BMC Bioinformatics

METHODOLOGY ARTICLE   **Open Access**

Time-lagged Ordered Lasso for network inference

Phan Nguyen[1] and Rosemary Braun[1,2]*

Abstract

Background: Accurate gene regulatory networks can be used to explain the emergence of different phenotypes, disease mechanisms, and other biological functions. Many methods have been proposed to infer networks from gene expression data but have been hampered by problems such as low sample size, inaccurate constraints, and incomplete characterizations of regulatory dynamics. Since expression regulation is dynamic, time-course data can be used to infer causality, but these datasets tend to be short or sparsely sampled. In addition, temporal methods typically assume that the expression of a gene at a time point depends on the expression of other genes at only the immediately preceding time point, while other methods include additional time points without any constraints to account for their temporal distance. These limitations can contribute to inaccurate networks with many missing and anomalous links.

Results: We adapted the time-lagged Ordered Lasso, a regularized regression method with temporal monotonicity constraints, for de novo reconstruction. We also developed a semi-supervised method that embeds prior network information into the Ordered Lasso to discover novel regulatory dependencies in existing pathways. R code is available at https://github.com/pn51/laggedOrderedLassoNetwork.

Conclusions: We evaluated these approaches on simulated data for a repressilator, time-course data from past DREAM challenges, and a HeLa cell cycle dataset to show that they can produce accurate networks subject to the dynamics and assumptions of the time-lagged Ordered Lasso regression.

Keywords: Gene network reconstruction, Network inference, Gene regulation, Lasso, Regularization, Penalized regression, Time course data

Background

A major challenge in systems biology is understanding the structure and function of the molecular interaction networks that regulate cellular processes. Gene regulatory networks (GRNs) are abstractions of these networks [1] in which nodes correspond to genes and edges to interactions, providing a high-level overview of the topology of gene-gene interactions and their purposes. A comprehensive GRN can improve our understanding of its role in the emergence of different phenotypes, disease mechanisms, and other biological processes and how it may be perturbed for therapeutic purposes [2–5]. Despite burgeoning research, constructing accurate GRN models

remains a challenge. Because of the large number of genes in a genome, experimental validation of every possible interaction is an arduous task. Therefore, computational methods are preferred to screen for probable dependencies based on high-throughput expression measurements. Elucidating edges from these datasets with GRN reconstruction methods can involve a combination of ad hoc heuristics and interaction criteria as well as imposing modeling assumptions on the expression dynamics of a GRN and inferring models that preserve those assumptions [6, 7]. Additional insights into these interactions may be obtained by ascertaining the quantitative models that describe the dynamics of these interactions [8–12]. Rather than predicting edges, these methods attempt to estimate the parameters that describe the stochastic kinetics of the chemical reactions that underlie the connections between genes to provide detailed models that govern the observed expression dynamics. In both cases, accurate methods can

*Correspondence: rbraun@northwestern.edu
[1]Department of Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, IL, USA
[2]Biostatistics Division, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA

offer new experimental directions to verify novel interactions and identify deficiencies in currently known GRNs and models.

However, computational approaches for GRN reconstruction pose another set of challenges. Since every ordered pair of genes presents the possibility of an edge, an exponentially large space of GRNs needs to be considered. Furthermore, while high-throughput sequencing technologies have advanced significantly and can simultaneously measure the expression levels of thousands of genes in an efficient and affordable manner, dataset sample sizes still tend to be very small compared to the number of genes. This disparity results in clusters with many genes that have similar expression profiles, allowing many GRNs to plausibly account for the observed patterns of expression in a dataset. In addition to GRN reconstruction being an underdetermined problem, other issues such as missing data, gene expression stochasticity, confounding, and incomplete characterizations of the gene regulatory dynamics can also adversely affect GRN predictions. While the wealth of gene expression data has been a boon to understanding GRNs, there is still a demand for accurate and interpretable GRN inference methods that properly address these problems with promising modeling assumptions and efficient algorithms.

Most GRN reconstruction methods can be broadly classified into two categories. De novo approaches attempt to infer GRNs solely from expression data. Specifically, edges between genes are inferred by deriving edge confidence scores based on similarities between expression profiles [13–15], statistical measures of causality [16], or estimations of the strength of influence between genes based on an assumed model for gene expression, including regression-based methods that model the expression of a gene as a linear function of its regulators [17, 18], probabilistic graphical models that estimate the conditional dependence between genes [19], Boolean networks that discretize the expression data into binary states that are used to learn Boolean functions and their associated networks [20–24], and random forests that can learn nonlinear dependencies using ensembles of decision trees [25]. Approaches to filter out false positive edges arising from confounding or indirect interactions have also been proposed [26, 27]. The other major approaches are semi-supervised methods, which incorporate information about a network. For example, experimentally derived evidence for regulatory dependencies between genes has been compiled in databases such as KEGG [28, 29] and REACTOME [30]. While these descriptions are incomplete, they can be used to refine partially known GRNs with additional evidence from transcriptomic data. Unlike de novo approaches, semi-supervised methods attempt to refine GRNs by leveraging knowledge of a partially known GRN with an expression dataset in order

to identify concordances and discrepancies between an expression model on the GRN and an observed expression dataset. One common approach to developing these methods has been to modify a de novo algorithm to bias the selection of known edges, which include methods that extend regression-based approaches [31, 32], random forest-based approaches [32], and Boolean network-based approaches [33, 34]. Despite fewer developments, semi-supervised approaches have the potential to reduce false positive predictions and improve GRN reconstructions.

