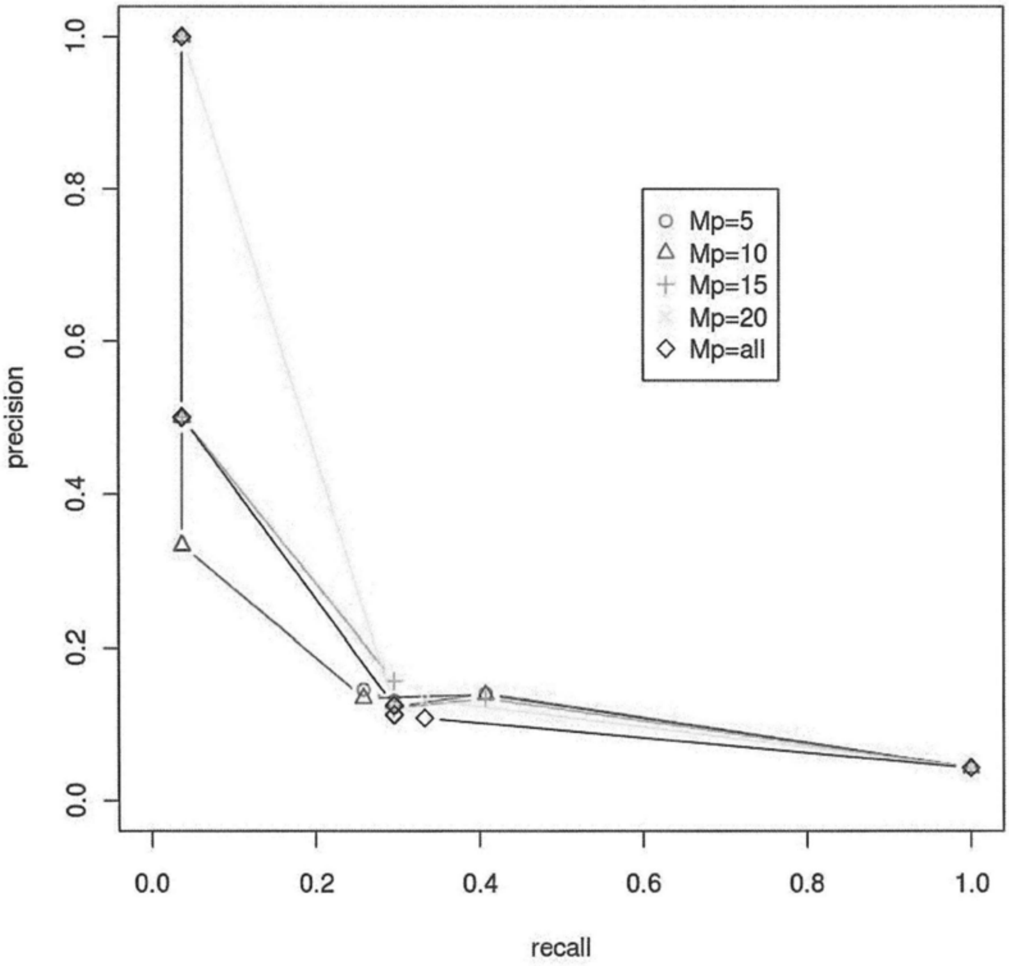


Figure B.2: **The PR-curve of the prediction result based on mutual information.**

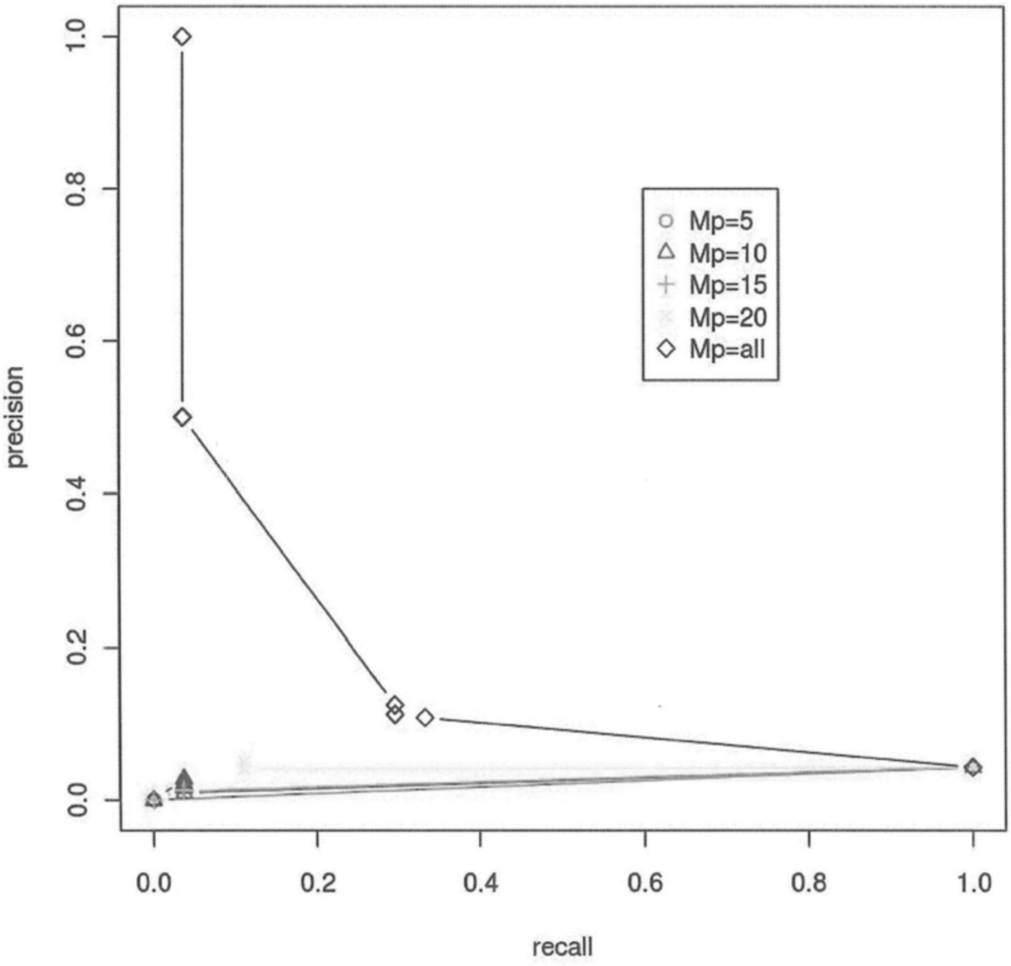


Figure B.3: **The PR-curve of the prediction result based on pearson correlation.**

