# Approximate inference of gene regulatory network models from RNA-Seq time series data

Article

Published Version

It is advisable to refer to the publisher's version if you intend to cite from the work.  See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1186/s12859-018-2125-2

Publisher: BioMed Central

www.reading.ac.uk/centaur

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

## METHODOLOGY ARTICLE

**Open Access**

# Approximate inference of gene regulatory network models from RNA-Seq time series data

Thomas Thorne

## Abstract

**Background:** Inference of gene regulatory network structures from RNA-Seq data is challenging due to the nature of the data, as measurements take the form of counts of reads mapped to a given gene. Here we present a model for RNA-Seq time series data that applies a negative binomial distribution for the observations, and uses sparse regression with a horseshoe prior to learn a dynamic Bayesian network of interactions between genes. We use a variational inference scheme to learn approximate posterior distributions for the model parameters.

**Results:** The methodology is benchmarked on synthetic data designed to replicate the distribution of real world RNA-Seq data. We compare our method to other sparse regression approaches and find improved performance in learning directed networks. We demonstrate an application of our method to a publicly available human neuronal stem cell differentiation RNA-Seq time series data set to infer the underlying network structure.

**Conclusions:** Our method is able to improve performance on synthetic data by explicitly modelling the statistical distribution of the data when learning networks from RNA-Seq time series. Applying approximate inference techniques we can learn network structures quickly with only moderate computing resources.

## Background

Methods for the inference of gene regulatory networks from RNA-Seq data are currently not as mature as those developed for microarray datasets. Normalised microarray data posses the desirable property of being approximately normally distributed so that they are readily amenable to various forms of inference, and in the literature many graphical modelling schemes have been explored that exploit the normality of the data [1–9].

The data generated by RNA-Seq studies on the other hand present a more challenging inference problem, as the data are no longer approximately normally distributed, and before normalisation take the form of non-negative integers. In the detection of differential expression in RNA-Seq data, negative binomial distributions have been applied [10–13], providing a good fit to the overdispersion typically seen in the data relative to a Poisson distribution. Following similar graphical modelling approaches as applied in the analysis of microarray data, it is natural to consider Poisson and negative binomially distributed graphical models. Unfortunately in many cases

when applying graphical modelling approaches with Poisson distributed observations, only models that represent negative conditional dependencies are available, or inference is significantly complicated due to lack of conjugacy between distributions [14]. Poisson graphical models have been applied successfully in the analysis of miRNA regulatory interactions [15, 16], but we might expect to improve on these by modelling the overdispersion seen in typical RNA-Seq data sets with a negative binomial model.

One specific case of interest in the analysis of RNA-Seq data is the study of time series in a manner that takes into account the temporal relationships between data points. Previous work in the literature has developed sophisticated models for the inference of networks from microarray time series data [4, 5], but whilst methods have been developed for the analysis of differential behaviour in RNA-Seq time series [17, 18], little attention has been given to the task of learning networks from such data. Although existing nonparametric methods applicable to time series may be applied [19, 20], these were not specifically designed for application to RNA-Seq data, and also require time consuming approaches such as Markov Chain Monte Carlo schemes. There are also existing information theoretic methods, for example those of [21], but

Correspondence: t.thorne@reading.ac.uk
Department of Computer Science, University of Reading, Reading, UK

again these were designed for application to microarray data, and are not designed for time series data and learning of directed networks.

Here we present a method for the inference of networks from RNA-Seq time series data through the application of a Dynamic Bayesian Network (DBN) approach, that models the RNA-Seq count data as being negative binomially distributed conditional on the expression levels of a set of predictors. Whilst there has been work applying negative binomial regularised regression approaches in the literature [22], here we specifically consider the problem of learning networks from RNA-Seq data, and apply the horseshoe prior [23, 24], that has been shown to have advantages in robustness and adaptivity over other regularisation methods.

## Methods

### Dynamic Bayesian Networks

In a DBN framework [25], considering only edges between time points, we can model a sequence of observations using a first order Markov model, where the value of a variable at time $t$ is dependent only on the values of a set of parent variables at time $t-1$. This is illustrated in Fig. 1 and can be written as

$$p\left(X_t^i | X_{t-1}\right) = p\left(X_t^i | X_{t-1}^{\mathrm{Pa}(i)}\right) \tag{1}$$

where Pa($i$) is the set of parents of variable $i$ in the network. In our case we wish to model the expression level of a gene conditional on a set of parent genes that have some influence on it. To learn the set of parent variables of a given gene, it is possible to perform variable selection in a Markov Chain Monte Carlo framework, proposing to add or remove genes to the parent set in a Metropolis-Hastings sampler. Another option is to consider all possible sets of parent genes as suggested in [20].
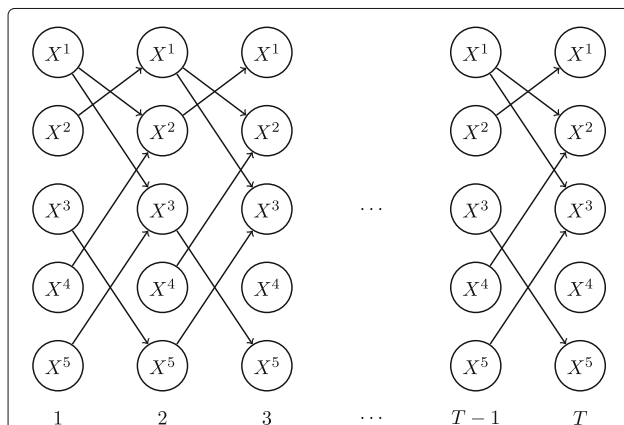


**Fig. 1** DBN of five random variables $X^1, \ldots, X^5$ over $T$ time steps. Variables are conditionally independent when conditioned on their parent variables (incoming arrows)

However for even modestly sized sets of genes (e.g. 50) this can be computationally expensive, and so instead we consider applying a sparse regression approach to learn a set of parents for each gene. This approach considers the contribution of all possible parent genes in a regression framework but encourages sparsity in the coefficients so that only a small set are non-zero.

### Sparse negative binomial regression

Given data $D$ consisting of $M$ columns and $L$ rows, with columns corresponding to genes and rows to time points, we seek to learn a parent set for each gene. To do so we can employ a regularised regression approach that enforces sparsity of the regression coefficients, and only take predictors (genes) whose coefficients are significantly larger than zero as parents. To simplify the presentation, below we consider the regression of the counts for a single gene $i$, $y = D_{2:L,i}$, conditional on the counts of the remaining $W = M-1$ genes $X = D_{1:(L-1),-i}$. The matrix $X$ is supplemented with a column vector **1** to include a constant term in the regression. Where there are multiple replicates for each time point these can be adjusted appropriately.

