Molecular BioSystems RSCPublishing

Supplementary Information

Inferring cellular regulatory networks with Bayesian Model Averaging for Linear Regression (BMALR)

Xun Huang[a] and Zhike Zi[a,b*]

[a]*BIOSS Centre for Biological Signalling Studies, University of Freiburg, 79104, Freiburg, Germany.*
[b]*Present address: Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany*
[*]*Corresponding author, E-mail: zhike.zi@molgen.mpg.de*

This Supplementary Information contains the supplementary text for the derivation of Equation 6 in the main text, other technical details for calculating probability density of the data, as well as Supplementary Figures.

1. The derivation of probability density in Gaussian Bayesian network

Given a Gaussian Bayesian network, we can uniquely generate a multivariate (nonsingular) normal distribution and vice versa[1, 2]. Suppose we have an n-dimensional random vector (X), which follows a multivariate normal distribution:

$$\rho(X \mid u,r) = N(X;u,r^{-1}) \tag{S1}$$

where the parameter u is the mean vector of this distribution, and the parameter r is the inverse covariance matrix of this distribution. To make the model decomposable, the global parameter independence and parameter modularity should be satisfied[1, 3]. Therefore, the parameter u is assumed to follow a multivariate normal distribution, and the parameter r is assumed to follow a Wishart distribution[1]. Then, the prior distributions of u and r are given below:

$$\rho(u \mid r) = N(u;\mu,(\upsilon \cdot r)^{-1}) \tag{S2}$$

$$\rho(r) = Wishart(r;\alpha,\beta) \tag{S3}$$

where the hyperparameter μ is the mean vector of the distribution of u. The hyperparameter υ is a positive number. The hyperparameter α is the degree of freedom of the Wishart distribution, and α should be greater than (n-1). The hyperparameter β is an n-by-n scale matrix, and should be positive-definite.
If a dataset D is generated by the random vector X, which satisfies the distribution mentioned above, there is a close-formed expression of $\rho(D_W)$ that is the probability density of the data restricted to the subset W of the variables in X. In addition, we use the parameter M to stand for the sample size of D, l_W for the number of elements in W. Then, the probability density of D_W will satisfy the following equation[1, 2]:

$$\rho(D_W) = \left(\frac{1}{2\pi}\right)^{\frac{Ml_W}{2}} \left(\frac{\upsilon}{\upsilon+M}\right)^{\frac{l_W}{2}} \left(\frac{c(l_W,\alpha_W))}{c(l_W,\alpha_W+M)}\right) \left(\frac{\left|\beta_W\right|^{\frac{\alpha_W}{2}}}{\left|\beta_W^*\right|^{\frac{\alpha_W+M}{2}}}\right) \tag{S4}$$

where

$$c(n,\alpha) = \left[2^{n\alpha/2}\pi^{n(n-1)/4}\prod_{i=1}^{n}\Gamma(\frac{\alpha+1-i}{2})\right]^{-1} \tag{S5}$$

$$\beta_W = \left((\beta^{-1})^{(W)}\right)^{-1} \tag{S6}$$

$$\beta_W^* = \beta_W + s_W + (\frac{\upsilon M}{\upsilon + M})(\overline{X}_W - \mu^{(W)})(\overline{X}_W - \mu^{(W)})^T \quad \text{(S7)}$$

$$s_W = \sum_{i=1}^{M} (X_W[i] - \overline{X}_W)(X_W[i] - \overline{X}_W)^T \quad \text{(S8)}$$

$$\alpha_W = \alpha - n + l_W \quad \text{(S9)}$$

2. Default parameter settings for probability density in Gaussian Bayesian network

To calculate the probability density of data $\rho(D_W)$ with the Equation S5 listed above, we need to specify four parameters μ, υ, α, and β, which are the hyperparameters shown in Equation S3 and Equation S4. The parameter μ is the mean vector of the parameter u in Equation S2. In BMALR, we use the mean vector of the real sample ($\overline{X}$) as the default value for the parameter μ. The parameter υ is the size of the hypothetical sample upon which we base our prior belief concerning the value of u[1, 2]. Therefore, we use the sample size M as the default value for the parameter υ. The parameter α can be regarded as the effective sample size for determining the parameter r in Equation S2, and it should be greater than $(n\text{-}1)$[1, 2]. In BMALR, we use n as the default value for parameter α. The parameter β can be regarded as covariance matrix in the hypothetical sample upon which we based on our prior belief of u. It is an n-by-n positive definite matrix. We assume the prior probability of all the edges is uniformly distributed for simplification. With such flat prior, β will be a diagonal matrix. Therefore, we set the default value of β as H, where H is the diagonal matrix of the covariance matrix of the sample data.
In summary, the parameter settings for calculating probability density of the data, $\rho(D_W)$, is listed as below:

$$\mu = \overline{X} \quad \text{(S10)}$$

$$\upsilon = M \quad \text{(S11)}$$

$$\alpha = n \quad \text{(S12)}$$

$$\beta = H \quad \text{(S13)}$$

where H is a n-by-n diagonal matrix, and $H(i,i)$ = variance(X_i), for i = 1, 2 , … , n.
With the default parameter settings in Equation S10-S13, Equation S4 can reformate as below:

$$\rho(D_W) = \left(\frac{1}{2\pi}\right)^{\frac{M \cdot l_W}{2}} \left(\frac{1}{2}\right)^{\frac{l_W}{2}} \left(\frac{c(l_W, l_W)}{c(l_W, l_W + M)}\right) \left(\frac{\left|H^{(W)}\right|^{\frac{M}{2}}}{\left|E^{(W)} + M \cdot R^{(W)}\right|^{\frac{l_W + M}{2}}}\right) \quad \text{(S14)}$$

where $E^{(W)}$ is a l_W-by-l_W identity matrix.

The likelihood that the local structure G_{Pa} generates the data $D_{Pa,Xi}$, $P(D_{Pa,Xi}|G_{Pa})$ can be rewritten as:

$$\begin{aligned} P(D_{Pa,X_i} \mid G_{Pa}) &= \frac{\rho(D_{Pa,Xi})}{\rho(D_{Pa})} \\ &= \left(\frac{1}{2\pi}\right)^{\frac{M \cdot (l_{Pa,Xi} - l_{Pa})}{2}} \left(\frac{1}{2}\right)^{\frac{l_{Pa,Xi} - l_{Pa}}{2}} \left(\frac{c(l_{Pa,Xi}, l_{Pa,Xi}) c(l_{Pa}, l_{Pa} + M)}{c(l_{Pa}, l_{Pa}) c(l_{Pa,Xi}, l_{Pa,Xi} + M)}\right) \left(\frac{\left|H^{(Pa)}\right|^{\frac{M}{2}} \left|E^{(Pa)} + M \cdot R^{(Pa)}\right|^{\frac{l_{Pa} + M}{2}}}{\left|H^{(Pa,Xi)}\right|^{\frac{M}{2}} \left|E^{(Pa,Xi)} + M \cdot R^{(Pa,Xi)}\right|^{\frac{l_{Pa,Xi} + M}{2}}}\right) \\ &= \left(\frac{1}{2\pi}\right)^{\frac{M}{2}} \left(\frac{1}{2}\right)^{\frac{1}{2}} \left(\frac{c(l_{Pa,Xi}, l_{Pa,Xi}) c(l_{Pa}, l_{Pa} + M)}{c(l_{Pa}, l_{Pa}) c(l_{Pa,Xi}, l_{Pa,Xi} + M)}\right) \left(\frac{\left|E^{(Pa)} + M \cdot R^{(Pa)}\right|^{\frac{l_{Pa} + M}{2}}}{\left|H^{(Xi)}\right|^{\frac{M}{2}} \left|E^{(Pa,Xi)} + M \cdot R^{(Pa,Xi)}\right|^{\frac{l_{Pa,Xi} + M}{2}}}\right) \end{aligned} \quad \text{(S15)}$$

To further increase the efficiency, we make an approximation: $\frac{c(l_{Pa,Xi}, l_{Pa,Xi}) c(l_{Pa}, l_{Pa} + M)}{c(l_{Pa}, l_{Pa}) c(l_{Pa,Xi}, l_{Pa,Xi} + M)} \approx 1$, therefore,

$$P(D_{Pa,X_i} \mid G_{Pa}) \approx \left(\frac{1}{2\pi}\right)^{\frac{M}{2}} \left(\frac{1}{2}\right)^{\frac{1}{2}} \left(H^{(Xi)}\right)^{-\frac{M}{2}} \left(\frac{\left|E^{(Pa)} + M \cdot R^{(Pa)}\right|^{\frac{l_{Pa} + M}{2}}}{\left|E^{(Pa,Xi)} + M \cdot R^{(Pa,Xi)}\right|^{\frac{l_{Pa,Xi} + M}{2}}}\right) \quad \text{(S16)}$$