Bibliography

[1] K. Nakashima, Y. Ito, and K. Yamaguchi-Shinozaki, 2009, "Transcriptional regulatory networks in response to abiotic stresses in arabidopsis and grasses," *Plant Physiology*, vol. 149, no. 1, pp. 88 –95.

[2] T. I. Lee, N. J. Rinaldi, F. Robert, D. T. Odom, Z. Bar-Joseph, G. K. Gerber, N. M. Hannett, C. T. Harbison, C. M. Thompson, I. Simon, J. Zeitlinger, E. G. Jennings, H. L. Murray, D. B. Gordon, B. Ren, J. J. Wyrick, J. Tagne, T. L. Volkert, E. Fraenkel, D. K. Gifford, and R. A. Young, Oct. 2002, "Transcriptional regulatory networks in saccharomyces cerevisiae," *Science (New York, N.Y.)*, vol. 298, no. 5594, pp. 799–804.

[3] E. Davidson and M. Levin, 2005, "Gene regulatory networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 14, p. 4935.

[4] M. A. Kohanski, D. J. Dwyer, J. Wierzbowski, G. Cottarel, and J. J. Collins, Nov. 2008, "Mistranslation of membrane proteins and Two-Component system activation trigger Antibiotic-Mediated cell death," *Cell*, vol. 135, no. 4, pp. 679–690.

[5] H. Yoon, J. E. McDermott, S. Porwollik, M. McClelland, and F. Heffron, 2009, "Coordinated regulation of virulence during systemic infection of salmonella enterica serovar typhimurium," *PLoS Pathog*, vol. 5, no. 2, p. e1000306.

[6] R. Bonneau, M. T. Facciotti, D. J. Reiss, A. K. Schmid, M. Pan, A. Kaur, V. Thorsson, P. Shannon, M. H. Johnson, J. C. Bare, W. Longabaugh, M. Vuthoori, K. Whitehead, A. Madar, L. Suzuki, T. Mori, D. Chang, J. DiRuggiero, C. H. Johnson, L. Hood, and N. S. Baliga, Dec. 2007, "A predictive model for transcriptional control of physiology in a free living cell," *Cell*, vol. 131, no. 7, pp. 1354–1365.

[7] R. De Smet and K. Marchal, 2010, "Advantages and limitations of current network inference methods," *Nat Rev Micro*, vol. 8, no. 10, pp. 717–729.

[8] H. D. VanGuilder, K. E. Vrana, and W. M. Freeman, Apr. 2008, "Twenty-five years of quantitative PCR for gene expression analysis," *BioTechniques*, vol. 44, no. 5, pp. 619–626.

[9] M. K. Udvardi, T. Czechowski, and W. Scheible, 2008, "Eleven golden rules of quantitative RT-PCR," *The Plant Cell Online*, vol. 20, no. 7, pp. 1736 –1737.

[10] E. Spackman and D. L. Suarez, "Type a influenza virus detection and quantitation by Real-Time RT-PCR," in *Avian Influenza Virus*, E. Spackman, Ed. Totowa, NJ: Humana Press, 2008, vol. 436, pp. 19–26.

[11] J. Gertsch, M. Güttinger, O. Sticher, and J. Heilmann, Aug. 2002, "Relative quantification of mRNA levels in jurkat t cells with RT-real time-PCR (RT-rt-PCR): new possibilities for the screening of anti-inflammatory and cytotoxic compounds," *Pharmaceutical Research*, vol. 19, no. 8, pp. 1236–1243.

[12] M. D. Adams, J. M. Kelley, J. D. Gocayne, M. Dubnick, M. H. Polymeropoulos, H. Xiao, C. R. Merril, A. Wu, B. Olde, and R. F. Moreno, Jun. 1991, "Complementary DNA sequencing: expressed sequence tags and human genome project," *Science (New York, N.Y.)*, vol. 252, no. 5013, pp. 1651–1656.

[13] L. H. Augenlicht and D. Kobrin, 1982, "Cloning and screening of sequences expressed in a mouse colon tumor," *Cancer Research*, vol. 42, no. 3, pp. 1088 –1093.

[14] D. A. Kulesh, D. R. Clive, D. S. Zarlenga, and J. J. Greene, 1987, "Identification of interferon-modulated proliferation-related cDNA sequences," *Proceedings of the National Academy of Sciences*, vol. 84, no. 23, pp. 8453 –8457.

[15] M. Schena, D. Shalon, R. W. Davis, and P. O. Brown, Oct. 1995, "Quantitative monitoring of gene expression patterns with a complementary DNA microarray," *Science (New York, N.Y.)*, vol. 270, no. 5235, pp. 467–470.

[16] D. A. Lashkari, J. L. DeRisi, J. H. McCusker, A. F. Namath, C. Gentile, S. Y. Hwang, P. O. Brown, and R. W. Davis, Nov. 1997, "Yeast microarrays for genome wide parallel genetic and gene expression analysis," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 24, pp. 13 057–13 062.