The counts $y_t$ are then modelled as following a negative binomial distribution with mean $\exp(X\beta)_t$ and dispersion $\omega$, where $\beta$ is a vector of regression coefficients $\beta_w$ and a constant term $\beta_c$. The model for a gene $i$ is then

$$y_t \sim \mathrm{NB}\left(s_t \exp(X_{t-1}\beta)_t, \omega\right), \tag{2}$$

where we have applied the NB2 formulation of the negative binomial distribution, and $s_t$ is a scaling factor for each sample to account for sequencing depth. The $s_t$ can be estimated from the data by considering the sum of counts for each sample, or by the more robust approach of [11] where the median of ratios is used. We place a straightforward normal prior on $\beta_c$ and to enforce sparsity of the $\beta_w$ we apply a horseshoe prior [23, 24], assuming that $\beta_w \sim \mathcal{N}(0, \zeta_w^2)$, and placing a half-Cauchy prior on the $\zeta_w^2$,

$$\beta_w \sim \mathcal{N}\left(0, \zeta_w^2\right) \tag{3}$$

$$\zeta_w \sim \mathrm{C}^+(0, \tau). \tag{4}$$

Then as in [24] we set a prior on $\tau$ that allows the degree of shrinkage to be learnt from the data

$$\tau \sim \mathrm{C}^+(0, \sigma) \tag{5}$$

$$p\left(\sigma^2\right) \propto \frac{1}{\sigma^2}. \tag{6}$$

An example of the sparsity induced in the $\beta_w$ can be seen in figure 8 in Appendix 2. Finally we place a gamma

prior on the dispersion parameter $\omega$. This gives a joint probability (subsuming the model parameters into $\theta$) of

$$p(y, \theta | X) = \prod_i p(y_i | X, \beta, \omega) p(\omega)$$
$$\prod_w p\left(\beta_w | \zeta_w^2\right) p\left(\zeta_w^2 | \tau\right) p(\tau | \sigma) p\left(\sigma^2\right) \quad (7)$$

### Variational Inference

We now apply a variational inference [26–30] scheme to learn approximate posterior distributions over the model parameters. In a Bayesian setting variational inference aims to approximate the posterior $p(\theta | x)$ with a distrubtion $q(\theta)$. To do so we attempt to minimise the Kullback-Leibler (KL) divergence between the two, defined as

$$\mathrm{KL}(q(\theta) || p(\theta | x)) = \int q(\theta) \log \frac{q(\theta)}{p(\theta | x)} d\theta \quad (8)$$
$$= \mathbb{E}_q \left[\log q(\theta)\right] - \mathbb{E}_q \left[\log p(\theta, x)\right]$$
$$+ \log p(x). \quad (9)$$

As the KL divergence is bounded below by zero, it follows from 9 that

$$\log p(x) = \mathrm{KL}(q(\theta) || p(\theta | x)) - \mathbb{E}_q \left[\log q(\theta)\right] \quad (10)$$
$$+ \mathbb{E}_q \left[\log p(\theta, x)\right]$$
$$\log p(x) \geq \mathbb{E}_q \left[\log p(\theta, x)\right] - \mathbb{E}_q \left[\log q(\theta)\right], \quad (11)$$

and so we can define a lower bound on the logarithm of the model evidence as

$$\mathcal{L}(q) = \mathbb{E}_q \left[\log p(\theta, x)\right] - \mathbb{E}_q \left[\log q(\theta)\right]. \quad (12)$$

To make the problem of minimising the KL divergence tractable we can consider a mean field approximation where the posterior is approximated by a series of independent distributions $q(\theta_i)$ on some partition of the parameters,

$$p(\theta | x) \approx q(\theta) = \prod_i q(\theta_i). \quad (13)$$

Under the mean field assumption it can be shown that to minimise the KL divergence between $p(\theta | x)$ and $q(\theta)$, or equivalently to maximise the model evidence lower bound (or ELBO) $\mathcal{L}(q)$, the optimal form for each $q(\theta_i)$ is

$$\log \hat{q}(\theta_i) = \mathbb{E}_{q_{j \neq i}} \left[\log p(\theta, x)\right] + \text{const.} \quad (14)$$

where the expectation is over the remaining $q(\theta_{j \neq i})$. In many cases this formalism is sufficient to derive a coordiante ascent algorithm to maximise the ELBO where the variational parameters of the $\hat{q}(\theta_i)$ are updated iteratively.

Unfortunately in our model the optimal distribution $\hat{q}$ for the regression coefficients $\beta_w$ does not have a tractable solution. However following [31] we can sidestep this problem by applying non-conjugate variational message passing [32], and we can then derive approximate posterior distributions for each of the model parameters

following a straightforward parameter update scheme. The full set of variational updates are given in Appendix 1.

Considering our model as a graphical model as in Fig. 2, we can decompose the terms of $\mathbb{E}_{q_{j \neq i}} \left[\log p(\theta, x)\right]$ in Eq. 14 into those dependent on $\theta_i$ by considering the neighbours of $\theta_i$. Then we can rewrite Eq. 14 as

$$\log \hat{q}(\theta_i) = \mathbb{E}_q \left[\log p(\theta_i | \theta_{\mathrm{Pa}_i})\right] + \sum_{k \in \mathrm{Ch}_i} \mathbb{E}_q \left[\log p(\theta_k | \theta_{\mathrm{Pa}_k \neq i})\right]$$
$$(15)$$

where $\mathrm{Ch}_i$ denotes the children of node $i$ in the graphical model. Considering each term on the right hand side of Eq. 15 as a message from another variable in the graphical model it is possible to derive $\hat{q}$ in the conjugate exponential family as in [33]. In the non-conjugate case, the messages can be approximated as in [32], derived for the negative binomial model in [31].

## Results

### Synthetic data

We apply our method to the task of inferring directed networks from simulated gene expression time series. The time series were generated by utilising the GeneNetWeaver [34] software to first generate subnetworks representative of the structure of the



**Fig. 2** Graphical model representation of our statistical model. Applying variational message passing, the approximating distribution $\hat{q}$ of a random variable can be updated based on messages passed from connected nodes
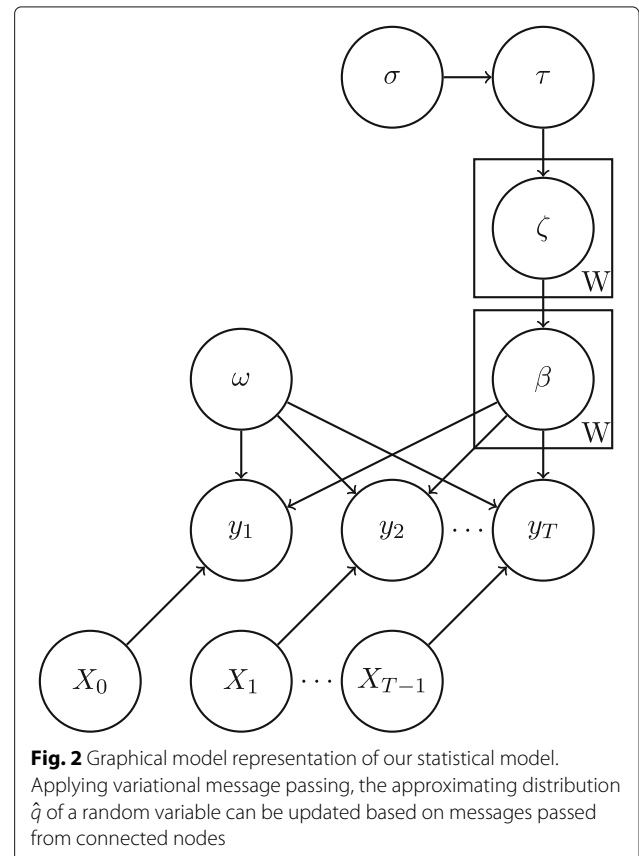
*Saccharomyces cerevisiae* gene regulatory network, and then simulating the dynamics of the networks under our DBN model. Subnetworks of 25 and 50 nodes were generated and used to simulate 20 time points with 3 replicates.

