

RESEARCH

Open Access



# MICRAT: a novel algorithm for inferring gene regulatory networks using time series gene expression data

Bei Yang<sup>1,2\*</sup>, Yaohui Xu<sup>1</sup>, Andrew Maxwell<sup>3</sup>, Wonryull Koh<sup>3</sup>, Ping Gong<sup>4</sup> and Chaoyang Zhang<sup>3\*</sup>

## Abstract

**Background:** Reconstruction of gene regulatory networks (GRNs), also known as reverse engineering of GRNs, aims to infer the potential regulation relationships between genes. With the development of biotechnology, such as gene chip microarray and RNA-sequencing, the high-throughput data generated provide us with more opportunities to infer the gene-gene interaction relationships using gene expression data and hence understand the underlying mechanism of biological processes. Gene regulatory networks are known to exhibit a multiplicity of interaction mechanisms which include functional and non-functional, and linear and non-linear relationships. Meanwhile, the regulatory interactions between genes and gene products are not spontaneous since various processes involved in producing fully functional and measurable concentrations of transcriptional factors/proteins lead to a delay in gene regulation. Many different approaches for reconstructing GRNs have been proposed, but the existing GRN inference approaches such as probabilistic Boolean networks and dynamic Bayesian networks have various limitations and relatively low accuracy. Inferring GRNs from time series microarray data or RNA-sequencing data remains a very challenging inverse problem due to its nonlinearity, high dimensionality, sparse and noisy data, and significant computational cost, which motivates us to develop more effective inference methods.

**Results:** We developed a novel algorithm, MICRAT (Maximal Information coefficient with Conditional Relative Average entropy and Time-series mutual information), for inferring GRNs from time series gene expression data. Maximal information coefficient (MIC) is an effective measure of dependence for two-variable relationships. It captures a wide range of associations, both functional and non-functional, and thus has good performance on measuring the dependence between two genes. Our approach mainly includes two procedures. Firstly, it employs maximal information coefficient for constructing an undirected graph to represent the underlying relationships between genes. Secondly, it directs the edges in the undirected graph for inferring regulators and their targets. In this procedure, the conditional relative average entropies of each pair of nodes (or genes) are employed to indicate the directions of edges. Since the time delay might exist in the expression of regulators and target genes, time series mutual information is combined to cooperatively direct the edges for inferring the potential regulators and their targets. We evaluated the performance of MICRAT by applying it to synthetic datasets as well as real gene expression data and compare with other GRN inference methods. We inferred five 10-gene and five 100-gene networks from the DREAM4 challenge that were generated using the gene expression simulator GeneNetWeaver (GNW). MICRAT was also used to reconstruct GRNs on real gene expression data including part of the DNA-damaged response pathway (SOS DNA repair network) and experimental dataset in *E. Coli*. The results showed that MICRAT significantly improved the inference accuracy, compared to other inference methods, such as TDBN, etc.

(Continued on next page)

\* Correspondence: [iebyang@zzu.edu.cn](mailto:iebyang@zzu.edu.cn); [chaoyang.zhang@usm.edu](mailto:chaoyang.zhang@usm.edu)

<sup>1</sup>School of Information & Engineering, Zhengzhou University, Zhengzhou 450000, China

<sup>3</sup>School of Computing, University of Southern Mississippi, Hattiesburg, MS 39406, USA

Full list of author information is available at the end of the article



(Continued from previous page)

**Conclusion:** In this work, a novel algorithm, MICRAT, for inferring GRNs from time series gene expression data was proposed by taking into account dependence and time delay of expressions of a regulator and its target genes. This approach employed maximal information coefficients for reconstructing an undirected graph to represent the underlying relationships between genes. The edges were directed by combining conditional relative average entropy with time course mutual information of pairs of genes. The proposed algorithm was evaluated on the benchmark GRNs provided by the DREAM4 challenge and part of the real SOS DNA repair network in *E. Coli*. The experimental study showed that our approach was comparable to other methods on 10-gene datasets and outperformed other methods on 100-gene datasets in GRN inference from time series datasets.

**Keywords:** Gene regulatory networks, Maximal information coefficient, Conditional relative average entropy, Time-series mutual information

## Background

Reconstruction of gene regulatory networks (GRNs), also known as reverse engineering [1] of GRNs, aims to infer the potential relationship between genes using gene expression data. Sometimes the expression of a gene is affected by others (named as its regulators) and meanwhile regulates its downstream target genes. Reconstruction of gene regulatory networks can help people understand the mechanism of gene interactions in the cell and reveal the mystery of life. The discovering of the regulation relationship between genes was initially achieved by biological experiments [2]. However, this approach was costly and progressed slowly in research, thus becoming a bottleneck that restricted the development of biological systems. With the development of biotechnology, such as gene chip microarray, the high-throughput data generated provides us with more opportunities to understand the potential regulatory relationships of genes. In recent years, developing machine learning and data mining algorithms and applying them to reconstruction of gene regulatory networks has become an active research topic in bioinformatics.

Different computational methods have been proposed to tackle the gene regulatory networks identification problems. The typical methods include regression models [3, 4], Bayesian networks [5–10], state space models [11–14] and information theory models [15–20], etc.

Several existing methods formulate gene regulatory networks as regression problems which are based on the assumption that all regulatory interactions are linear, meaning that the gene expression level of a target gene is a combination of the expression levels of its transcriptional factors. However, biological networks are known to exhibit multiple regulation mechanisms including non-linear interactions.

The Bayesian network model employed the joint probability distribution of gene expression data to construct a directed acyclic graph to represent the relationship between genes. A relative change ratio (RCR) method was proposed by Li et al. to preprocess the null-mutants steady state data in order to find the key genes and build GRNs, in which

these selected key genes have a higher potential than other genes to play very critical roles [10]. The LBN algorithm (local Bayesian network) [21] was designed by Liu et al. to improve the accuracy of GRN inference from gene expression data by exploring advantages of Bayesian network (BN) and conditional mutual information (CMI) methods. The LBN algorithm first uses CMI to construct an initial network or GRN. Then, the BN method is employed to generate a series of local BNs by selecting the k-nearest neighbors of each gene as its candidate regulatory genes. Integrating these local BNs forms a tentative network by performing CMI. The final network, or GRN, can be obtained by iteratively performing CMI and local BN on the tentative network. TDBN [7] was used to infer GRNs from time series trajectory data which were combined with the previous knowledge gained in the above step. Wang et al. developed the pLasso [22] method from a Bayesian perspective for the reconstruction of gene networks. This method assigns different prior distributions to different edge subsets according to a modified Bayesian information criterion that incorporates prior knowledge on both the network structure and the pathway information. It used information from the Pathway Commons (PC) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) as prior information for the network reconstruction. However, in reality there is no prior knowledge while inferring the interaction relationship between genes. Moreover, signaling pathways are dynamic events that take place over a given period of time. In order to identify these pathways, expression data over time is required. Dynamic Bayesian network (DBN) is an important and widely used approach for predicting the GRNs from time series expression data but has high computational costs and may not be able to infer a large GRN.