[17] C. Lausted, T. Dahl, C. Warren, K. King, K. Smith, M. Johnson, R. Saleem, J. Aitchison, L. Hood, and S. R. Lasky, 2004, "POSaM: a fast, flexible, open-source, inkjet oligonucleotide synthesizer and microarrayer," *Genome Biology*, vol. 5, no. 8, p. R58.

[18] T. Bammler, R. P. Beyer, S. Bhattacharya, G. A. Boorman, A. Boyles, B. U. Bradford, R. E. Bumgarner, P. R. Bushel, K. Chaturvedi, D. Choi, M. L. Cunningham, S. Deng, H. K. Dressman, R. D. Fannin, F. M. Farin, J. H. Freedman, R. C. Fry, A. Harper, M. C. Humble, P. Hurban, T. J. Kavanagh, W. K. Kaufmann, K. F. Kerr, L. Jing, J. A. Lapidus, M. R. Lasarev, J. Li, Y. Li, E. K. Lobenhofer, X. Lu, R. L. Malek, S. Milton, S. R. Nagalla, J. P. O'malley, V. S. Palmer, P. Pattee, R. S. Paules, C. M. Perou, K. Phillips, L. Qin, Y. Qiu, S. D. Quigley, M. Rodland, I. Rusyn, L. D. Samson, D. A. Schwartz, Y. Shi, J. Shin, S. O. Sieber, S. Slifer, M. C. Speer, P. S. Spencer, D. I. Sproles, J. A. Swenberg, W. A. Suk, R. C. Sullivan, R. Tian, R. W. Tennant, S. A. Todd, C. J. Tucker, B. Van Houten, B. K. Weis, S. Xuan, and H. Zarbl, May 2005, "Standardizing global gene

expression analysis between laboratories and across platforms," *Nature Methods*, vol. 2, no. 5, pp. 351–356.

[19] T. Tang, N. François, A. Glatigny, N. Agier, M. Mucchielli, L. Aggerbeck, and H. Delacroix, Oct. 2007, "Expression ratio evaluation in two-colour microarray experiments is significantly improved by correcting image misalignment," *Bioinformatics (Oxford, England)*, vol. 23, no. 20, pp. 2686–2691.

[20] O. Aparicio, J. V. Geisberg, and K. Struhl, Sep. 2004, "Chromatin immunoprecipitation for determining the association of proteins with specific genomic sequences in vivo," *Current Protocols in Cell Biology / Editorial Board, Juan S. Bonifacino ... [et Al*, vol. Chapter 17, p. Unit 17.7.

[21] M. J. Buck and J. D. Lieb, Mar. 2004, "ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments," *Genomics*, vol. 83, no. 3, pp. 349–360.

[22] T. E. Royce, J. S. Rozowsky, P. Bertone, M. Samanta, V. Stolc, S. Weissman, M. Snyder, and M. Gerstein, Aug. 2005, "Issues in the analysis of oligonucleotide tiling microarrays for transcript mapping," *Trends in genetics : TIG*, vol. 21, no. 8, pp. 466–475.

[23] D. S. Johnson, A. Mortazavi, R. M. Myers, and B. Wold, 2007, "Genome-Wide mapping of in vivo Protein-DNA interactions," *Science*, vol. 316, no. 5830, pp. 1497 –1502.

[24] R. Jothi, S. Cuddapah, A. Barski, K. Cui, and K. Zhao, 2008, "Genome-wide identification of in vivo protein–DNA binding sites from ChIP-Seq data," *Nucleic Acids Research*, vol. 36, no. 16, pp. 5221 –5231.

[25] B. E. Bernstein, M. Kamal, K. Lindblad-Toh, S. Bekiranov, D. K. Bailey, D. J. Huebert, S. McMahon, E. K. Karlsson, r. Kulbokas, Edward J, T. R. Gingeras, S. L. Schreiber, and E. S.

Lander, Jan. 2005, "Genomic maps and comparative analysis of histone modifications in human and mouse," *Cell*, vol. 120, no. 2, pp. 169–181.

[26] M. Bansal, V. Belcastro, A. Ambesi-Impiombato, and D. di Bernardo, 2007, "How to infer gene networks from expression profiles," *Mol Syst Biol*, vol. 3.

[27] S. A. Kauffman. Oxford University Press, USA, Jun. 1993, *The Origins of Order: Self-Organization and Selection in Evolution*.

[28] J. Vohradsky, Sep. 2001, "Neural model of the genetic network," *The Journal of Biological Chemistry*, vol. 276, no. 39, pp. 36 168–36 173.

[29] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, Aug. 2002, "Stochastic gene expression in a single cell," *Science (New York, N.Y.)*, vol. 297, no. 5584, pp. 1183–1186.

[30] W. J. Blake, M. KAErn, C. R. Cantor, and J. J. Collins, Apr. 2003, "Noise in eukaryotic gene expression," *Nature*, vol. 422, no. 6932, pp. 633–637.

[31] A. Arkin, J. Ross, and H. H. McAdams, Aug. 1998, "Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected escherichia coli cells," *Genetics*, vol. 149, no. 4, pp. 1633–1648.

[32] J. M. Raser and E. K. O'Shea, Sep. 2005, "Noise in gene expression: origins, consequences, and control," *Science (New York, N.Y.)*, vol. 309, no. 5743, pp. 2010–2013.

[33] M. B. Elowitz and S. Leibler, Jan. 2000, "A synthetic oscillatory network of transcriptional regulators," *Nature*, vol. 403, no. 6767, pp. 335–338.

[34] T. S. Gardner, C. R. Cantor, and J. J. Collins, Jan. 2000, "Construction of a genetic toggle switch in escherichia coli," *Nature*, vol. 403, no. 6767, pp. 339–342.