Synthetic count data were generated by constructing a negative binomial DBN model as in Eq. 2 corresponding to the generated subnetworks with randomised parameters $\beta$ sampled from a mixture of equally weighted $\mathcal{N}(0.3, 0.1)$ and $\mathcal{N}(-0.3, 0.1)$ distributions. The initial conditions and mean and dispersion parameters were randomly sampled from the empirically estimated means and dispersions of each gene from a publicly available RNA-seq count data set from the recount2 database [35] (accession *ERP003613*) consisting of 171 samples from 25 tissues in humans [36]. This was done so as to simulate the observed distributions of RNA-Seq counts in a real world data set.

We compare our approach against the Lasso as implemented in the `lars` R package [37], and the Gaussian regularised regression method in the `glmnet` R package [38]. For these methods the count data was first normalised, either by transforming the counts by the empirical cumulative distribution function of the data and subsequently mapping these to the quantiles of a $\mathcal{N}(0, 1)$ distribution, or by applying the `rlog` function of the `DESeq2` R package [13] to normalise the counts. We also applied the regularised Poisson regression method implemented of the `glmnet` R package to the count data, and the `mpath` R package [22] that performs penalised negative binomial regression. Finally we also applied a multinomial regularised regression from the `glmnet` R package to discretised data that were binned into 4 distinct levels by quantiles, to give a discrete DBN model. The degree of regularisation was in each case selected using cross validation as implemented in the respective software packages.

In Figs. 3 and 4 we show the partial area under the receiver operating curve (AUC-ROC) with a cutoff of 0.95 and corrected to fit the range 0 to 1, and area under the precision recall curve (AUC-PR), as calculated by the `PRROC` R package [39], and Matthews Correlation Coefficient (MCC), for the various methods to be benchmarked. For the MCC, edges were predicted as those where zero
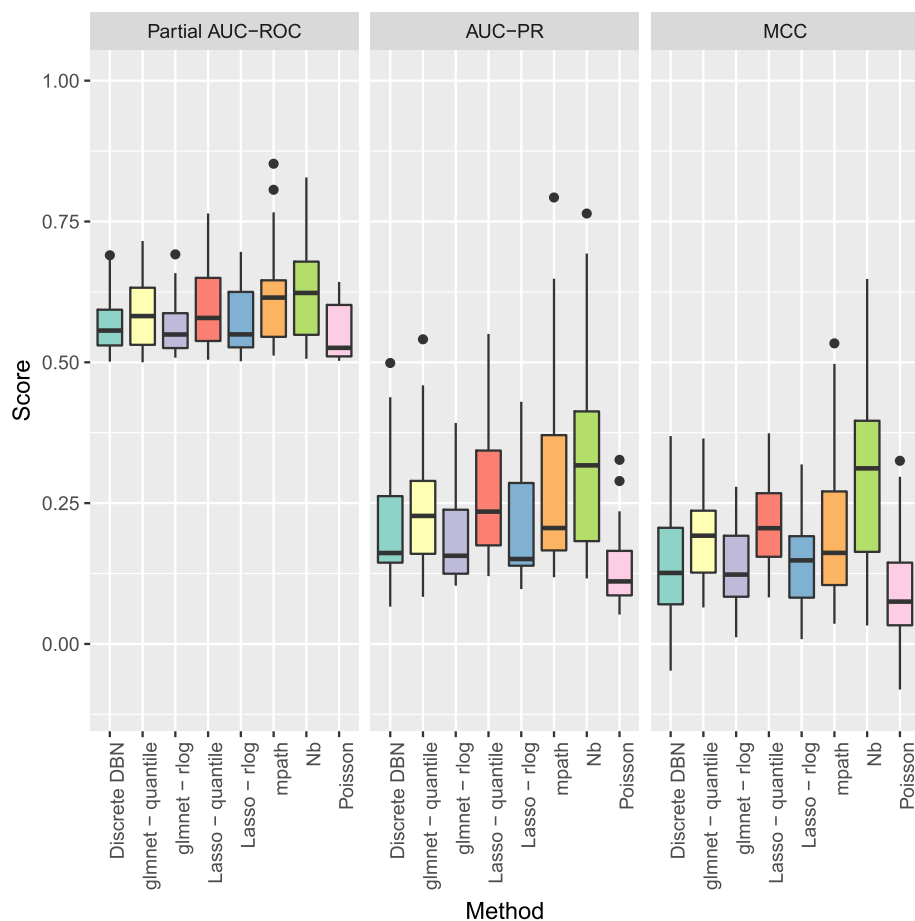


**Fig. 3** Boxplots of partial AUC-ROC, AUC-PR, and MCC for our method (Nb) and the competing methods benchmarked when learning directed networks of 25 nodes from synthetic data, for 5 subnetworks sampled from the *S. cerevisiae* gene regulatory network
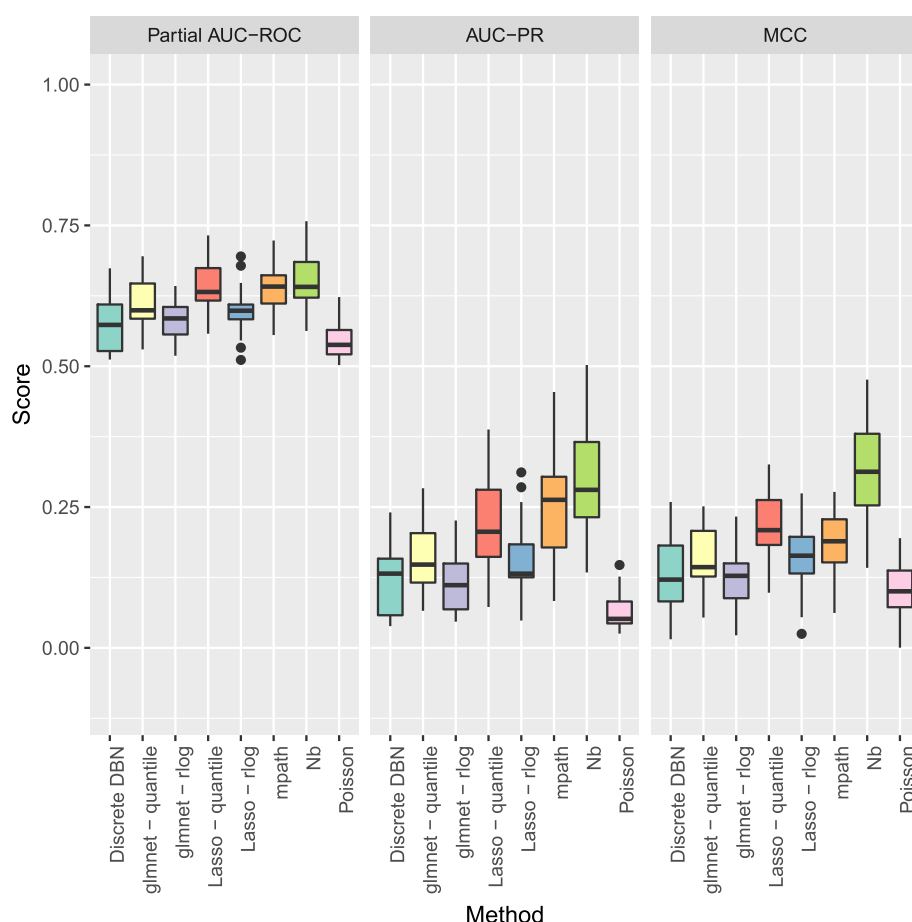
**Fig. 4** Boxplots of partial AUC-ROC, AUC-PR, and MCC for our method (Nb) and the methods benchmarked when learning directed networks of 50 nodes from synthetic data, for 5 subnetworks sampled from the *S. cerevisiae* gene regulatory network