The posterior probability of an edge feature f is computed with

$$P(f \mid D) \approx \sum_{Pa \in S_i} f(G_{Pa}) P(G_{Pa} \mid D_{Pa,X_i}) = \frac{\sum_{Pa \in S_i} f(G_{Pa}) P(D_{Pa,X_i} \mid G_{Pa}) P(G_{Pa})}{\sum_{Pa \in S_i} P(D_{Pa,X_i} \mid G_{Pa}) P(G_{Pa})} \quad \text{(S17)}$$

It is worth noting that replacing $P(D_{Pa,X_i} \mid G_{Pa})$ in equation S17 with equation S16 leads to the cancellation of the item $\left(\frac{1}{2\pi}\right)^{\frac{M}{2}} \left(\frac{1}{2}\right)^{\frac{1}{2}} \left(H^{(Xi)}\right)^{-\frac{M}{2}}$. Therefore, the item $\left(\frac{1}{2\pi}\right)^{\frac{M}{2}} \left(\frac{1}{2}\right)^{\frac{1}{2}} \left(H^{(Xi)}\right)^{-\frac{M}{2}}$ can be ignored in the implementation of the algorithm. Therefore, we made the following simplification for calculating $\rho(D_W)$:

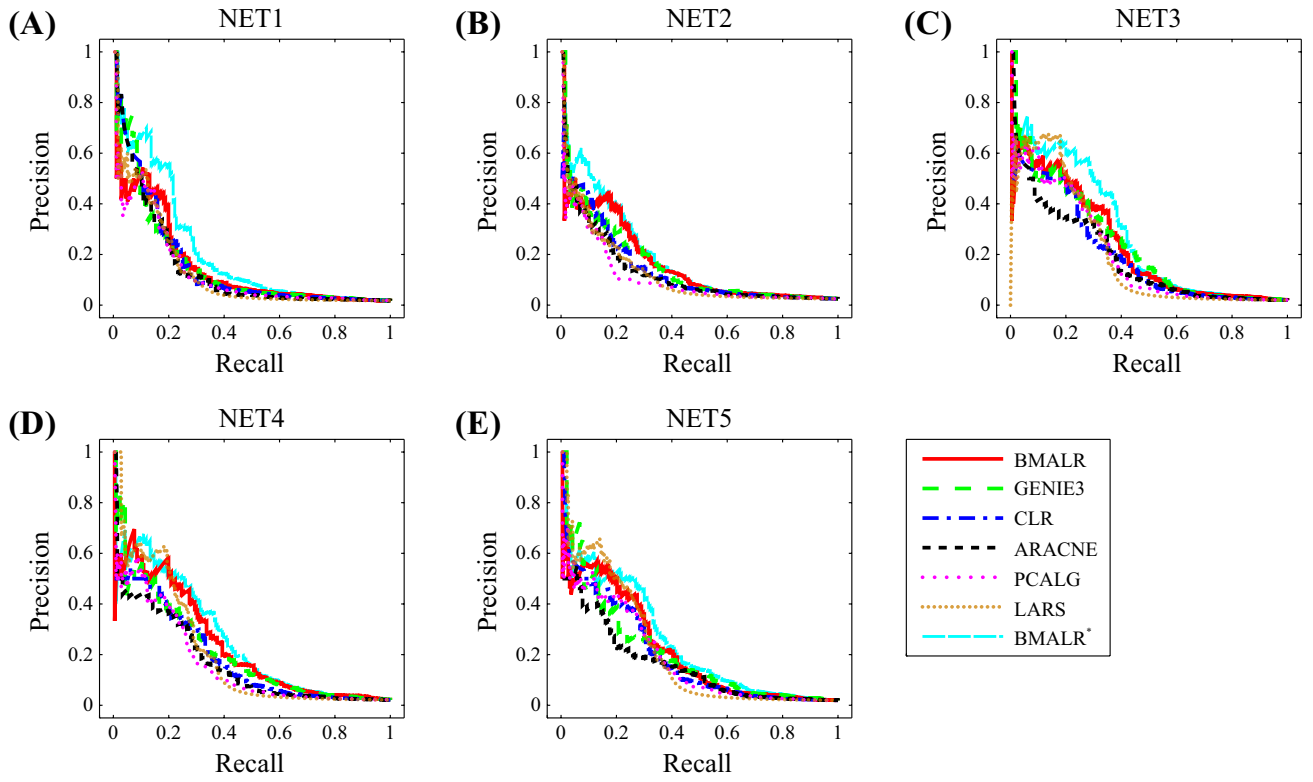
$$\rho(D_W) \propto \left(\frac{1}{2\pi}\right)^{\frac{M \cdot l_W}{2}} \left(\frac{1}{2}\right)^{\frac{l_W}{2}} \left(\frac{\left|H^{(W)}\right|^{-\frac{M}{2}}}{\left|E^{(W)} + M \cdot R^{(W)}\right|^{\frac{l_W + M}{2}}} \right) \propto \left|E^{(W)} + M \cdot R^{(W)}\right|^{\frac{-l_W - M}{2}} \quad \text{(S18)}$$