[35] D. T. Gillespie, 1976, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," *Journal of Computational Physics*, vol. 22, no. 4, pp. 403–434.

[36] M. R. Roussel and R. Zhu, Nov. 2006, "Validation of an algorithm for delay stochastic simulation of transcription and translation in prokaryotic gene expression," *Physical Biology*, vol. 3, no. 4, pp. 274–284.

[37] A. Ribeiro, R. Zhu, and S. A. Kauffman, Nov. 2006, "A general modeling strategy for gene regulatory networks with stochastic dynamics," *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology*, vol. 13, no. 9, pp. 1630–1639.

[38] "Wikipedia," http://en.wikipedia.org/wiki/Gene_regulatory_network. [Online]. Available: http://en.wikipedia.org/wiki/Gene_regulatory_network

[39] G. Karlebach and R. Shamir, 2008, "Modelling and analysis of gene regulatory networks," *Nat Rev Mol Cell Biol*, vol. 9, no. 10, pp. 770–780.

[40] M. B. Eisen, P. T. Spellman, P. O. Brown, and D. Botstein, Dec. 1998, "Cluster analysis and display of genome-wide expression patterns," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 25, pp. 14 863–14 868.

[41] A. Ben-Dor, R. Shamir, and Z. Yakhini, 1999, "Clustering gene expression patterns," *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology*, vol. 6, no. 3-4, pp. 281–297.

[42] U. Alon, N. Barkai, D. A. Notterman, K. Gish, S. Ybarra, D. Mack, and A. J. Levine, Jun. 1999, "Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 12, pp. 6745–6750.

[43] S. Tavazoie, J. D. Hughes, M. J. Campbell, R. J. Cho, and G. M. Church, Jul. 1999, "Systematic determination of genetic network architecture," *Nature Genetics*, vol. 22, no. 3, pp. 281–285.

[44] R. Shamir and R. Sharan, 2001, "Algorithmic approaches to clustering gene expression data," *CURRENT TOPICS IN COMPUTATIONAL BIOLOGY*, pp. 269—300.

[45] A. Brazma and J. Vilo, Aug. 2000, "Gene expression data analysis," *FEBS Letters*, vol. 480, no. 1, pp. 17–24.

[46] P. D'haeseleer, S. Liang, and R. Somogyi, Aug. 2000, "Genetic network inference: from co-expression clustering to reverse engineering," *Bioinformatics (Oxford, England)*, vol. 16, no. 8, pp. 707–726.

[47] J. H. Ward, 1963, "Hierarchical grouping to optimize an objective function," *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 236–244.

[48] A. Fernández and S. Gómez, Jun. 2008, "Solving Non-Uniqueness in agglomerative hierarchical clustering using multidendrograms," *Journal of Classification*, vol. 25, no. 1, pp. 43–65.

[49] J. B. MacQueen, "Some methods for classification and analysis of MultiVariate observations," in *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, L. M. L. Cam and J. Neyman, Eds., vol. 1. University of California Press, 1967, pp. 281–297.

[50] R. Nock and F. Nielsen, Aug. 2006, "On weighting clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 8, pp. 1223–1235.

[51] M. Ester, H.-p. Kriegel, J. S, and X. Xu, 1996, "A density-based algorithm for discovering clusters in large spatial databases with noise," pp. 226—231.

[52] M. Ankerst, M. M. Breunig, H.-p. Kriegel, and J. S, 1999, "OPTICS: ordering points to identify the clustering structure," pp. 49—60.

[53] H. Kriegel, P. Kröger, and A. Zimek, Mar. 2009, "Clustering high-dimensional data: A survey on subspace clustering, pattern-based clustering, and correlation clustering," *ACM Trans. Knowl. Discov. Data*, vol. 3, no. 1, p. 1:1–1:58.

[54] E. Segal, M. Shapira, A. Regev, D. Pe'er, D. Botstein, D. Koller, and N. Friedman, Jun. 2003, "Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data," *Nature Genetics*, vol. 34, no. 2, pp. 166–176.

[55] I. Priness, O. Maimon, and I. Ben-Gal, 2007, "Evaluation of gene-expression clustering via mutual information distance measure," *BMC Bioinformatics*, vol. 8, no. 1, p. 111.

[56] R. Gentleman, V. Carey, W. Huber, R. Irizarry, and S. Dudoit. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2005, *Bioinformatics and Computational Biology Solutions Using R and Bioconductor (Statistics for Biology and Health).*

[57] M. G. Kendall, 1938, "A new measure of rank correlation," *Biometrika*, vol. 30, no. 1/2, pp. 81–93.

[58] S. Kullback and R. A. Leibler, Mar. 1951, "On information and sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86.

[59] S. Ghosh, K. P. Burnham, N. F. Laubscher, G. E. Dallal, L. Wilkinson, D. F. Morrison, M. W. Loyer, B. Eisenberg, S. Kullback, I. T. Jolliffe, and J. S. Simonoff, 1987, "Letter to the editor: The Kullback–Leibler distance," *The American Statistician*, vol. 41, no. 4, pp. 338–341.

[60] C. E. Shannon, Jan. 2001, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 5, p. 3–55.

[61] A. A. Alizadeh, M. B. Eisen, R. E. Davis, C. Ma, I. S. Lossos, A. Rosenwald, J. C. Boldrick, H. Sabet, T. Tran, X. Yu, J. I. Powell, L. Yang, G. E. Marti, T. Moore, J. Hudson, J, L. Lu, D. B. Lewis, R. Tibshirani, G. Sherlock, W. C. Chan, T. C. Greiner, D. D. Weisenburger, J. O. Armitage, R. Warnke, R. Levy, W. Wilson, M. R. Grever, J. C. Byrd, D. Botstein, P. O. Brown,

and L. M. Staudt, Feb. 2000, "Distinct types of diffuse large b-cell lymphoma identified by gene expression profiling," *Nature*, vol. 403, no. 6769, pp. 503–511.