Furthermore, other methods for inferring GRNs have been proposed. Patrilia et al. proposed a new algorithm, namely iRafNet [23], which integrated different data types under a unified random forest framework. The key idea of iRafNet is to introduce a weighted sampling scheme within random forest to incorporate information

from other sources of data. Specifically, the model considers the expression of each gene as a function of the expression of other genes. In literature [24], a fast method was developed for inferring gene regulatory relationships from just knockdown data. It used a simple linear regression model focusing on single regulator-target gene pairs based on knockdown data, and allowed the incorporation of prior knowledge about the relationships and generates posterior probabilities.

Mutual information (MI) is another measurement of the dependence of two variables and is often used to assess the relationship between genes. Margolin [16] and Sales [17] employed MI between two genes to measure their associations in order to reconstruct gene regulatory networks. But they could not tackle the problem that one gene is regulated by multiple genes. PCA\_CMI [18] was proposed by Zhang et al. for inferring GRNs from gene expression data considering the non-linear dependence and topological structure of GRNs by employing path consistency algorithm based on conditional mutual information (CMI). It discovered the associations between related genes without identifying the regulators and their targets.

Reshef et al. presented a measure of dependence for two-variable relationships, named the maximal information coefficient (MIC) [25], which captures a wide range of functional and nonfunctional associations. MIC has better performance on measuring the dependence of two-variable relationships in comparison with other methods. Thus, we concentrate our effort on identifying the interactions between genes based on this method.

In this paper, we developed a novel algorithm, MICRAT (Maximal Information coefficient with Conditional Relative Average entropy and Time series mutual information), for inferring GRNs from time series gene expression data. This approach employed MIC as measurement for assessing the correlation between each pair of genes while constructing an undirected graph to represent the underlying relationship between genes. It subsequently combined conditional relative average entropy with time series mutual information to determine the potential regulator and its target genes.

## Methods

Information theory has been used to reconstruct GRNs from gene expression data. Mutual information is generally used as a useful criterion for measuring the dependence between variables (genes)  $X$  and  $Y$  [18]. For gene expression data, variable  $X$  is a vector, in which the element  $x$  denotes gene expression value in different conditions (samples/time points). In this section, we introduced several basic concepts of information theory and graph theory which will be used in the proposed method.

### Mutual information and conditional mutual information

For two random variables (genes)  $X$  and  $Y$ , the mutual information of  $X$  and  $Y$  is defined as:

$$I(X, Y) = H(X) + H(Y) - H(X, Y) \tag{1}$$

where  $H(X)$  and  $H(Y)$  are the entropy of the random variables  $X$  and  $Y$ , respectively;  $H(X, Y)$  is the joint entropy of  $X$  and  $Y$ .

For a discrete variable  $X$ , the entropy  $H(X)$  measures the average uncertainty of variable  $X$ , and can be computed by

$$H(X) = -\sum_{x \in X} p(x) \log p(x) \tag{2}$$

where  $p(x)$  is the probability that the variable  $X$  takes  $x$ ;

The joint entropy of  $X$  and  $Y$  is denoted by

$$H(X, Y) = -\sum_{x \in X, y \in Y} p(x, y) \log p(x, y) \tag{3}$$

where  $p(x, y)$  is the joint probability of  $X$  and  $Y$ .

The mutual information of  $X$  and  $Y$  is calculated by

$$I(X, Y) = \sum_{x \in X, y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \tag{4}$$

The conditional mutual information [16] of  $X$  and  $Y$  given  $Z$ ,  $I(X, Y|Z)$  is defined as

$$I(X, Y|Z) = \sum_{x \in X, y \in Y, z \in Z} p(x, y, z) \log \frac{p(x, y|z)}{p(x|z)p(y|z)} \tag{5}$$

where  $p(x, y, z)$  is the joint probability of variables  $X$ ,  $Y$  and  $Z$ ;  $p(x|z)$  is the conditional probability that the variable  $x$  holds under the condition that the variable  $z$  is established; similarly,  $p(x, y|z)$  is the conditional probability that the variable  $x$  and the variable  $y$  are established under the condition that the variable  $z$  is satisfied.

In this paper, the probability of a variable (gene) is estimated with the Gaussian kernel probability density estimator [26] as follows [18].

$$P(X_i) = \frac{1}{n} \sum_{j=1}^n \frac{1}{(2\pi)^{m/2} |C|^{m/2}} \exp\left(-\frac{1}{2}(X_j - X_i)^T C^{-1} (X_j - X_i)\right) \tag{6}$$

where  $C$  is the covariance matrix of variable  $X$ ,  $|C|$  is the determinant of matrix  $C$ ,  $n$  is the number of samples (time points), and  $m$  is the number of variables (genes).

Then the entropy of variable  $X$  can be computed by using (2) and (6).

$$H(X) = \log\left[(2\pi e)^{\frac{m}{2}} |C|^{\frac{1}{2}}\right] = \frac{1}{2} \log(2\pi e)^m |C| \tag{7}$$

With formulas (1) and (7), the mutual information of variables  $X$  and  $Y$  can be obtained as given in Eq. (8).

$$I(X, Y) = \frac{1}{2} \log \frac{|C(X) \cdot | \cdot |C(Y) |}{|C(X, Y) |} \tag{8}$$

Similarly, the conditional mutual information of variables  $X$  and  $Y$  given variable  $Z$  can be calculated by Eq. (9).

$$I(X, Y|Z) = \frac{1}{2} \log \frac{|C(X, Z) \cdot | \cdot |C(Y, Z) |}{|C(Z) \cdot | \cdot |C(X, Y, Z) |} \tag{9}$$

MI can be used for measuring the dependence of two variables (genes) and CMI can help determine the dependence between two variables given another variable. Generally speaking, a higher MI (or CMI) indicates closer relationship between variables.

**Maximal information coefficient**

The maximal information coefficient (MIC) is an effective measurement of interesting relationship between pairs of variables. MIC captures a wide range of functional and nonfunctional associations, and has two heuristic properties: generality and equitability. The generality ensures that with sufficient sample size MIC captures a wide range of interesting relationships, not limited to specific function types, or even to all functional relationships; and the equitability means that MIC gives similar scores to equally noisy relationships of different types. These two heuristic properties of MIC indicate that it is suitable to infer various underlying relationships.

The Maximal Information Coefficient of a set  $D$  of two-variable data with sample size  $n$  and grid size less than  $B(n)$  is given by literature [25].

$$MIC(D) = \max_{XY < B(n)} M(D)_{X,Y} = \max_{XY < B(n)} \frac{I^*(D, X, Y)}{\log(\min(X, Y))} \tag{10}$$

where the dataset  $D$  is composed of the pairs of values  $\langle x, y \rangle$ , and  $x \in X, y \in Y$ ;  $M(D)_{X, Y}$  is the characteristic matrix of dataset  $D$  of variable  $X$  and  $Y$  with entries

$$M(D)_{x,y} = \frac{I^*(D, x, y)}{\log(\min(x, y))} \tag{11}$$

where  $I^*(D, x, y)$  is the maximum mutual information achieved by  $x$  and  $y$ ;  $n$  is size of  $D$ ;  $\log(\min(x, y))$  is used to normalize the maximum mutual information;  $B(n)$  is the grid size, and  $B(n) = n^{0.6}$ , which was found to work well in practice [25].