In both cases, most methods rely on static expression data. Alternatively, since expression regulation is a dynamic process, time-course data can be used to infer causality. However, temporal data tends to exhibit high autocorrelation and is usually only gathered for a few time points and subjects. In addition, many temporal methods typically assume that the expression of a gene at a time point depends on its regulators at only the immediately preceding time point, while other methods include additional time points but do not impose any constraints to account for their temporal distance. For instance, pairwise Granger causality [16, 35] tests the predictive capability of the past values of a predictor in estimating the present values of a target variable by comparing regression models with and without the predictor, but ignores the influence of other potential predictors and does not discriminate the effect of different lags. To account for multiple causes, Lasso-Granger [36] uses lasso regularization to identify causal predictors, but also neglects temporal distance when constructing linear models. These limitations can result in predicted GRNs with many missing and anomalous links.

In this paper, we first describe a de novo approach for GRN reconstruction based on the Ordered Lasso [37], a recently published regularization method that uses monotonicity constraints on the coefficients of a linear model to reflect the relative importance of the model features and has natural applications to time-lagged regression. Since partial knowledge of the dependencies between genes is available, we also describe a semi-supervised method that embeds prior network information into the Ordered Lasso to facilitate the discovery of novel edges in existing pathways. These methods establish several novel contributions and results. Notably, our methods are the first to consider a time-ordered constraint on regulatory influence for GRN inference. In addition, we can accommodate prior knowledge of regulatory interactions to infer novel and anomalous edges in a semi-supervised manner. The performance of our methods can also be shown to increase monotonically with the maximum lag of an expression model, thus obviating the need to find the optimal lag parameter. Furthermore, our methods have a demonstrated ability to make novel inferences that are later validated by experiment.

The organization of the rest of this paper is as follows. In the "Methods" section, we briefly review the time-lagged Ordered Lasso and describe suitable assumptions for dynamic gene expression on a GRN. In particular, we assume that each gene linearly depends on the lagged expression of its regulators at multiple preceding time points and enforce a monotonicity constraint on the lagged variables so that the regulatory influence of a lagged variable on the gene decreases as the lag of the variable increases. We then describe the adaptations of the time-lagged Ordered Lasso for de novo and semi-supervised GRN reconstruction. In the "Results" section, we apply the methods to simulated data for a repressilator, time-course datasets from past DREAM challenges, and a HeLa cell cycle dataset that has been used for benchmarking. We show that the de novo algorithm can derive accurate GRNs that reflect the dynamics and assumptions of the time-lagged Ordered Lasso while obviating the need for heuristics that optimize the maximum lag of dependence. We also show that by embedding a partially known GRN into the dynamics of the time-lagged Ordered Lasso, the semi-supervised method can accurately predict novel edges that account for the discrepancies between the prior knowledge of the regulatory connections and the observed dynamics of a gene expression dataset. In the "Discussion" section, we conclude and discuss possible extensions.

Methods

Time-lagged Ordered Lasso

The main difficulties in fitting models for gene expression are the high dimensionality and small sample size of an expression dataset. Due to the large number of genes relative to the number of samples, fitting even simple one-lag models in which the expression of a gene depends on the expression of other genes at a previous time point may be an underdetermined problem wherein many models plausibly fit the dataset and result in overfitting or difficulties with model selection and interpretation. Higher-order lagged models in which the dependence extends to multiple preceding time points provide more flexibility by accounting for long-range and multiple-lag dependencies, but the additional variables that are introduced further compound the problems encountered in the one-lag model. Furthermore, the lagged variables of a predictor tend to have high autocorrelation, especially when the temporal resolution of the data is small. Therefore, additional reasonable modeling assumptions must be imposed to ensure that accurate, interpretable models can still be feasibly learned.

To this end, one useful approach to prevent overfitting, improve model interpretability, and produce accurate predictions is the lasso or ℓ_1-regularized regression [38]. The lasso performs regularization and produces sparse solutions by minimizing the mean squared error of a regression model while also penalizing the sum of the absolute value of the model coefficients. By imposing constraints on the size of the coefficients, the lasso forces many of the coefficients to zero, leaving a few non-zero coefficients whose corresponding variables may be deemed relevant to predicting the output variable. Consequently, the lasso may be used for variable selection and to reduce overfitting.