was not contained in the 95% credible interval of the corresponding regression coefficients, and for the Lasso and `glmnet` methods, non-zero coefficients were taken as predicted edges. As the count data were generated by a stochastic model, we repeated benchmarking on each network 5 times with resampled negative binomial means and dispersions and simulated count data. Running time for our algorithm was under 10 minutes for the 50 node networks considered.

For networks of 25 nodes in Fig. 3, our method shows an improved performance over the competing methods in terms of the AUC-PR, and also in terms of the MCC. Although the distinction between the approaches is less marked for the AUC-ROC, this is to be expected as the simulated biological network structures have far fewer (< 10%) true positives than true negatives, a situation in which the AUC-ROC does not distinguish performance as well as AUC-PR [39, 40].

As can be seen in Fig. 4 performance for larger networks of 50 nodes is also improved over competing methods in terms of AUC-PR and MCC. For the competing methods,

quantile normalisation for the Lasso and `glmnet` appear to outperform normalisation using the rlog function of `DESeq2`. As the only other method applying a negative binomial distribution, `mpath` is the closest method to our approach, but it appears that the application of the horseshoe shrinkage prior delivers improved performance. It is clear that, as might be expected, taking into account the distributional properties observed in RNA-Seq data improves on the performance of methods based on assumptions that do not hold for RNA-Seq count data.

## Neural progenitor cell differentiation

To illustrate an application of our model to a real world RNA-Seq data set, we consider a publicly available RNA-Seq time course data set available from the recount2 database [35], accession *SRP041159*. The data consist of RNA-Seq counts from neuronal stem cells for 3 replicates over 7 time points after the induction of neuronal differentiation [41]. To select a subset of genes to analyse we performed a differential expression test between time points

using the DESeq2 R package [13], and selected the 25 genes with the largest median fold-change between time points that were also differentially expressed between all time points.

Applying our method and selecting edges with a posterior probability > 0.95 produced the network shown in Fig. 5, where it can be seen that there are four genes (MCUR1, PARP12, COL17A1, CDON) acting as hubs, suggesting these genes may be important in neuronal differentiation. Within the network MCUR1 appears to influence the transcription of a large number of genes with many outgoing edges, whilst PARTP12, COL17A1 and CDON have both incoming and outgoing edges. This may suggest a more fundamental role of MCUR1 in controlling neuronal differentiation.

For each node we also calculate the betweenness centrality, which is the fraction of the total number of shortest paths between nodes in the network that pass through a given node. This gives a measure of the importance of a node in the network, as nodes with a larger betweenness centrality will disrupt more paths within the network if deleted, and act as bottlenecks that connect modules within the network. Looking at the betweenness centrality of each node it appears that PARP12 and CDON, and to a lesser extent COL17A1, are important carriers of information within the network. Of these genes playing a central role in the network, CDON has been shown to

be promote neuronal differentiation through the activation of p38MAPK pathway [42, 43] and inhibition of Wnt signalling [44], whilst MCUR1 is known to bind to MCU [45], that in turn has been shown to influence neuronal differentiation [46].

## Discussion and conclusions

We have developed a fast and robust methodology for the inference of gene regulatory networks from RNA-Seq data that specifically models the observed count data as being negative binomially distributed. Our approach outperforms other sparse regression based methods in learning directed networks from time series data.

Another approach to network inference from RNA-Seq data could be to further develop mutual information based methodologies with this specific problem in mind. Mutual information based methods have the benefit of being independent of any specific model of the distribution of the data, and so could help sidestep issues in parametric modelling of RNA-Seq data. However this comes at the cost of abandoning the simplifying assumptions that are made by applying a parametric model that provides a reasonable fit to the data, and presents challenges of its own in the reliable estimation of the mutual information between random variables.
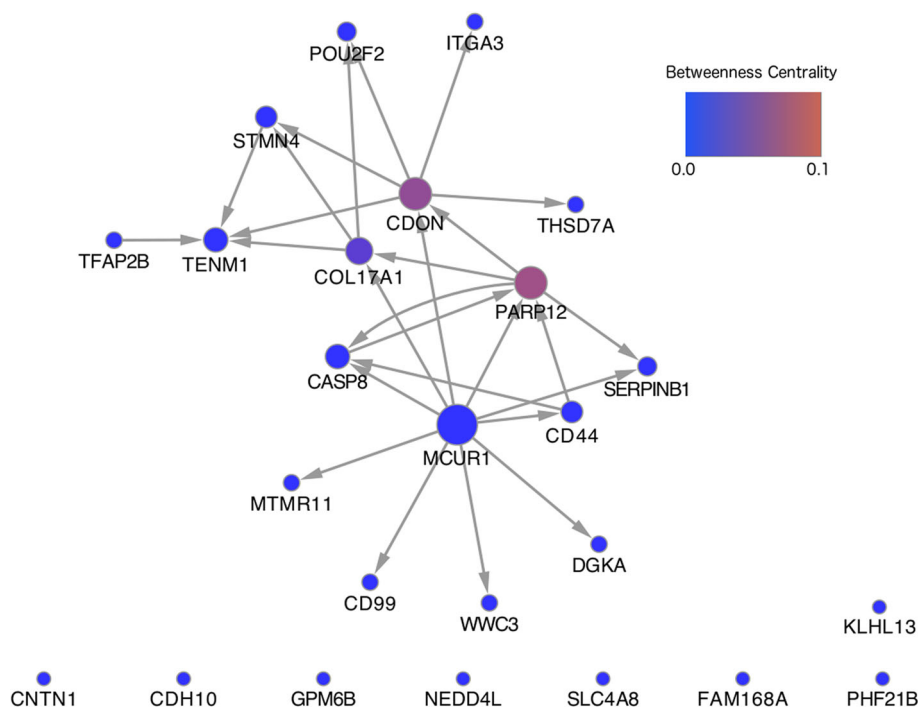


**Fig. 5** DBN inferred from the human neuronal differentiation time series data set. Edges were selected using a posterior probability cut-off of 0.95