**Conditional relative average entropy**

The conditional relative average entropy (CRAE) [27] is defined as:

$$CRAE(X \rightarrow Y) = \frac{H(Y|X)}{H(Y) \cdot |Y|} \tag{12}$$

where  $|Y|$  is the number of values of variable  $Y$ , and

$H(Y)$  is the entropy of the variable  $Y$ ;  $H(Y|X)$  is the conditional entropy of the variable  $Y$  under the condition of the given variable  $X$ .

**Time series mutual information**

In most cases, there is a time delay during the gene regulations or at least the gene expressions change simultaneously for regulatory and target genes. Thus the time series mutual information could contribute to the orientation of the regulation.

Suppose that we have  $T$  time points and the expression levels of  $N$  genes are measured at each time point. The time series gene expression data can be summarized as a  $T \times N$  matrix  $X = (x_1, x_2, \dots, x_T)^T$  whose  $i^{\text{th}}$  row vector  $x_i = (x_{i1}, x_{i2}, \dots, x_{iN})^T$  corresponds to a gene expression level vector measured at timepoint  $i$ .

For given genes  $X$  and  $Y$ , the time series mutual information from  $X$  to  $Y$  during time points 1 to  $T$  is defined as

$$I_T(X \rightarrow Y) = I(X_{1 \rightarrow (T-tdelay)}, Y_{1+tdelay \rightarrow T}) \tag{13}$$

where  $tdelay(0 \leq tdelay \leq T)$  represents the time delay for gene  $X$  regulating gene  $Y$ ;  $X_{1 \rightarrow (T-tdelay)}$  is a vector representing the gene expression level of  $X$  from timepoint 1 to timepoint  $(T - tdelay)$ . Similarly,  $Y_{1+tdelay \rightarrow T}$  represents the gene expression level of  $Y$  from timepoint  $1 + tdelay$  to timepoint  $T$ . In this paper, we set the time delay to 1 since from the observation of our experiments there is no significant improvement to the accuracy with different settings.

**Gene regulatory networks**

In this work, a gene regulatory network is described by a directed graph (digraph). A graph  $G$  is defined by the pair  $(V(G), E(G))$ , where  $V(G)$  denotes the set of vertices (nodes) and  $E(G) \subseteq V(G) \times V(G)$  denotes the set of edges. In a digraph, an edge is defined by an ordered pair of vertices  $(x, y)$  denoting the edge direction, from vertex  $x$  pointing to vertex  $y$ . Here, the nodes represent genes while the edges describe the gene regulatory interactions. A directed edge  $(x, y)$  indicates that gene  $x$  regulates gene  $y$ , while an undirected edge  $(x, y)$  only indicates that there is an association between gene  $x$  and gene  $y$ .

**Algorithm of MICRAT**

We propose a novel algorithm called MICRAT for reconstructing gene regulatory networks using time series gene expression profiles. The algorithm mainly consists of two procedures: First, an undirected graph is generated representing the associations between genes based on the maximal information coefficient (MIC) of each pair of genes. Then, the conditional relative

average entropy combining with time series mutual information is used to determine the directions of the edge in the undirected graph. The flowchart of the method for reconstructing GRN is shown in Fig. 1, and more details are described in the following subsections.

**Generating an undirected gene interaction graph**

The types of relationship between genes vary diversely. It is difficult, even impossible to identify the gene associations with only one function. The maximal information coefficient could capture a wide range of dependence of two variables, both functional and nonfunctional. Thus it is suitable for measuring the associations between genes.

We refer to two gene expression data  $G_X$  and  $G_Y$  as variables  $X$  and  $Y$ , respectively. The association between them is measured by

$$I_D(X, Y) = MIC(X, Y) \tag{14}$$

where  $MIC(X, Y)$  denotes the maximal information coefficient of  $X$  and  $Y$ . A higher value of  $I_D(X, Y)$  indicates a closer relationship between  $X$  and  $Y$ , vice versa. The edges in the undirected graph are added based on a given a threshold  $\theta$  of  $I_D(X, Y)$ .

The generation of an undirected gene interaction graph includes the following five steps:

- (1) Initialize the gene regulatory network,  $G_D$ , as an undirected graph with no vertices and no edges. There are  $n$  genes in the dataset  $D$ .
- (2) For a given time series gene expression dataset  $D$ , calculate  $I_D(X, Y)$  for each pair of genes  $(x, y)$ .

- (3) Rank the pairs of genes by descending order of  $I_D(X, Y)$  obtained by step (2) to a list called  $RLL$ .
- (4) From the top of  $RLL$ , check each pair of genes  $(x, y)$ , add an undirected edge between vertices  $x$  and  $y$  if  $I_D(X, Y) \geq \theta$  where  $\theta$  is a threshold and will be discussed in the experiment section.
- (5) Delete redundant edges. There might be some redundant edges in the undirected graph especially when there are triangle cycles. Conditional mutual information of genes is used to determine whether to delete edges or not. For example, there is a triangle cycle including three genes  $X, Y$  and  $Z$  in the graph since  $I_D(X, Y), I_D(X, Z)$  and  $I_D(Y, Z)$  are not lower than  $\theta$  and three edges between  $X, Y$  and  $Z$  are included. We calculated three conditional mutual information values of  $I(X, Y | Z), I(X, Z | Y)$  and  $I(Y, Z | X)$ , and deleted edge  $(Y, Z)$  if  $I(Y, Z | X) = 0$ , which indicates that gene  $Y$  has no relationship with gene  $Z$ .

**Data discretization and edge-orientation**

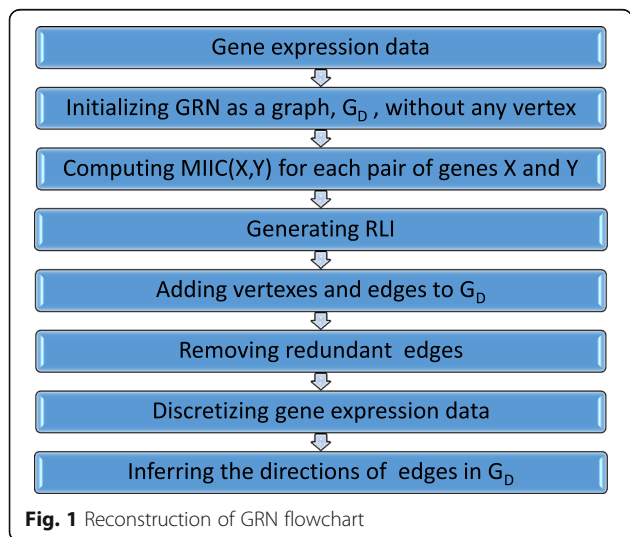
In our algorithm, the edges in the undirected graph are oriented by combining the conditional relative average entropy with time series mutual information of each pair of nodes. From the observation of our experiments, better performance was obtained on discretized data. Therefore, the gene expression data were discretized before being used for edge orientation.