We noticed that when the parameter M is too large (e.g. > 100), loss of precision often appears in the calculation of equation S18. In practice, one can set M with a value between 20 and 100 to avoid this problem and get an approximate value for equation S18 in case that the sample size is too large. In BMALR, we set $M = 50$.

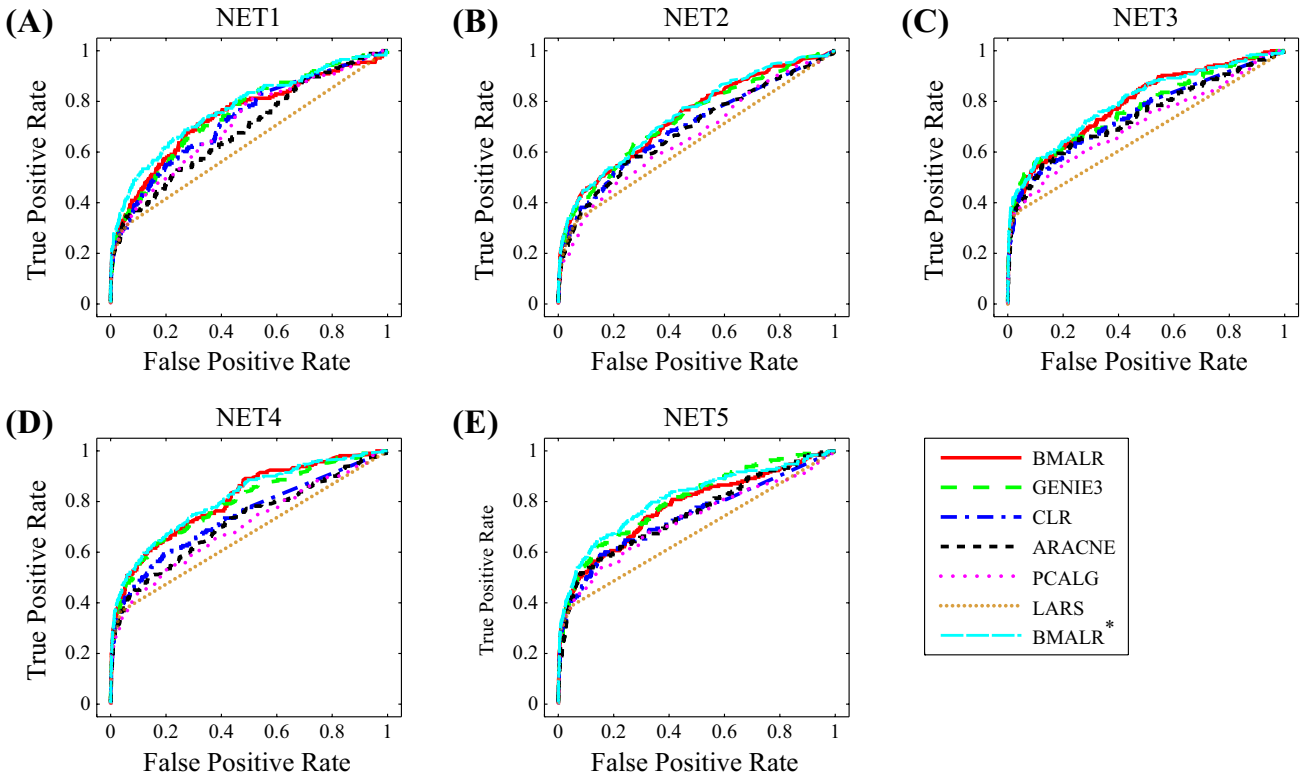
3. References

1. D. Geiger and D. Heckerman, *Ann Stat*, 2002, 30, 1412-1440.
2. R. E. Neapolitan, *Learning bayesian networks*, Pearson Prentice Hall Upper Saddle River, 2004.
3. D. Pe'er, *Sci STKE*, 2005, 2005, pl4.