## Appendix 1: Variational inference

From the results in [31] the model can be written as a Poisson-Gamma mixture, so that

$$p(y_t|\lambda_t) \sim \text{Pois}(\lambda_t) \qquad (16)$$

$$p(\lambda_t|x_t, \beta, \omega) \sim \text{Gamma}(\omega, \omega \exp[-X\beta]) \qquad (17)$$

and the horseshoe prior on $\beta$ represented using a mixture of inverse gamma distributions,

$$p(\beta_w|\zeta_w^2) \sim \mathcal{N}(0, \zeta_w^2) \qquad (18)$$

$$p(\zeta_w^2|a_w) \sim \text{InvGamma}\left(\frac{1}{2}, \frac{1}{a_w}\right) \qquad (19)$$

$$p(a_w|\tau^2) \sim \text{InvGamma}\left(\frac{1}{2}, \frac{1}{\tau^2}\right) \qquad (20)$$

$$p(\tau^2|b) \sim \text{InvGamma}\left(\frac{1}{2}, \frac{1}{b}\right) \qquad (21)$$

$$p(b|\sigma^2) \sim \text{InvGamma}\left(\frac{1}{2}, \frac{1}{\sigma^2}\right). \qquad (22)$$

### Mean field approximation

The mean field approximation of the posterior is then

$$\prod_i p(y_i|\lambda_i)p(\lambda_i|X_i, \beta, \omega)p(\omega) \prod_w p(\beta_w|\zeta_w^2) p(\zeta_w^2) p(\zeta_w^2|\tau)$$

$$p(\tau|\sigma^2) p(\sigma^2)$$

$$\approx \prod_i [q(\lambda_i)] q(\beta)q(\omega) \prod_w [q(\zeta_w^2)q(a_w)] q(\tau^2) q(b). \qquad (23)$$

The variational updates for $\lambda_t$ can be derived as

$$\log \hat{q}(\lambda_t) = \mathbb{E}_q\left[\log p(y_t|\lambda_t)p(\lambda_t|X_t, \beta, \omega)\right] + \text{const.}$$

$$= \mathbb{E}_q\left[\log \frac{\lambda_t^{y_t} e^{-\lambda_t}}{y_t!} \frac{(\omega \exp(-X_t\beta))^\omega \lambda_t^{\omega-1} e^{-\lambda_t \omega \exp(-X_t\beta)}}{\Gamma(\omega)}\right] + \text{const.}$$

$$= \mathbb{E}_q\left[y_t \log \lambda_t - \lambda_t + (\omega-1) \log \lambda_t - \lambda_t \omega \exp(-X_t\beta)\right] + \text{const.} \qquad (24)$$

$$\hat{q}(\lambda_t) \sim \text{Gamma}\left(y_t + \mathbb{E}_q[\omega], 1 + \mathbb{E}_q[\omega] \mathbb{E}_q[\exp(-X_t\beta)]\right) \qquad (25)$$

and due to the properties of the log-normal distribution

$$\mathbb{E}_q\left[\exp(-X_t\beta)\right] = \exp\left(-X_t\mathbb{E}[\beta] + \frac{1}{2}X_t\Sigma X_t^T\right), \qquad (26)$$

where $\Sigma$ is the covariance matrix of $\beta$ under $\hat{q}$.

As derived in [31], applying non-conjugate variational message passing, $\hat{q}(\beta) \sim \mathcal{N}(\mu, \Sigma)$ and the variational update for $\beta$ is

$$w = \exp\left(-X\mu + \frac{1}{2}\text{diagonal}\left(X\Sigma X^T\right)\right) \qquad (27)$$

$$\Sigma = \left[\omega X^T \text{diag}(\mathbb{E}[\lambda] \cdot w)X + M\right]^{-1} \qquad (28)$$

$$M = \text{diag}\left(\mathbb{E}\left[\frac{1}{\sigma_w^2}\right]\right) \qquad (29)$$

$$\mu = \mu + \Sigma\left[\omega X^T(\mathbb{E}[\lambda] \cdot w - 1) - M\mu\right], \qquad (30)$$

and for the dispersion $\omega$ we apply numerical integration as described in [31].

Then for the horseshoe prior on $\beta$, the variational updates are

$$\log \hat{q}(\zeta_w^2) = \mathbb{E}_q\left[\log p(\beta_w|\zeta_w^2) p(\zeta_w^2)\right] + \text{const.}$$

$$= \mathbb{E}_q\left[-\frac{1}{2} \log \zeta_w^2 - \frac{\beta_w^2}{2\zeta_w^2} + (-\alpha - 1) \log \zeta_w^2 - \frac{\gamma}{\zeta_w^2}\right] + \text{const.} \qquad (31)$$

$$\hat{q}(\zeta_w^2) \sim \text{InvGamma}\left(1, \frac{1}{2}\mathbb{E}[\beta_w^2] + \mathbb{E}_q[a_w]\right) \qquad (32)$$

$$\log \hat{q}(a_w) = \mathbb{E}_q\left[\log p(\zeta_w^2|a_w) p(a_w|\tau^2)\right] + \text{const.}$$

$$= \mathbb{E}_q\left[-\frac{1}{a_w\zeta_w^2} - \frac{1}{2} \log a_w - \frac{3}{2} \log a_w - \frac{1}{\tau^2 a_w}\right] + \text{const.} \qquad (33)$$

$$\hat{q}(a_w) \sim \text{InvGamma}\left(1, \mathbb{E}_q\left[\frac{1}{\zeta_w^2}\right] + \mathbb{E}_q\left[\frac{1}{\tau^2}\right]\right) \qquad (34)$$

$$\log \hat{q}(\tau^2) = \mathbb{E}_q\left[\sum_w \log p(a_w|\tau^2) + \log p(\tau^2|b)\right] + \text{const.}$$

$$= \mathbb{E}_q\left[-\sum_w \left(\frac{1}{2} \log \tau^2 + \frac{1}{a_w\tau^2}\right) - \frac{3}{2} \log \tau^2 - \frac{1}{b\tau^2}\right] + \text{const.} \qquad (35)$$

$$\hat{q}(\tau^2) \sim \text{InvGamma}\left(\frac{1}{2} + \frac{W}{2}, \mathbb{E}_q\left[\frac{1}{b}\right] + \sum_w \mathbb{E}_q\left[\frac{1}{a_w}\right]\right) \qquad (36)$$

$$\log \hat{q}(b) = \mathbb{E}_q\left[\log p(\tau^2|b) p(b|\sigma^2)\right] + \text{const.}$$

$$= \mathbb{E}_q\left[-\frac{1}{2} \log b - \frac{1}{\tau^2 b} - \frac{3}{2} \log b - \frac{1}{\sigma^2 b}\right] + \text{const.} \qquad (37)$$

$$\hat{q}(b) \sim \text{InvGamma}\left(1, \mathbb{E}_q\left[\frac{1}{\tau^2}\right] + \mathbb{E}_q\left[\frac{1}{\sigma^2}\right]\right) \qquad (38)$$