The gene expression data are first normalized by standard fraction  $Z$ -Score and then discretized according to a given threshold. The standard fraction  $Z$ -score of the gene  $X$  at the time point  $t_j$  is defined as:

$$Z_{x,j} = \frac{|x_j - \mu|}{\sigma} \tag{15}$$

where  $\mu$  denotes the mean value of the gene expression data at all time points of gene  $X$ ,  $\sigma$  is the standard deviation, and  $x_j$  denotes the expression value of the gene  $X$  at the time point  $t_j$ . Given a threshold  $k$ , if  $Z_{x,j} \geq k$ , then the gene expression data at time  $t_j$  is denoted as 1; otherwise the gene expression data is denoted as 0.

All edges in the undirected graph are oriented by the following procedure: Given two vertices  $X$  and  $Y$  of an edge in the undirected graph, if  $CRAE(X \rightarrow Y) + I_T(X \rightarrow Y) > CRAE(Y \rightarrow X) + I_T(Y \rightarrow X)$ , then the orientation is  $X \rightarrow Y$ , indicating that gene  $X$  regulates gene  $Y$ ; While if  $CRAE(X \rightarrow Y) + I_T(X \rightarrow Y) < CRAE(Y \rightarrow X) + I_T(Y \rightarrow X)$ , then the orientation is  $Y \rightarrow X$ , which means gene  $Y$  regulates gene  $X$ . In the case of  $CRAE(X \rightarrow Y) + I_T(X \rightarrow Y) = CRAE(Y \rightarrow X) + I_T(Y \rightarrow X)$ , the higher value of  $I_T(X \rightarrow Y)$  (or  $I_T(Y \rightarrow X)$ ) indicates  $X \rightarrow Y$  (or  $Y \rightarrow X$ ). The pseudo code of algorithm MICRAT is given in Fig. 2.



**Fig. 1** Reconstruction of GRN flowchart

**Input:** Time series gene expression dataset  $D$  including  $n$  genes and each gene profile composed of  $T$  time points, MIC value threshold  $\theta$

**Output:** Gene regulatory network  $G_D$

```

1  $G_D \leftarrow \emptyset$ .
2 for  $i \leftarrow 1$  to  $n - 1$ 
3   for  $j \leftarrow i + 1$  to  $n$ 
4      $mic_{i,j} = mic(D(:, i), D(:, j))$ ;
5     if  $mic_{i,j} > \theta$  and  $I(i, j | \alpha) \neq 0$  then
6        $dir(i, j) = CRAE(D(i, j)) + I_T(i \rightarrow j)$ ;
7        $dir(j, i) = CRAE(D(j, i)) + I_T(j \rightarrow i)$ ;
8       if  $dir(i, j) > dir(j, i)$  then
9          $G_D(i, j) = 1$ ;
10      elseif  $dir(i, j) < dir(j, i)$  then
11         $G_D(j, i) = 1$ ;
12      elseif  $I_T(X \rightarrow Y) \geq I_T(Y \rightarrow X)$  then
13         $G_D(i, j) = 1$ ;
14      else
15         $G_D(j, i) = 1$ ;
16 return  $G_D$ .
```

**Fig. 2** Pseudo code of algorithm MICRAT

## Results and discussion

In this section, we evaluate the performance of algorithm MICRAT by applying it to synthetic datasets as well as a real gene expression dataset for inferring GRNs. Several experiments were carried out to demonstrate the performance of our approach by comparison with other related methods.

For synthetic datasets, five 10-genes and five 100-genes time series expression datasets from the DREAM4 challenge were used, respectively [28]. The real life gene expression dataset involves an eight-gene subnetwork, part of a DNA-damage response pathway (SOS pathway) in the bacteria *E. Coli*.

Generally, there are several measurements for evaluating the performance of GRN inference methods, including *Precision*, *Recall*, *Accuracy (ACC)*, *F<sub>score</sub>*, *Matthews correlation coefficient (MCC)*, etc. They are defined as follows:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$ACC = \frac{TP + TN}{TP + FP + TN + FN}$$

$$F_{score} = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

where  $TP$ ,  $FP$ ,  $TN$ ,  $FN$  are the numbers of true positives, false positives, true negatives and false negatives, respectively. In this study, true positives indicate correctly

inferred edges which exist in golden standard networks, while false positives denote wrongly inferred edges which do not exist in golden standard networks, and so on for true negatives and false negatives, respectively. Usually, there exists an inverse relationship between precision and recall. It is possible to increase one at the cost of reducing the other. Therefore, *Accuracy*, *F<sub>score</sub>* and *Matthews correlation coefficient* are widely used as balanced evaluation measures. Here we use *ACC*, *F<sub>score</sub>* and *MCC* as measurements for evaluating the performance of our method.

In order to set the best value of  $\theta$ , the threshold of MIC of two gene expression data while measuring the independence between genes, we test MICRAT by varying  $\theta$  from 0 to 1 with interval of 0.01 on both 10-gene expression datasets and 100-gene expression datasets. Empirical results showed that with the increase of  $\theta$ , the precision of the model becomes higher while the recall goes down. According to the experimental results, the comparatively good performance of MICRAT could be obtained while the thresholds were set to 0.43 for 10-gene datasets and 0.15 for 100-gene datasets, respectively.

### Experiments on simulation datasets

We first evaluated MICRAT on simulation datasets which were generated based on benchmarking networks. Many tools have been developed for assessing the effectiveness of GRN inference methods. The DREAM4 challenge introduced a critical performance assessment framework of methods for GRN inference and presented an in silico benchmark suite as a blinded, community-wide challenge within the context of the DREAM project. In this challenge, the gene expression datasets with noise and their golden standard (benchmark) networks were given. The golden standard networks were selected from source

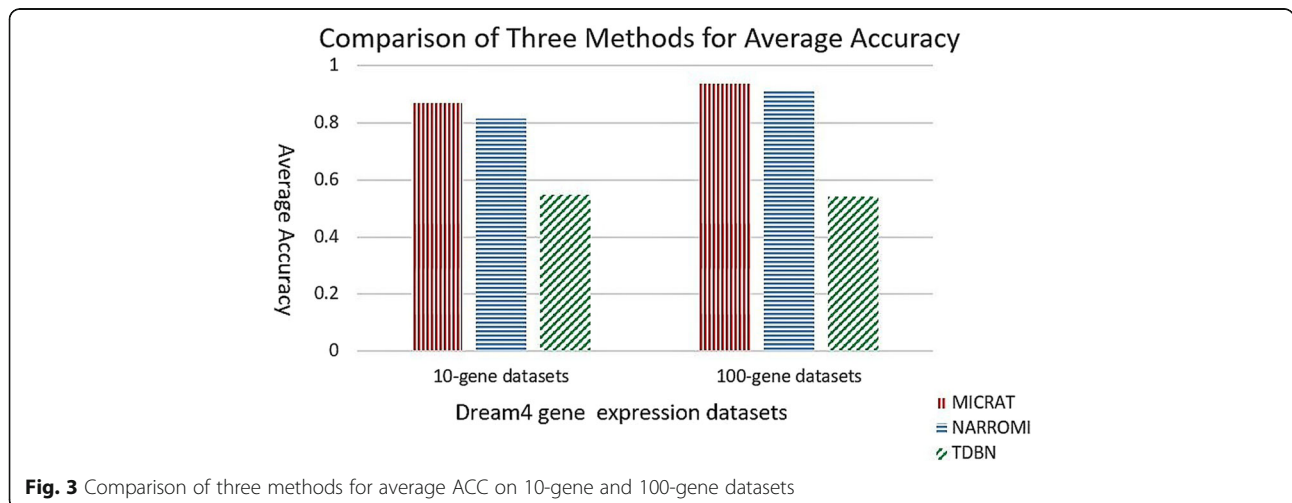
**Table 1** Experimental results for three methods on both 10-gene datasets and 100-gene datasets