[62] A. Ben-Dor, L. Bruhn, N. Friedman, I. Nachman, M. Schummer, and Z. Yakhini, 2000, "Tissue classification with gene expression profiles," *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology*, vol. 7, no. 3-4, pp. 559–583.

[63] E. P. Xing and R. M. Karp, "CLIFF: clustering of high-dimensional microarray data via iterative feature filtering using normalized cuts." in *ISMB (Supplement of Bioinformatics)'01*, 2001, pp. 306–315.

[64] X. Chen, S. T. Cheung, S. So, S. T. Fan, C. Barry, J. Higgins, K. Lai, J. Ji, S. Dudoit, I. O. Ng, M. van de Rijn, D. Botstein, and P. O. Brown, Jun. 2002, "Gene expression patterns in human liver cancers," *Mol. Biol. Cell*, vol. 13, no. 6, pp. 1929–1939.

[65] J. H. M. van Delft, E. van Agen, S. G. J. van Breda, M. H. Herwijnen, Y. C. M. Staal, and J. C. S. Kleinjans, Jul. 2004, "Discrimination of genotoxic from non-genotoxic carcinogens by gene expression profiling," *Carcinogenesis*, vol. 25, no. 7, pp. 1265–1276.

[66] G. S. Michaels, D. B. Carr, M. Askenazi, S. Fuhrman, X. Wen, and R. Somogyi, 1998, "Cluster analysis and data visualization of large-scale gene expression data," *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, pp. 42–53.

[67] W. Geng, P. Cosman, J. Baek, C. C. Berry, and W. R. Schafer, Nov. 2003, "Quantitative classification and natural clustering of caenorhabditis elegans behavioral phenotypes," *Genetics*, vol. 165, no. 3, pp. 1117–1126.

[68] M. Reich, K. Ohm, M. Angelo, P. Tamayo, and J. P. Mesirov, Jul. 2004, "GeneCluster 2.0: an advanced toolset for bioarray analysis," *Bioinformatics*, vol. 20, no. 11, pp. 1797 –1798.

[69] H. Herzel and I. Große, Jul. 1995, "Measuring correlations in symbol sequences," *Physica A: Statistical and Theoretical Physics*, vol. 216, no. 4, pp. 518–542.

[70] J. Kurths, C. O. Daub, J. Weise, J. Selbig, and Steuer, 2002, "The mutual information: detecting and evaluating dependencies between variables." *Bioinformatics*, vol. 18 Suppl 2, no. 2, pp. S231–40.

[71] A. J. Butte and I. S. Kohane, 2000, "Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements," *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, pp. 418–429.

[72] R. Herwig, A. J. Poustka, C. Müller, C. Bull, H. Lehrach, and J. O'Brien, Nov. 1999, "Large-scale clustering of cDNA-fingerprinting data," *Genome Research*, vol. 9, no. 11, pp. 1093–1105.

[73] C. Daub, R. Steuer, J. Selbig, and S. Kloska, 2004, "Estimating mutual information using b-spline functions - an improved similarity measure for analysing gene expression data," *BMC Bioinformatics*, vol. 5, no. 1, p. 118.

[74] Y. Ko, C. Zhai, and S. Rodriguez-Zas, 2009, "Inference of gene pathways using mixture bayesian networks," *BMC Systems Biology*, vol. 3, no. 1, p. 54.

[75] Y. Ko, C. Zhai, and S. L. Rodriguez-Zas, 2010, "Discovery of gene network variability across samples representing multiple classes," *International Journal of Bioinformatics Research and Applications*, vol. 6, no. 4, pp. 402 – 417.

[76] D. Klein-Marcuschamer, C. N. S. Santos, H. Yu, and G. Stephanopoulos, May 2009, "Mutagenesis of the bacterial RNA polymerase alpha subunit for improvement of complex phenotypes," *Applied and Environmental Microbiology*, vol. 75, no. 9, pp. 2705–2711.

[77] B. P. Cormack, G. Bertram, M. Egerton, N. A. Gow, S. Falkow, and A. J. Brown, Feb. 1997, "Yeast-enhanced green fluorescent protein (yEGFP)a reporter of gene expression in candida albicans," *Microbiology (Reading, England)*, vol. 143 (Pt 2), pp. 303–311.

[78] E. Andrianantoandro, S. Basu, D. K. Karig, and R. Weiss, 2006, "Synthetic biology: new engineering rules for an emerging discipline," *Molecular Systems Biology*, vol. 2, p. 2006.0028.

[79] G. M. Church, 2005, "From systems biology to synthetic biology," *Mol Syst Biol*, vol. 1.

[80] J. Pleiss, Nov. 2006, "The promise of synthetic biology," *Applied Microbiology and Biotechnology*, vol. 73, no. 4, pp. 735–739.

[81] U. Alon, 2003, "Biological networks: The tinkerer as an engineer," *Science*, vol. 301, no. 5641, pp. 1866 –1867.

[82] H. Kobayashi, M. Kærn, M. Araki, K. Chung, T. S. Gardner, C. R. Cantor, and J. J. Collins, 2004, "Programmable cells: Interfacing natural and engineered gene networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 22, pp. 8414 –8419.

[83] Y. N. Kaznessis, 2007, "Models for synthetic biology," *BMC Systems Biology*, vol. 1, p. 47.

[84] J. Carrera, G. Rodrigo, and A. Jaramillo, Apr. 2009, "Model-based redesign of global transcription regulation," *Nucleic Acids Research*, vol. 37, no. 5, p. e38.