$$\log \hat{q}(\sigma^2) = \mathbb{E}_q\left[\log p(b|\sigma^2) p(\sigma^2)\right] + \text{const.}$$

$$= \mathbb{E}_q\left[-\frac{1}{2} \log \sigma^2 - \frac{1}{b\sigma^2} - \log \sigma^2\right] + \text{const.} \qquad (39)$$

$$\hat{q}(\sigma^2) \sim \text{InvGamma}\left(\frac{1}{2}, \mathbb{E}_q\left[\frac{1}{b}\right]\right). \qquad (40)$$
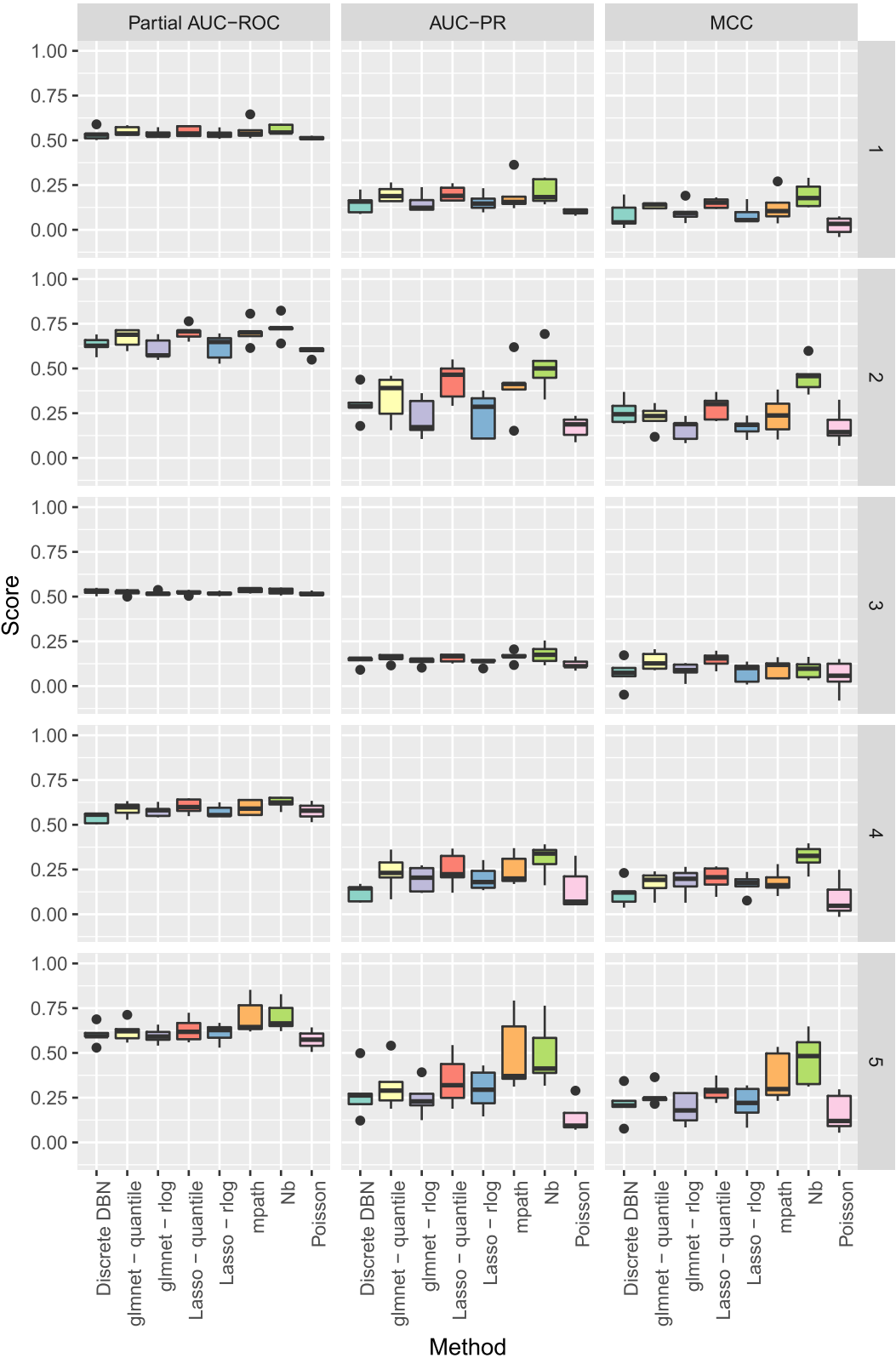
## Appendix 2: Supplemental Figures