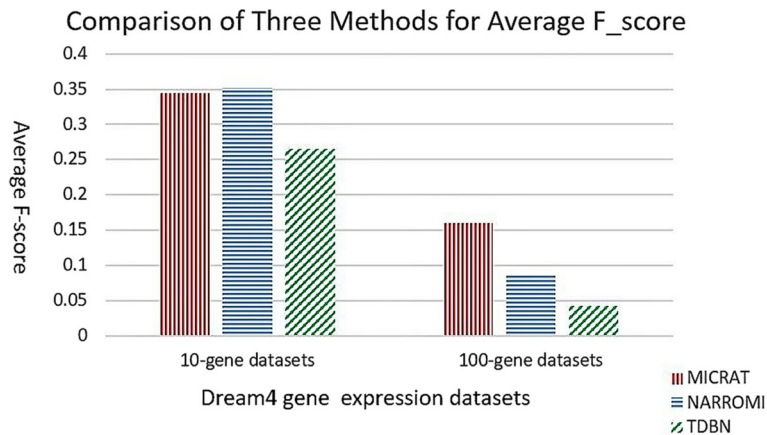
| Dataset    | Accuracy |         |        | F_score |         |        | MCC   |         |        |
|------------|----------|---------|--------|---------|---------|--------|-------|---------|--------|
|            | TDBN     | NARROMI | MICRAT | TDBN    | NARROMI | MICRAT | TDBN  | NARROMI | MICRAT |
| 10-gene-1  | 0.59     | 0.81    | 0.88   | 0.27    | 0.30    | 0.40   | 0.06  | 0.19    | 0.42   |
| 10-gene-2  | 0.54     | 0.76    | 0.83   | 0.33    | 0.29    | 0.19   | 0.12  | 0.15    | 0.15   |
| 10-gene-3  | 0.57     | 0.81    | 0.88   | 0.24    | 0.34    | 0.33   | 0.00  | 0.23    | 0.42   |
| 10-gene-4  | 0.51     | 0.82    | 0.88   | 0.24    | 0.40    | 0.25   | 0.03  | 0.30    | 0.28   |
| 10-gene-5  | 0.52     | 0.88    | 0.87   | 0.25    | 0.40    | 0.43   | 0.07  | 0.34    | 0.36   |
| 100-gene-1 | 0.52     | 0.92    | 0.93   | 0.05    | 0.09    | 0.16   | 0.04  | 0.08    | 0.17   |
| 100-gene-2 | 0.57     | 0.91    | 0.93   | 0.06    | 0.06    | 0.12   | 0.04  | 0.03    | 0.10   |
| 100-gene-3 | 0.48     | 0.91    | 0.94   | 0.04    | 0.10    | 0.19   | 0.01  | 0.09    | 0.18   |
| 100-gene-4 | 0.57     | 0.92    | 0.94   | 0.03    | 0.09    | 0.16   | -0.03 | 0.07    | 0.15   |
| 100-gene-5 | 0.56     | 0.91    | 0.94   | 0.04    | 0.09    | 0.16   | 0.01  | 0.08    | 0.16   |

networks of real species [28]. They are widely used as a benchmark for the evaluation of GRN inference methods. Here, we tested our algorithm on the DREAM4 challenge time series gene expression data in networks of size 10 and 100 genes, respectively.

As mentioned above, MICRAT found the associations between pairs of genes in the beginning. Then the conditional relative average entropy in cooperation with time series mutual information was used to orient the edges in the undirected graph and obtain the GRN. For evaluating the performance of our approach, we compared our algorithm with TDBN [7, 10] and NARROMI [19]. TDBN is a typical method with significant impact on inferring GRNs from time series gene expression data. NARROMI is quite related to our method and it significantly outperformed other methods, such as LP [29], LASSO [30, 31], ARACNE [16] and GENIE3 [32]. This procedure was tested on DREAM4 time series gene expression datasets with 10-gene networks and 100-gene networks, respectively. Experimental results were given in Table 1. First, we carried out MICRAT on five

10-gene datasets. There are  $21 \times 5 = 105$  time points in each of the datasets. We computed ACC, F\_score and MCC on each of the five datasets as evaluation measures for the compared methods. Here, we chose 0.43 as the threshold value of MIC to determine the dependence between genes in the inferred network for our method, and the threshold of Z\_score was set to  $k = 1.2$  for discretization of gene expression data when the edge is oriented. For NARROMI, the threshold is set to 0.05 which is the default value given in literature [19]. For TDBN, all the parameters were chosen as default in their literature. The accuracy of our approach is higher than those of the other methods in four datasets and slightly lower than that of NARROMI for dataset 5. The left part of Fig. 3 shows the average accuracies on five 10-gene datasets for three methods. It can be observed that MICRAT performed better than the others. For the F\_score measurement, MICRAT has the best performance on two datasets, but the worst performance on one dataset. Figure 4 shows that the average F\_scores on five 10-gene datasets is somewhat inferior to NARROMI but





**Fig. 4** Comparison of three methods for average F\_score on 10-gene and 100-gene datasets

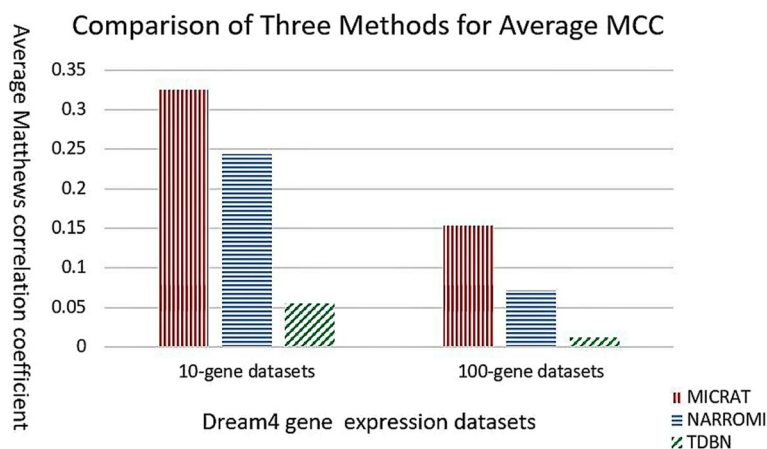
superior to TDBN. From Table 1 it can be observed that MICRAT is only slightly inferior to NARROMI on one dataset but has better performance on others for MCC. Figure 5 gives the average Matthews correlation coefficients of the three methods on five 10-gene datasets and five 100-gene datasets, respectively. We can see that MICRAT is superior to NARROMI and significantly outperforms TDBN.