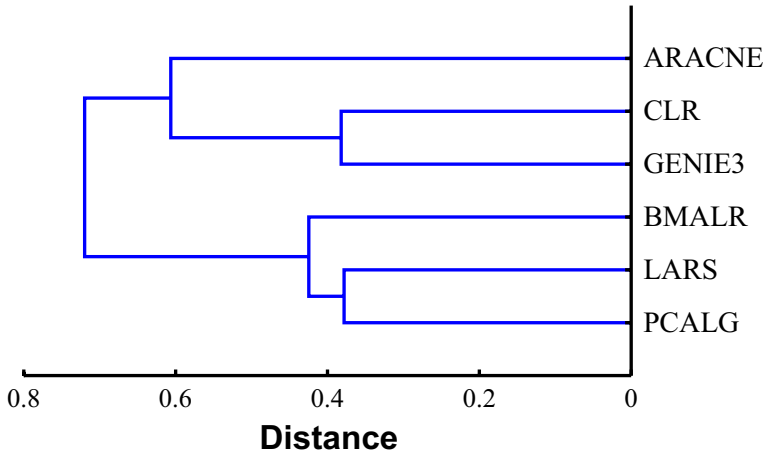
4. Supplementary Figures



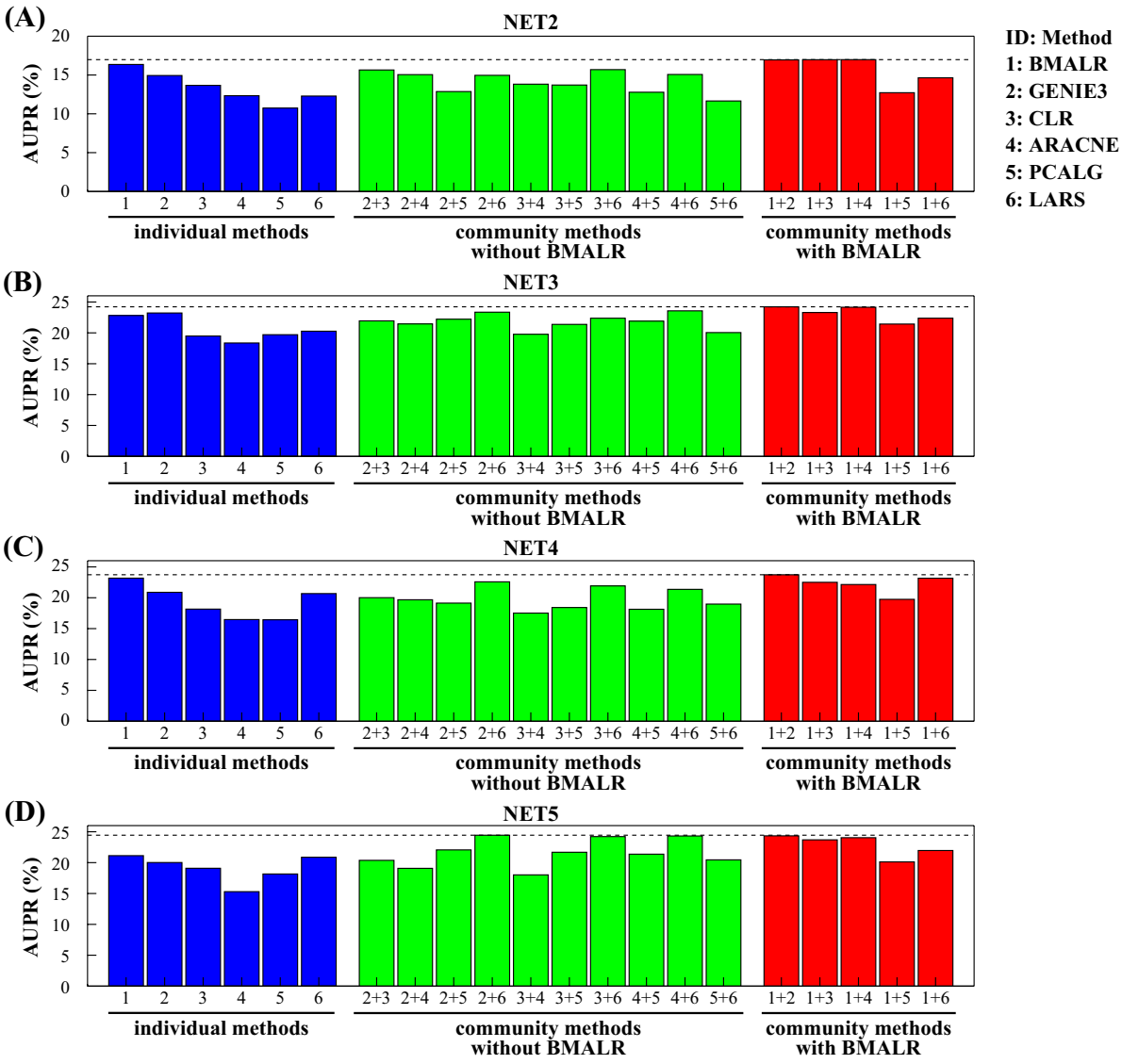
Supplementary Figure 1 PR curves for DREAM4 *in silico* Size100 Multifactorial Challenge. BMALR* denotes the results of BMALR with the log transformation of the datasets. (A) For gold standard network 1 (NET1). (B) For gold standard network 2 (NET2). (C) For gold standard network 3 (NET3). (D) For gold standard network 4 (NET4). (E) For gold standard network 5 (NET5).