**Fig. 6** Metrics calculated for networks of 25 nodes separated by individual network structure for the 5 different networks considered. Each bar plot corresponds to 5 simulated data sets from a single network structure
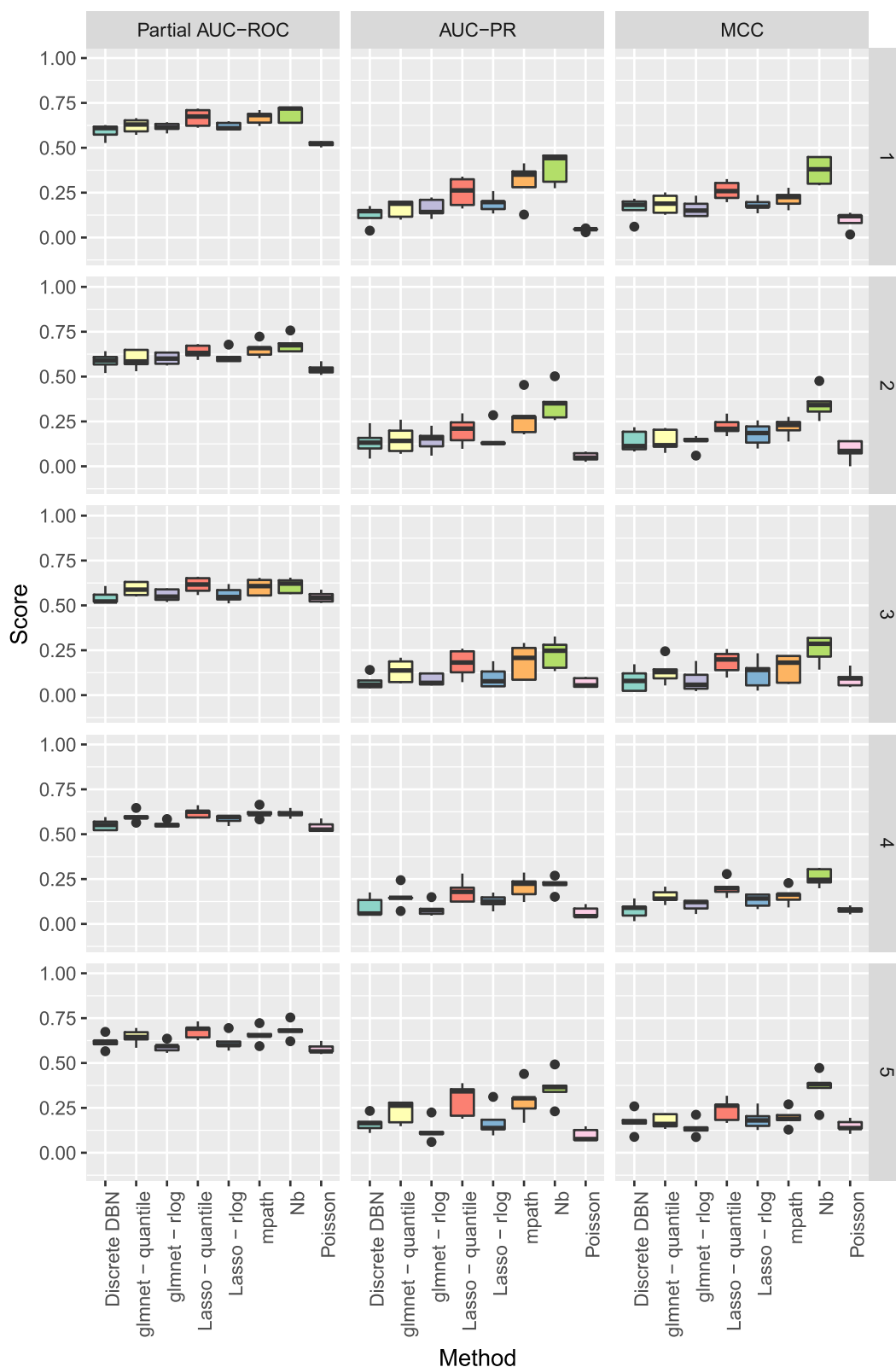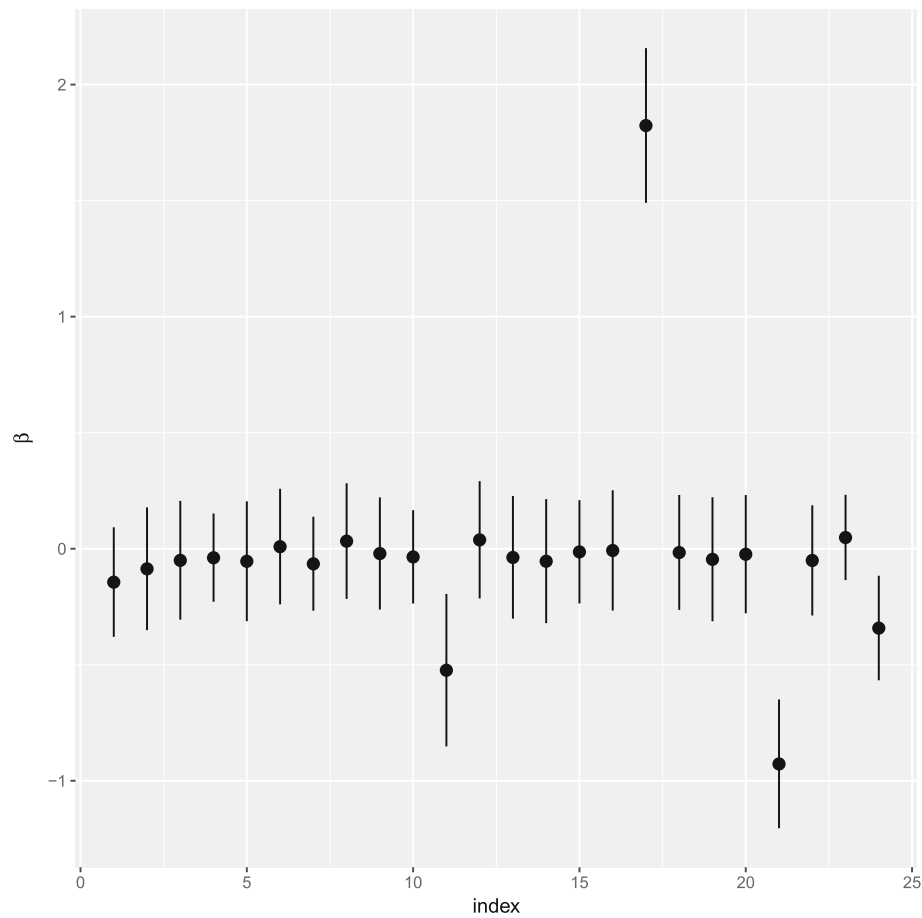
**Fig. 7** Metrics calculated for networks of 50 nodes separated by individual network structure for the 5 different networks considered. Each bar plot corresponds to 5 simulated data sets from a single network structure

**Fig. 8** Posterior means and standard deviations for the regression coefficients $\beta$ for a single node when applied to the NPC data considered in "Neural progenitor cell differentiation" section