We also applied our algorithm to the five DREAM4 100-gene datasets. Each dataset has  $21 \times 10 = 210$  time points. We also computed *ACC*, *F\_score* and *MCC* on each of the five datasets as evaluation measures for MICRAT and the other two methods. Experimental results were showed in Table 1 and the right part of Figs. 3, 4 and 5. In these experiments, we chose 0.15 as the threshold of MIC to determine the regulation relationship between genes in the inferred network. For NARROMI, the threshold is also set to 0.05, the same value as in literature [19]. As shown in Table 1 and the right part of Figs. 3, 4 and 5, all the three

measures of *ACC*, *F\_score* and *MCC* in our method are superior to NARROMI and TDBN on each of the five 100-gene time series dataset. The results show that our approach has better performance than other methods on the 100-gene datasets. Thus MICRAT has the potential to infer large GRNs using time series expression data with a large number of genes.

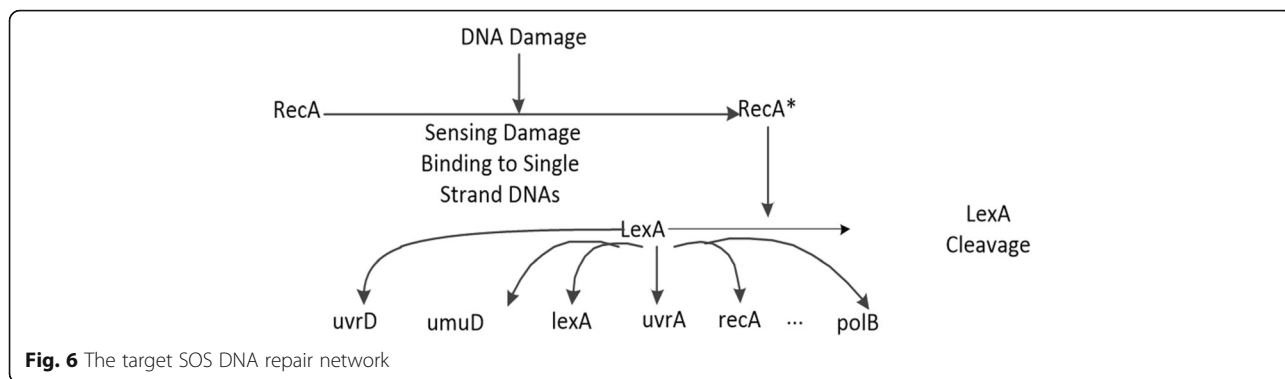
**Experiments on real gene expression dataset**

To validate our method on a real biological gene regulatory network, we analyzed the well-known SOS DNA repair network in *E. Coli*. This GRN is well known for its responsibility of repairing the DNA if it gets damaged. It is the largest, most complex and best understood DNA damage-inducible network to be characterized to date [33]. We applied our MICRAT algorithm to an eight-gene network, part of the SOS DNA repair network in the bacteria *E. coli* [34–36]. The expression data sets of the network were obtained from Uri Aron Lab [36]. These data are kinetics of 8 genes namely *uvrD*, *lexA*,



**Fig. 5** Comparison of three methods for average MCC on 10-gene and 100-gene datasets



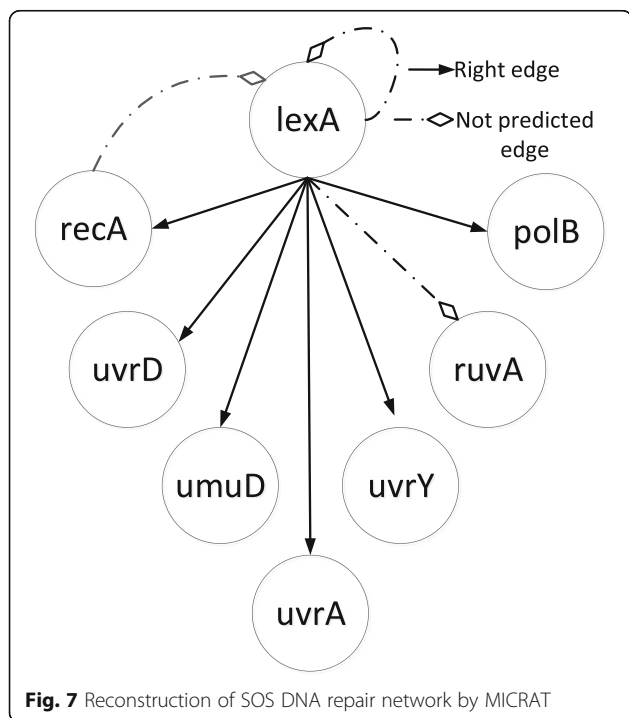


umuD, recA, uvrA, uvrY, ruvA and polB. Four experiments were done at various UV light intensities (Exp.1 and 2:  $5Jm^{-2}$ , Exp. 3 and 4:  $20Jm^{-2}$ ). In each experiment, the 8 genes were monitored at 50 instants which are evenly spaced by 6 min intervals. In order to assess the effectiveness of our algorithm, we compared the inferred network with the known interactions between these eight genes. In our experiment, the threshold of MIC was set to  $\theta=0.1$  for measuring the dependence of the pair of genes, and the threshold of  $Z\_score$  was also set to  $k=1.2$  for discretization of gene expression data while the edge is oriented. Figure 6 gives the real SOS DNA repair network. Figure 7 shows part of inferred GRN inferred by our approach. In this Figure, one can see that our method finds 6 out of the 9 edges in the target network and identifies lexA as the ‘hub’ gene for this

network. The exact ground truth for this network is not precisely known, hence it is not possible to calculate well-known performance measures [33]. By using the known interactions obtained from literature [33, 37, 38], Table 2 showed the comparison of our algorithm with other methods such as Perrin [39], BANJO [40], GlobalMIT [41] and Morshed [33] for correct predictions on the regulatory relationships between these eight genes. From Table 2, it could be observed that our method identified most interactions except  $lexA \rightarrow ruvA$ ,  $lexA \rightarrow lexA$  and  $recA \rightarrow lexA$ , where the latter two are not inferred since our method does not consider the double regulation between two genes and the gene that is self-regulated. That is the limitation of our method, which motivates us to continue further study for improvement.

**Conclusions**

In this work, a novel algorithm, MICRAT, for inferring GRNs from time series gene expression data was proposed by taking into account dependence and time delay of expressions of a regulator and its target genes. This approach employed maximal information coefficients for reconstructing an undirected graph to represent the



**Table 2** Comparison of interactions inferred by MICRAT with other methods on SOS DNA dataset

| Regulators | Target genes | MICRAT  | Perrin [39] | BANJO [40] | GlobalMIT [41] | Morshed [33] |
|------------|--------------|---------|-------------|------------|----------------|--------------|
| lexA       | recA         | correct | correct     |            | correct        | correct      |
|            | lexA         |         | correct     |            |                |              |
|            | umuD         | correct |             |            | correct        | correct      |
|            | uvrD         | correct |             |            |                | correct      |
|            | uvrA         | correct | correct     | correct    | correct        | correct      |
|            | polB         | correct |             |            | correct        |              |
|            | uvrY         | correct |             |            | correct        |              |
| recA       | lexA         |         | correct     | correct    |                |              |

underlying relationships between genes. The edges were directed by combining conditional relative average entropy with time course mutual information of pairs of genes. The performance of our proposed algorithm was evaluated using synthetic datasets from benchmark GRNs provided by the DREAM4 challenge and the real, experimental dataset SOS DNA repair network in *E. Coli*. Experimental study showed that our approach was comparable to other methods on 10-gene datasets and outperformed other methods on 100-gene datasets in GRN inference from time series datasets. In the follow-up study, we plan to test the algorithm on large-scale datasets, including time course mRNA-Seq, human gene expression data, etc. and compare our approach with more recently proposed algorithms so as to evaluate its robustness and improve its performance.