Supplementary Figure 2 ROC curves for DREAM4 *in silico* Size100 Multifactorial Challenge. BMALR* denotes the results of BMALR with the log transformation of the datasets. (A) For gold standard network 1 (NET1). (B) For gold standard network 2 (NET2). (C) For gold standard network 3 (NET3). (D) For gold standard network 4 (NET4). (E) For gold standard network 5 (NET5).



Supplementary Figure 3 Similarity of the inference methods clustered by their predicted networks on the DREAM5 network inference sub-challenge.



Supplementary Figure 4 The performance (area under precision-recall curve, AUPR) of individual methods and community methods with combinations of every two individual methods on DREAM4 *in silico* size 100 multifactorial sub-challenge. (A) For gold standard network 2 (NET2). (B) For gold standard network 3 (NET3). (C) For gold standard network 4 (NET4). (D) For gold standard network 5 (NET5). The dashed horizontal line denotes the highest performance level in all the methods.

Supplementary Tables

Supplementary Table 1 The influence of *maxFanIn* on the performance of BMALR in DREAM4 *in silico* size 100 multifactorial sub-challenge.

MaxFanIn	NET1		NET2		NET3		NET4		NET5		Score	Time (second)
	AUPR	AUROC	AUPR	AUROC	AUPR	AUROC	AUPR	AUROC	AUPR	AUROC		
2	0.155	0.745	0.166	0.737	0.231	0.792	0.234	0.808	0.214	0.778	39.4	1.8E+00
3	0.151	0.742	0.163	0.735	0.232	0.798	0.231	0.806	0.214	0.790	39.4	5.7E+01
4	0.148	0.743	0.159	0.732	0.230	0.791	0.227	0.796	0.212	0.793	38.6	1.8E+03

Supplementary Table 2 The influence of *maxFanIn* on the performance of BMALR in DREAM5 network inference sub-challenge with *in silico* datasets.

maxFanIn	AUPR	AUROC	Optimal *F*-score	Time (second)
2	0.320	0.809	0.372	1.3E+02
3	0.321	0.809	0.374	8.3E+03

Supplementary Table 3 The influence of *maxFanIn* on the performance of BMALR in the benchmark of the T-cell signaling network.

Method	AUPR	AUROC	Optimal F-score	*p*-value of Fisher's exact test	Time (second)
2	0.343	0.720	0.500	2.70E-04	1.3E-01
3	0.345	0.726	0.500	2.70E-04	1.4E-01
4	0.337	0.709	0.465	9.44E-04	1.6E-01
5	0.338	0.709	0.462	9.44E-04	2.0E-01
6	0.338	0.709	0.462	9.44E-04	2.2E-01
7	0.338	0.709	0.462	9.44E-04	2.4E-01