## Publisher's Note

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

## References

1. Werhli AV,  Grzegorczyk M,  Husmeier D. Comparative evaluation of reverse engineering gene regulatory networks with relevance networks, graphical Gaussian models and Bayesian networks. Bioinformatics. 2006;22(20):2523–31.
2. Husmeier D,  Werhli AV. Bayesian integration of biological prior knowledge into the reconstruction of gene regulatory networks with Bayesian networks. Comput Syst Bioinforma Life Sci Soc Comput Syst Bioinforma Conf. 2007;6:85–95.
3. Opgen-Rhein R,  Strimmer K. From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data. BMC Syst Biol. 2007;1(1):37.
4. Lèbre S. Inferring dynamic Bayesian network with low order independencies. Stat Appl Genet Mole Biol. 2009;8(1):1–38.
5. Lèbre S,  Becq J,  Devaux F,  Stumpf MP,  Lelandais G. Statistical inference of the time-varying structure of gene-regulation networks. BMC Syst Biol. 2010;4(1):130.
6. Grzegorczyk M,  Husmeier D. Improvements in the reconstruction of time-varying gene regulatory networks: dynamic programming and regularization by information sharing among genes. Bioinformatics. 2011;27(5):693–9.
7. Thorne T,  Stumpf MPH. Inference of temporally varying Bayesian networks. Bioinformatics. 2012;28(24):3298–305.
8. Thorne T,  Fratta P,  Hanna MG,  Cortese A,  Plagnol V,  Fisher EM,  Stumpf MP. Graphical modelling of molecular networks underlying sporadic inclusion body myositis. Mole BioSyst. 2013;9(7):1736–42.
9. Wang T,  Ren Z,  Ding Y,  Fang Z,  Sun Z,  MacDonald ML,  Sweet RA,  Wang J,  Chen W. FastGGM: An Efficient Algorithm for the Inference of Gaussian Graphical Model in Biological Networks. PLOS Comput Biol. 2016;12(2):e1004755.
10. Hardcastle TJ,  Kelly KA. baySeq: Empirical Bayesian methods for identifying differential expression in sequence count data. BMC Bioinforma. 2010;11(1):422.
11. Anders S,  Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11(10):R106.
12. Robinson MD,  McCarthy DJ,  Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–40.
13. Love MI,  Huber W,  Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
14. Inouye DI,  Yang E,  Allen GI,  Ravikumar P. A review of multivariate distributions for count data derived from the Poisson distribution. Wiley Interdisc Rev Comput Stat. 2017;4(3):e1398.
15. Allen GI,  Liu Z. A Local Poisson Graphical Model for inferring networks from sequencing data. IEEE Transac NanoBiosci. 2013;12(3):189–98.
16. Gallopin M,  Rau A,  Jaffrézic F. A hierarchical poisson log-normal model for network inference from RNA sequencing data. PLOS ONE. 2013;8(10):e77503.
17. Äijö T,  Butty V,  Chen Z,  Salo V,  Tripathi S,  Burge CB,  Lahesmaa R,  Lähdesmäki H. Methods for time series analysis of RNA-seq data with application to human Th17 cell differentiation. Bioinformatics. 2014;30(12):i113–20.
18. Jo K,  Kwon H-B,  Kim S. Time-series RNA-seq analysis package (TRAP) and its application to the analysis of rice, Oryza sativa L. ssp. Japonica, upon drought stress. Methods. 2014;67(3):364–72.
19. Christopher DLW,  Penfold A. How to infer gene networks from expression profiles, revisited. Interface Focus. 2011;1(6):857–70.
20. Penfold CA,  Buchanan-Wollaston V,  Denby KJ,  Wild DL. Nonparametric Bayesian inference for perturbed and orthologous gene regulatory networks. Bioinformatics. 2012;28:i233–41.
21. Meyer PE,  Lafitte F,  Bontempi G. minet: A R/Bioconductor Package for Inferring Large Transcriptional Networks Using Mutual Information. BMC Bioinformatics. 2008;9(1):461.
22. Wang Z,  Ma S,  Zappitelli M,  Parikh C,  Wang C-Y,  Devarajan P. Penalized count data regression with application to hospital stay after pediatric cardiac surgery. Stat Methods Med Res. 2016;25(6):2685–703.
23. Carvalho CM,  Polson NG,  Scott JG. Handling Sparsity via the Horseshoe. AISTATS. Proc Mach Learn Res. 2009;5:73–80.
24. Carvalho CM,  Polson NG,  Scott JG. The horseshoe estimator for sparse signals. Biometrika. 2010;97(2):465–80.
25. Koller D,  Friedman N. Probabilistic Graphical Models. Cambridge: MIT Press; 2009.
26. MacKay DJC. Developments in Probabilistic Modelling with Neural Networks —Ensemble Learning. In: Machine Learning. London: Springer London; 1995.  p. 191–8.
27. MacKay DJC. Information Theory, Inference and Learning Algorithms. Cambridge: Cambridge University Press; 2003.
28. Bishop CM. Pattern Recognition and Machine Learning. New York: Springer Verlag; 2006.
29. Barber D. Bayesian Reasoning and Machine Learning. Cambridge: Cambridge University Press; 2012.
30. Murphy KP. Machine Learning A Probabilistic Perspective. Cambridge: MIT Press; 2012.
31. Luts J. Variational Inference for Count Response Semiparametric Regression. Bayesian Analysis. 2015;10(4):991–1023, Wand, MP.
32. Knowles DA,  Minka T. Non-conjugate Variational, Message Passing for Multinomial and Binary Regression. In: Proceedings of the 24th International Conference on Neural Information Processing Systems; 2011.  p. 1701–9.
33. Winn J,  Bishop CM. Variational Message Passing. J Mach Learn Res. 2005;6(Apr):661–94.
34. Schaffter T,  Marbach D,  Floreano D. GeneNetWeaver: in silico benchmark generation and performance profiling of network inference methods. Bioinformatics. 2011;27(16):2263–70.
35. Collado-Torres L,  Nellore A,  Kammers K,  Ellis SE,  Taub MA,  Hansen KD,  Jaffe AE,  Langmead B,  Leek JT. Reproducible RNA-seq analysis using recount2. Nature Biotechnology. 2017;35(4):319–21.
36. Fagerberg L,  Hallström BM,  Oksvold P,  Kampf C,  Djureinovic D,  Odeberg J,  Habuka M,  Tahmasebpoor S,  Danielsson A,  Edlund K,  Asplund A,  Sjöstedt E,  Lundberg E,  Szigyarto CA-K,  Skogs M,  Takanen JO,  Berling H,  Tegel H,  Mulder J,  Nilsson P,  Schwenk JM,  Lindskog C,  Danielsson F,  Mardinoglu A,  Sivertsson Å,  von Feilitzen K,  Forsberg M,  Zwahlen M,  Olsson I,  Navani S,  Huss M,  Nielsen J,  Pontén F,  Uhlén M. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. Molecular &, Cellular Proteomics. 2014;13(2):397–406.
37. Hastie T,  Efron B. lars: Least Angle Regression, Lasso and Forward Stagewise; 2013. URL https://CRAN.R-project.org/package=lars. R package version 1.2.
38. Friedman J,  Hastie T,  Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw. 2010;33(1):1–22.

39. Grau J, Grosse I, Keilwagen J. PRROC: computing and visualizing precision-recall and receiver operating characteristic curves in R. Bioinformatics. 2015;31(15):2595–7.

40. Davis J, Goadrich M. The relationship between Precision-Recall and ROC curves. New York: ACM; 2006.

41. Sauvageau M, Goff LA, Lodato S, Bonev B, Groff AF, Gerhardinger C, Sanchez-Gomez DB, Hacisuleyman E, Li E, Spence M, Liapis SC, Mallard W, Morse M, MR Swerdel, Ecclessis MFD, Moore JC, Lai V, Gong G, Yancopoulos GD, Frendewey D, Kellis M, Hart RP, Valenzuela DM, Arlotta P, Rinn JL. Multiple knockout mouse models reveal lincRNAs are required for life and brain development. eLife. 2013;2:360.

42. Zhang W, Yi M-J, Chen X, Cole F, Krauss RS, Kang J-S. Cortical thinning and hydrocephalus in mice lacking the immunoglobulin superfamily member CDO. Mole Cell Biol. 2006;26(10):3764–72.

43. Oh J-E, Bae G-U, Yang Y-J, Yi M-J, Lee H-J, Kim B-G, Krauss RS, Kang J-S. Cdo promotes neuronal differentiation via activation of the p38 mitogen-activated protein kinase pathway. FASEB J. 2009;23(7):2088–99.

44. Jeong M-H, Ho S-M, Vuong TA, Jo S-B, Liu G, Aaronson SA, Leem Y-E, Kang J-S. Cdo suppresses canonical Wnt signalling via interaction with Lrp6 thereby promoting neuronal differentiation. Nature Communications. 2014;5:5:w455.

45. Mallilankaraman K, Cárdenas C, Doonan PJ, Chandramoorthy HC, Irrinki KM, Golenár T, Csordás G, Madireddi P, Yang J, Müller M, Miller R, Kolesar JE, Molgó J, Kaufman B, Hajnóczky G, Foskett JK, Madesh M. MCUR1 is an essential component of mitochondrial Ca2+ uptake that regulates cellular metabolism. Nature Cell Biology. 2012;14(12):1336–43.

46. Rharass T, Lemcke H, Lantow M, Kuznetsov SA, Weiss DG, Panáková D. Ca2+-mediated mitochondrial reactive oxygen species metabolism augments Wnt/$\beta$-catenin pathway activation to facilitate cell differentiation. J Biol Chem. 2014;289(40):7–27951.