#### Abbreviations

ACC: Accuracy; CMI: Conditional mutual information; CRAE: Conditional relative average entropy; GRN: Gene regulatory network; MCC: Matthews correlation coefficient; MI: Mutual information; MIC: Maximal information coefficient

#### Funding

This work was partially supported by a subaward of the National Science Foundation EPS 0903787 and the U.S. Army Environmental Quality/Installation (EQI) Basic Research Program. Permission was granted by the Chief of Engineers to publish this information. It was also partially supported by the Key Research Projects of Universities in Henan Province of China under grant number 17A520015. The publication cost of this article was funded by the U.S. Army Environmental Quality/Installation (EQI) Basic Research Program.

#### Availability of data and materials

The datasets used and/or analysed during the current study are available at <http://dreamchallenges.org/project/dream4-in-silico-network-challenge/>. The codes of the algorithm of MICRAT are available from the corresponding author on reasonable request.

#### About this supplement

This article has been published as part of *BMC Systems Biology Volume 12 Supplement 7, 2018: From Genomics to Systems Biology*. The full contents of the supplement are available online at <https://bmcsystbiol.biomedcentral.com/articles/supplements/volume-12-supplement-7>.

#### Authors' contributions

BY, CZ and PG conceived the project. YX and BY developed and implemented the MICRAT algorithms and conducted the experimental study. BY and CZ wrote the paper. PG, AM and WK participated in the analysis and provided suggestions for the document. All authors read and approved the final manuscript.

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

#### Author details

<sup>1</sup>School of Information & Engineering, Zhengzhou University, Zhengzhou 450000, China. <sup>2</sup>Center of Precision Medicine, Zhengzhou University, Zhengzhou 450000, China. <sup>3</sup>School of Computing, University of Southern Mississippi, Hattiesburg, MS 39406, USA. <sup>4</sup>Environmental Lab, US Army Engineer Research and Development Center, Vicksburg, MS 39180, USA.

Published: 14 December 2018

#### References

- Vijesh N, Chakrabarti SK, Sreekumar J. Modeling of gene regulatory networks: a review. *J Biomed Sci Eng.* 2013;06(1):223–31. [http://file.scirp.org/pdf/JBISE\\_2013022716483315.pdf](http://file.scirp.org/pdf/JBISE_2013022716483315.pdf). Accessed 03 July 2018.
- Lemmens K, Dhollander T, Bie TD, et al. Inferring transcriptional modules from ChIP-chip, motif and microarray data. *Genome Biol.* 2006;7(5):R37. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.279.1345&rep=rep1&type=pdf>. Accessed 03 July 2018.
- Singh N, Vidyasagar M. bLARS: an algorithm to infer gene regulatory networks. *IEEE/ACM Trans Comput Biol Bioinform.* 2016;13(2):1. <https://www.computer.org/csdl/trans/tb/2016/02/07138615.pdf>. Accessed 03 July 2018.
- Haury AC, et al. TIGRESS: trustful inference of gene regulation using stability selection. *BMC Syst Biol.* 2012;6(1):145. <https://core.ac.uk/download/pdf/51228782.pdf>. Accessed 03 July 2018.
- Friedman N, Linal M, Nachman I, et al. Using Bayesian network to analyze expression data. *J Comput Biol.* 2000;7(3–4):601–20. [www.cs.huji.ac.il/~nir/Papers/FLNP1Full.pdf](http://www.cs.huji.ac.il/~nir/Papers/FLNP1Full.pdf). Accessed 03 July 2018.
- Murphy K, Mian S. Modeling gene expression data using dynamic Bayesian networks. Technical Report, Computer Science Division. Berkeley: University of California; 1999. [www.cs.ubc.ca/~murphyk/Papers/ismb99.pdf](http://www.cs.ubc.ca/~murphyk/Papers/ismb99.pdf). Accessed 03 July 2018.
- Zou M, Conzen SD. A new dynamic Bayesian network (DBN) approach for identifying gene regulatory networks from time course microarray data. *Bioinformatics.* 2005;21(1):71–9. <https://academic.oup.com/bioinformatics/article/21/1/71/212416>. Accessed 03 July 2018.
- Yeung KY, Dombek KM, Lo K, et al. Construction of regulatory networks using expression time-series data of a genotyped population. *Proc Natl Acad Sci U S A.* 2011;108(48):19436–41. [www.pnas.org/content/108/48/19436.full](http://www.pnas.org/content/108/48/19436.full). Accessed 03 July 2018.
- Haoni L, et al. Learning the structure of gene regulatory networks from time series gene expression data. *BMC Genomics.* 2011;12(Suppl 5):S13. <https://www.ncbi.nlm.nih.gov/pubmed/22369588>. Accessed 03 July 2018.
- Li P, et al. Gene regulatory network inference and validation using relative change ratio analysis and time-delayed dynamic Bayesian network. *EURASIP J Bioinform Syst Biol.* 2014;2014:12. <https://link.springer.com/article/10.1186/s13637-014-0012-3>. Accessed 03 July 2018.
- Wu X, et al. State space model with hidden variables for reconstruction of gene regulatory networks. *BMC Syst Biol.* 2011;5(Suppl 3):S3. <http://europepmc.org/articles/PMC3287571>. Accessed 03 July 2018.
- Wu FX. Gene regulatory network modelling: a state-space approach. *Int J Data Min Bioinform.* 2008;2(1):1–14. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.192.4775&rep=rep1&type=pdf>. Accessed 03 July 2018.
- Osamu H, Ryo Y, Seiya I, Rui Y, Tomoyuki H, Charnock-Jones DS, Cristin P, Satoru M. Statistical inference of transcriptional module-based gene networks from time course gene expression profiles by using state space models. *Bioinformatics.* 2008;24:932–42. <https://academic.oup.com/bioinformatics/article/24/7/932/295736>. Accessed 03 July 2018.
- Kojima K, Rui Y, Seiya I, Mai Y, Masao N, Ryo Y, Tepei S, Kazuko U, Tomoyuki H, Noriko G, Satoru M. A state space representation of VAR models with sparse learning for dynamic gene networks. *Genome Inform.* 2009;22:56–68. <https://www.jsbi.org/pdfs/journal/1/IBSB09/IBSB09006.pdf>. Accessed 03 July 2018.
- Wang J, Chen B, Wang Y, et al. Reconstructing regulatory networks from the dynamic plasticity of gene expression by mutual information. *Nucleic Acids Res.* 2013;41(8):395–408. <https://academic.oup.com/nar/article/41/8/e97/2409387>. Accessed 03 July 2018.
- Margolin AA, Nemenman I, Basso K, et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics.* 2006;7(1):S7. <https://arxiv.org/pdf/q-bio/0410037>. Accessed 03 July 2018.
- Sales G, Romualdi C. Parmigene—a parallel R package for mutual information estimation and gene network reconstruction. *Bioinformatics.*