[85] J. J. Faith, M. E. Driscoll, V. A. Fusaro, E. J. Cosgrove, B. Hayete, F. S. Juhn, S. J. Schneider, and T. S. Gardner, Oct. 2007, "Many microbe microarrays database: uniformly normalized affymetrix compendia with structured experimental metadata," *Nucl. Acids Res.*, p. gkm815.

[86] R. A. Irizarry, B. Hobbs, F. Collin, Y. D. Beazer-Barclay, K. J. Antonellis, U. Scherf, and T. P. Speed, Apr. 2003, "Exploration, normalization, and summaries of high density oligonucleotide array probe level data," *Biostatistics (Oxford, England)*, vol. 4, no. 2, pp. 249–264.

[87] S. Gama-Castro, H. Salgado, M. Peralta-Gil, A. Santos-Zavaleta, L. Muñiz-Rascado, H. Solano-Lira, V. Jimenez-Jacinto, V. Weiss, J. S. García-Sotelo, A. López-Fuentes, L. Porrón-Sotelo, S. Alquicira-Hernández, A. Medina-Rivera, I. Martínez-Flores, K. Alquicira-Hernández,

R. Martínez-Adame, C. Bonavides-Martínez, J. Miranda-Ríos, A. M. Huerta, A. Mendoza-Vargas, L. Collado-Torres, B. Taboada, L. Vega-Alvarado, M. Olvera, L. Olvera, R. Grande, E. Morett, and J. Collado-Vides, 2010, "RegulonDB version 7.0: transcriptional regulation of escherichia coli k-12 integrated within genetic sensory response units (Gensor units)," *Nucleic Acids Research.*

[88] T. Van den Bulcke, K. Van Leemput, B. Naudts, P. van Remortel, H. Ma, A. Verschoren, B. De Moor, and K. Marchal, 2006, "SynTReN: a generator of synthetic gene expression data for design and analysis of structure learning algorithms," *BMC Bioinformatics*, vol. 7, no. 1, p. 43.

[89] R. D. C. Team. Vienna, Austria: R Foundation for Statistical Computing, 2009, *R: A language and environment for statistical computing.*

[90] C. Fraley and A. E. Raftery, 2000, "Model-based clustering, discriminant analysis, and density estimation," *JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION*, vol. 97, pp. 611–631.

[91] C. Fraley and A. Raftery, "MCLUST version 3 for r: Normal mixture modeling and Model-Based clustering," Tech. Rep., 2007.

[92] M. S. Cline, M. Smoot, E. Cerami, A. Kuchinsky, N. Landys, C. Workman, R. Christmas, I. Avila-Campilo, M. Creech, B. Gross, K. Hanspers, R. Isserlin, R. Kelley, S. Killcoyne, S. Lotia, S. Maere, J. Morris, K. Ono, V. Pavlovic, A. R. Pico, A. Vailaya, P. Wang, A. Adler, B. R. Conklin, L. Hood, M. Kuiper, C. Sander, I. Schmulevich, B. Schwikowski, G. J. Warner, T. Ideker, and G. D. Bader, 2007, "Integration of biological networks and gene expression data using cytoscape," *Nature Protocols*, vol. 2, no. 10, pp. 2366–2382.

[93] G. Schwarz, Mar. 1978, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464.

[94] A. P. Dempster, N. M. Laird, and D. B. Rubin, 1977, "Maximum likelihood from incomplete data via the EM algorithm," *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, vol. 39, no. 1, pp. 1—38.

[95] R. Ihaka and R. Gentleman, 1996, "R: A language for data analysis and graphics," *Journal of Computational and Graphical Statistics*, vol. 5, no. 3, pp. 314, 299.

[96] B. Ding, R. Gentleman, and V. Carey, 2011, "bioDist: different distance measures."

[97] P. Cohen, J. Cohen, S. G. West, and L. S. Aiken. Routledge Academic, Aug. 2002, *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, 3rd ed.

[98] M. Kanehisa, S. Goto, M. Furumichi, M. Tanabe, and M. Hirakawa, Jan. 2010, "KEGG for representation and analysis of molecular networks involving diseases and drugs," *Nucleic Acids Research*, vol. 38, no. Database issue, pp. D355–360.

[99] M. Kanehisa, S. Goto, M. Hattori, K. F. Aoki-Kinoshita, M. Itoh, S. Kawashima, T. Katayama, M. Araki, and M. Hirakawa, Jan. 2006, "From genomics to chemical genomics: new developments in KEGG," *Nucleic Acids Research*, vol. 34, no. Database issue, pp. D354–357.

[100] M. Kanehisa and S. Goto, Jan. 2000, "KEGG: kyoto encyclopedia of genes and genomes," *Nucleic Acids Research*, vol. 28, no. 1, pp. 27–30.

[101] A. J. Dombroski, B. D. Johnson, M. Lonetto, and C. A. Gross, Aug. 1996, "The sigma subunit of escherichia coli RNA polymerase senses promoter spacing," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 17, pp. 8858–8862.

[102] R. G. Martin, W. K. Gillette, S. Rhee, and J. L. Rosner, Nov. 1999, "Structural requirements for marbox function in transcriptional activation of mar/sox/rob regulon promoters in escherichia coli: sequence, orientation and spatial relationship to the core promoter," *Molecular Microbiology*, vol. 34, no. 3, pp. 431–441.

[103] D. F. Browning and S. J. W. Busby, 2004, "The regulation of bacterial transcription initiation," *Nat Rev Micro*, vol. 2, no. 1, pp. 57–65.

[104] L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, and R. Phillips, Apr. 2005, "Transcriptional regulation by the numbers: models," *Current Opinion in Genetics & Development*, vol. 15, no. 2, pp. 116–124.

[105] T. L. Bailey and C. Elkan, 1994, "Fitting a mixture model by expectation maximization to discover motifs in biopolymers," *Proceedings / ... International Conference on Intelligent Systems for Molecular Biology ; ISMB. International Conference on Intelligent Systems for Molecular Biology*, vol. 2, pp. 28–36.