- 2011;27(13):1876–7 (2). <https://academic.oup.com/bioinformatics/article/27/13/1876/184634>. Accessed 03 July 2018.
18. Zhang X, Zhao XM, He K, et al. Inferring gene regulatory networks from gene expression data by path consistency algorithm based on conditional mutual information. *Bioinformatics*. 2012;28(1):98–104. <https://doi.org/10.1093/bioinformatics/btq373>. Accessed 03 July 2018.
  19. Zhang X, et al. NARROMI: a noise and redundancy reduction technique improves accuracy of gene regulatory network inference. *Bioinformatics*. 2013;29(1):106–13. <https://academic.oup.com/bioinformatics/article/29/1/106/13>. Accessed 03 July 2018.
  20. Chaitankar V, Ghosh P, Perkins EJ, et al. A novel gene network inference algorithm using predictive minimum description length approach. *BMC Syst Biol*. 2010;4(Suppl 1):S7. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2880413>. Accessed 03 July 2018.
  21. Liu F, Zhang SW, Guo WF, Wei ZG, Chen L. Inference of gene regulatory network based on local Bayesian networks. *PLoS Comput Biol*. 2016;12(8):e1005024. <https://doi.org/10.1371/journal.pcbi.1005024> <http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005024>. Accessed 03 July 2018.
  22. Wang Z. Incorporating prior knowledge into gene network study. *Bioinformatics*. 2013;29(20):2633–40. <https://academic.oup.com/bioinformatics/article/29/20/2633/277562>. Accessed 03 July 2018.
  23. Petralia F, Wang P, Yang J, Zhidong T. Integrative random forest for gene regulatory network inference. *Bioinformatics*. 2015;31:197–205. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4542785/>. Accessed 03 July 2018.
  24. Young WC, Raftery AE, Yeung KY. A posterior probability approach for gene regulatory network inference in genetic perturbation data. *Math Biosci Eng*. 2017;13(6):1241–51. <https://arxiv.org/abs/1603.04835>. Accessed 03 July 2018.
  25. Reshef DN, Reshef YA, Finucane HK, et al. Detecting novel associations in large data sets. *Science*. 2011;334(6062):1518–24. <http://science.sciencemag.org/content/334/6062/1518>. Accessed 03 July 2018.
  26. Basso K, et al. Reverse engineering of regulatory networks in human B cells. *Nat Genet*. 2005;37:382–92. <https://s3-us-west-2.amazonaws.com/oww-files-public/b/b2/Basso.pdf>. Accessed 03 July 2018.
  27. Jiang J, Wang J, Yu H, et al. Poison identification based on Bayesian network: a novel improvement on K2 algorithm via Markov blanket[M]// advances in swarm intelligence: Springer Berlin Heidelberg; 2013. p. 173–82. [https://link.springer.com/chapter/10.1007/978-3-642-38715-9\\_21](https://link.springer.com/chapter/10.1007/978-3-642-38715-9_21). Accessed 03 July 2018.
  28. Dream4 In Silico Network Challenge. <http://dreamchallenges.org/project/dream4-in-silico-network-challenge/>. Accessed 03 July 2018.
  29. Wang Y, Joshi T, Zhang XS, et al. Inferring gene regulatory networks from multiple microarray datasets. *Bioinformatics*. 2006;22(22):2413–20. <https://academic.oup.com/bioinformatics/article/22/19/2413/240982>. Accessed 03 July 2018.
  30. Tibshirani R. Regression shrinkage and selection via the lasso. *J Royal Stat Soc*. 1996;58(1):267–88. <http://statweb.stanford.edu/~tibs/ftp/lasso-retro.pdf>. Accessed 03 July 2018.
  31. Geeven G, et al. Identification of context-specific gene regulatory networks with GEMULA-gene expression modeling using LAsso. *Bioinformatics*. 2012;28(2):214–21. <https://academic.oup.com/bioinformatics/article/28/2/214/198473>. Accessed 03 July 2018.
  32. Huynh-Thu VA, et al. Inferring regulatory networks from expression data using tree-based methods. *PLoS One*. 2010;5(9):e12776. <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0012776>. Accessed 03 July 2018.
  33. Morshed, et al. Simultaneous learning instantaneous and time-delayed genetic interactions using novel information theoretic scoring technique. *BMC Syst Biol*. 2012;6:62. [https://link.springer.com/chapter/10.1007%2F978-3-642-24958-7\\_29](https://link.springer.com/chapter/10.1007%2F978-3-642-24958-7_29). Accessed 03 July 2018.
  34. Ronen M, Rosenberg R, Shraiman BI, et al. Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics. *Proc Natl Acad Sci U S A*. 2002;99(16):10555–60. [www.pnas.org/lookup/doi/10.1073/pnas.152046799](http://www.pnas.org/lookup/doi/10.1073/pnas.152046799). Accessed 03 July 2018.
  35. Shenorr SS, Milo R, Mangan S, et al. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat Genet*. 2002;31(1):64–8. [https://www.weizmann.ac.il/mcb/UriAlon/sites/mcb/UriAlon/files/network\\_motifs\\_in\\_coli\\_0.pdf](https://www.weizmann.ac.il/mcb/UriAlon/sites/mcb/UriAlon/files/network_motifs_in_coli_0.pdf). Accessed 03 July 2018.
  36. Uri Alon's SOS Dataset webpage. <https://www.weizmann.ac.il/mcb/UriAlon/download/downloadable-data>. Accessed 03 July 2018.
  37. Noman N, Iba H. Inferring gene regulatory networks using differential evolution with local search heuristics. *IEEE/ACM Trans Comput Biol Bioinform*. 2007;4(4):643–7. <https://www.ncbi.nlm.nih.gov/pubmed/17975274>. Accessed 03 July 2018.
  38. Kamura S, et al. Inference of S-system models of genetic networks using a cooperative coevolutionary algorithm. *Bioinformatics*. 2005;21(7):1154–63. <https://academic.oup.com/bioinformatics/article/21/7/1154/268773>. Accessed 03 July 2018.
  39. Perrin BE, Ralaivola L, Mazurie A, et al. Gene networks inference using dynamic Bayesian networks. *Bioinformatics*. 2003;19(suppl 2):ii138–48. [https://academic.oup.com/bioinformatics/article/19/suppl\\_2/ii138/180436](https://academic.oup.com/bioinformatics/article/19/suppl_2/ii138/180436). Accessed 03 July 2018.
  40. Yu J, Smith VA, Wang PP, et al. Advances to Bayesian network inference for generating causal networks from observational biological data. *Bioinformatics*. 2004;20(18):3594–603. <https://www.semanticscholar.org/paper/Advances-to-Bayesian-network-inference-for-causal-Yu-Smith/adcbcb725490d64d12b4f795e1e381ca6b8de4b4>. Accessed 03 July 2018.
  41. Vinh NX, Chetty M, Coppel R, et al. GlobalMIT: learning globally optimal dynamic bayesian network with the mutual information test criterion. *Bioinformatics*. 2011;27(19):2765–6. <https://academic.oup.com/bioinformatics/article/27/19/2765/231220>. Accessed 03 July 2018.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