[106] R. Münch, K. Hiller, H. Barg, D. Heldt, S. Linz, E. Wingender, and D. Jahn, Jan. 2003, "PRODORIC: prokaryotic database of gene regulation," *Nucleic Acids Research*, vol. 31, no. 1, pp. 266–269.

[107] R. d'Ari, Apr. 1985, "The SOS system," *Biochimie*, vol. 67, no. 3-4, pp. 343–347.

[108] A. R. Fernández De Henestrosa, T. Ogi, S. Aoyagi, D. Chafin, J. J. Hayes, H. Ohmori, and R. Woodgate, Mar. 2000, "Identification of additional genes belonging to the LexA regulon in escherichia coli," *Molecular Microbiology*, vol. 35, no. 6, pp. 1560–1572.

[109] R. Brent and M. Ptashne, Apr. 1980, "The lexA gene product represses its own promoter," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 77, no. 4, pp. 1932–1936.

[110] W. B. Muse and R. A. Bender, Mar. 1998, "The nac (nitrogen assimilation control) gene from escherichia coli," *Journal of Bacteriology*, vol. 180, no. 5, pp. 1166–1173.

[111] N. L. Craig and J. W. Roberts, 1980, "E. coli recA protein-directed cleavage of phage [lambda] repressor requires polynucleotide," *Nature*, vol. 283, no. 5742, pp. 26–30.

[112] T. Ogawa, H. Wabiko, T. Tsurimoto, T. Horii, H. Masukata, and H. Ogawa, 1979, "Characteristics of purified recA protein and the regulation of its synthesis in vivo," *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 43, pp. 909 –915.

[113] A. Matsushiro, K. Sato, H. Miyamoto, T. Yamamura, and T. Honda, Apr. 1999, "Induction of prophages of enterohemorrhagic escherichia coli O157:H7 with norfloxacin," *Journal of Bacteriology*, vol. 181, no. 7, pp. 2257–2260.

[114] Z. Ghahramani and M. J. Beal, 2000, "Variational inference for bayesian mixture of factor analysers," *In Advances in Neural Information Processing Systems 12*, vol. 12, pp. 449—455.

[115] D. Husmeier, Nov. 2000, "The bayesian evidence scheme for regularizing probability-density estimating neural networks," *Neural Computation*, vol. 12, no. 11, pp. 2685–2717.

[116] A. Penalver Benavent, F. Escolano Ruiz, and J. Saez, 2009, "Learning gaussian mixture models with Entropy-Based criteria," *Neural Networks, IEEE Transactions on*, vol. 20, no. 11, pp. 1756–1771.

[117] D. J. C. MacKay, "Introduction to monte carlo methods," in *Learning in Graphical Models*, ser. NATO Science Series, M. I. Jordan, Ed. Kluwer Academic Press, 1998, pp. 175–204.

[118] Y. X. Zhu, L. Y. Kang, W. Luo, C. H. Li, L. Yang, and Y. Yang, Jun. 1996, "Multiple transcription factors are required for activation of human interleukin 9 gene in t cells," *Journal of Biological Chemistry*, vol. 271, no. 26, pp. 15 815 –15 822.

[119] H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono, and M. Kanehisa, Jan. 1999, "KEGG: kyoto encyclopedia of genes and genomes," *Nucleic Acids Research*, vol. 27, no. 1, pp. 29–34.

[120] H. Muller, J. Chiou, and X. Leng, 2008, "Inferring gene expression dynamics via functional regression analysis," *BMC Bioinformatics*, vol. 9, no. 1, p. 60.

[121] C. Balsalobre, J. Johansson, and B. E. Uhlin, Aug. 2006, "Cyclic AMP-dependent osmoregulation of crp gene expression in escherichia coli," *Journal of Bacteriology*, vol. 188, no. 16, pp. 5935–5944.

[122] J. Johansson, C. Balsalobre, S. Y. Wang, J. Urbonaviciene, D. J. Jin, B. Sondén, and B. E. Uhlin, Aug. 2000, "Nucleoid proteins stimulate stringently controlled bacterial promoters: a link between the cAMP-CRP and the (p)ppGpp regulons in escherichia coli," *Cell*, vol. 102, no. 4, pp. 475–485.

[123] X. Mao, Y. Huo, M. Buck, A. Kolb, and Y. Wang, 2007, "Interplay between CRP-cAMP and PII-Ntr systems forms novel regulatory network between carbon metabolism and nitrogen assimilation in escherichia coli," *Nucleic Acids Research*, vol. 35, no. 5, pp. 1432–1440.

[124] Z. Zhang, G. Gosset, R. Barabote, C. S. Gonzalez, W. A. Cuevas, and J. S. Milton H, Feb. 2005, "Functional interactions between the carbon and iron utilization regulators, crp and fur, in escherichia coli," *Journal of Bacteriology*, vol. 187, no. 3, pp. 980–990.

[125] K. Nishino, Y. Senda, and A. Yamaguchi, Mar. 2008, "CRP regulator modulates multidrug resistance of escherichia coli by repressing the mdtEF multidrug efflux genes," *The Journal of Antibiotics*, vol. 61, no. 3, pp. 120–127.

[126] S. E. Finkel and R. C. Johnson, Nov. 1992, "The fis protein: it's not just for DNA inversion anymore," *Molecular Microbiology*, vol. 6, no. 22, pp. 3257–3265.

[127] A. Travers, R. Schneider, and G. Muskhelishvili, Feb. 2001, "DNA supercoiling and transcription in escherichia coli: The FIS connection," *Biochimie*, vol. 83, no. 2, pp. 213–217.

[128] M. Isalan, C. Lemerle, K. Michalodimitrakis, C. Horn, P. Beltrao, E. Raineri, M. Garriga-Canut, and L. Serrano, Apr. 2008, "Evolvability and hierarchy in rewired bacterial gene networks," *Nature*, vol. 452, no. 7189, pp. 840–845.

[129] A. P. Benavent, F. E. Ruiz, and J. M. S. Martinez, "EBEM: an entropy-based EM algorithm for gaussian mixture models," in *Pattern Recognition, International Conference on*, vol. 2. Los Alamitos, CA, USA: IEEE Computer Society, 2006, pp. 451–455.

[130] J. J. Faith, B. Hayete, J. T. Thaden, I. Mogno, J. Wierzbowski, G. Cottarel, S. Kasif, J. J. Collins, and T. S. Gardner, Jan. 2007, "Large-Scale mapping and validation of escherichia coli transcriptional regulation from a compendium of expression profiles," *PLoS Biol*, vol. 5, no. 1, p. e8.

[131] J. Yu, V. A. Smith, P. P. Wang, A. J. Hartemink, and E. D. Jarvis, Dec. 2004, "Advances to bayesian network inference for generating causal networks from observational biological data," *Bioinformatics (Oxford, England)*, vol. 20, no. 18, pp. 3594–3603.

[132] D. Husmeier, Nov. 2003, "Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic bayesian networks," *Bioinformatics*, vol. 19, no. 17, pp. 2271–2282.

[133] M. Bansal, G. D. Gatta, and D. di Bernardo, Apr. 2006, "Inference of gene regulatory networks and compound mode of action from time course gene expression profiles," *Bioinformatics (Oxford, England)*, vol. 22, no. 7, pp. 815–822.

[134] A. A. Margolin, I. Nemenman, K. Basso, C. Wiggins, G. Stolovitzky, R. D. Favera, and A. Califano, 2006, "ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context," *BMC Bioinformatics*, vol. 7 Suppl 1, p. S7.

[135] Y. Li, Y. Zhu, X. Bai, H. Cai, W. Ji, and D. Guo, Nov. 2009, "ReTRN: a retriever of real transcriptional regulatory network and expression data for evaluating structure learning algorithm," *Genomics*, vol. 94, no. 5, pp. 349–354.

[136] N. Friedman, M. Linial, I. Nachman, and D. Pe'er, "Using bayesian networks to analyze expression data," in *Proceedings of the fourth annual international conference on Computational molecular biology*. Tokyo, Japan: ACM, 2000, pp. 127–135.

[137] D. Heckerman, D. Geiger, and D. M. Chickering, Sep. 1995, "Learning bayesian networks: The combination of knowledge and statistical data," *Machine Learning*, vol. 20, no. 3, pp. 197–243.

[138] S. Russell and P. Norvig. Prentice Hall, Dec. 2002, *Artificial Intelligence: A Modern Approach*, 2nd ed.

[139] J. Gill. Boca Raton: Chapman & Hall/CRC, 2008, *Bayesian methods : a social and behavioral sciences approach*, 2nd ed.

[140] C. Robert. New York: Springer, 2004, *Monte Carlo statistical methods*, 2nd ed.

[141] S. Kirkpatrick, J. Gelatt, C D, and M. P. Vecchi, May 1983, "Optimization by simulated annealing," *Science (New York, N.Y.)*, vol. 220, no. 4598, pp. 671–680.

[142] V. Černý, Jan. 1985, "Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm," *Journal of Optimization Theory and Applications*, vol. 45, no. 1, pp. 41–51.

[143] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, 1953, "Equation of state calculations by fast computing machines," *The Journal of Chemical Physics*, vol. 21, no. 6, p. 1087.

[144] C. Chow and C. Liu, May 1968, "Approximating discrete probability distributions with dependence trees," *IEEE Transactions on Information Theory*, vol. 14, no. 3, pp. 462– 467.

[145] S. Ma, Q. Gong, and H. J. Bohnert, Nov. 2007, "An arabidopsis gene network based on the graphical gaussian model," *Genome Research*, vol. 17, no. 11, pp. 1614–1625.

[146] X. Wu, Y. Ye, and K. R. Subramanian, 2003, "Interactive analysis of gene interactions using graphical gaussian model," *ACM SIGKDD WORKSHOP ON DATA MINING IN BIOINFORMATICS, 3:63–69.*

[147] H. Li and J. Gui, Apr. 2006, "Gradient directed regularization for sparse gaussian concentration graphs, with applications to inference of genetic networks," *Biostatistics*, vol. 7, no. 2, pp. 302 –317.

[148] J. Schaefer, R. Opgen-Rhein, and K. Strimmer, 2006, "Reverse engineering genetic networks using the GeneNet package," *R News*, vol. 6/5, pp. 50–53.

[149] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander, Oct. 1999, "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring," *Science (New York, N.Y.)*, vol. 286, no. 5439, pp. 531–537.

[150] A. Djebbari and J. Quackenbush, 2008, "Seeded bayesian networks: Constructing genetic networks from microarray data," *BMC Systems Biology*, vol. 2, no. 1, p. 57.

[151] S. G. Bøttcher and C. Dethlefsen, 2003, "DEAL: a package for learning bayesian networks," *JOURNAL OF STATISTICAL SOFTWARE*, vol. 8, pp. 200—3.

[152] X. Chen, M. Chen, and K. Ning, 2006, "BNArray: an r package for constructing gene regulatory networks from microarray data by using bayesian network," *Bioinformatics*, vol. 22, no. 23, pp. 2952 –2954.

[153] R. C. Gentleman, V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A. J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J. Y. H. Yang,

and J. Zhang, 2004, "Bioconductor: open software development for computational biology and bioinformatics," *Genome Biology*, vol. 5, no. 10, p. R80.