In certain regression problems, an order constraint may be imposed to reflect the relative importance of the features. Recently, the Ordered Lasso was introduced to solve ℓ_1-regularized linear regression problems with monotonicity constraints on the coefficients [37], with a primary application to time-lagged regression. Specifically, a time-lagged order assumption may be imposed wherein recent data is assumed to be more predictive of the future than older data is; as the lag of a predictor increases, its influence decreases. To reflect this attenuation, the magnitude of a coefficient can be forced to monotonically decay with increasing temporal distance from a response variable. Additional algorithmic details about the Ordered Lasso and time-lagged Ordered Lasso may be found in [37].

Like the ordinary lasso, the time-lagged Ordered Lasso can facilitate feature selection and model interpretability. Since the ℓ_1 penalty forces many of the coefficients to zero, a lagged variable may be considered relevant if it has a non-zero coefficient. In addition, because of the monotonicity constraint on the lagged features of a predictor, all of the coefficients may be equal to zero beyond a certain lag. Therefore, the time-lagged Ordered Lasso can also provide insight into the maximum effective lag or range of influence of each predictor on the response.

De novo reconstruction

To adapt the time-lagged Ordered Lasso for de novo GRN reconstruction, we impose several assumptions on the dynamic model for the expression of a gene. We first assume that the expression of each gene is linearly dependent on the expression of its regulators at multiple preceding time points, a common assumption in many reconstruction methods for time series expression data. Furthermore, to reflect the importance of recent expression data, we assume that as the temporal distance between a target gene variable and a lagged variable of a predictor gene increases, the regulatory influence of the lagged variable on the target decreases, a justifiable assumption for many expression datasets. For example, since expression data tends to be sparsely sampled at distant time points, it is unreasonable to expect expression data at highly distant time points in the past to be strongly influential on the current expression level of a gene. For each gene i in a time-course expression dataset, we then

fit an expression model with maximum lag $l_{\max}$ allowed by the data and lasso regularization by solving the following problem using the time-lagged Ordered Lasso:

$$
\min_{\{w_{ji,k}\}} \quad \frac{1}{2} \sum_{t=1}^{T} \left(x_i(t) - \sum_{j=1}^{p} \sum_{k=1}^{l_{\max}} w_{ji,k} x_j(t-k\Delta t) \right)^2
$$

$$
+ \lambda \sum_{j=1}^{p} \sum_{k=1}^{l_{\max}} \left| w_{ji,k} \right|
$$

$$
\text{subject to} \quad \left| w_{ji,1} \right| \geq \left| w_{ji,2} \right| \geq \cdots \geq \left| w_{ji,l_{\max}} \right|,
$$

$$
\tag{1}
$$

where $x_i(t)$ is the expression of gene i at time t and the monotonicity constraint $\left| w_{ji,1} \right| \geq \cdots \geq \left| w_{ji,l_{\max}} \right|$ encodes the time-lagged order assumption of the expression model. We then predict an edge from gene j to gene i if any of the coefficients $w_{ji,1}, \ldots, w_{ji,l_{\max}}$ of the lagged variables of j are non-zero. Because of the monotonicity constraint, in effect, this only requires checking that the first lagged variable is non-zero. However, this does not imply that the higher-order lagged variables have no bearing on the edges that are predicted; the additional lagged variables of one gene may better explain a target gene's evolution in expression than the lagged variables of multiple other genes in a lower-lag model will, thereby eliminating the corresponding edges and potentially lowering the false positive rate in the higher-lag model.

Although a gene's expression may in reality depend non-linearly on its regulators, we use a simplified linear model for several reasons. First, having too many terms may be computationally restrictive; if p is the number of genes in the network, for each term we wish to consider, $pl_{\max}$ additional lagged variables need to be added to the model. In addition, the low sampling rates and time coverage of a dataset may be insufficient to accurately characterize these terms without overfitting or signal aliasing. Therefore, linearity serves as a simplifying assumption, deterrent to prevent overfitting, and preemptive measure to reduce computational overhead. We expect this approximation to be adequate for most applications, especially when detailed dynamics are difficult to observe due to the short time coverage and sparse sampling of a dataset.

To assess prediction accuracy across different values of λ, we test the method against known/synthetic networks and compute the area under the curve (AUC) of the receiver operating characteristic (ROC) curve as λ is varied. Since edges may potentially enter, leave, and re-enter a model as λ decreases, to ensure the ROC curve increases monotonically, we consider an edge to be predicted at a given value of λ if it enters at that value or larger. This can be viewed as applying a threshold on λ and merging the predicted networks for that value and

larger. Here, the AUC may be interpreted as the probability that a randomly chosen edge is ranked higher or enters a model earlier than a randomly chosen non-edge. Additional details on choosing λ may be found in Additional file 1: Section S-2.

Semi-supervised reconstruction

Since partial knowledge of the dependencies between genes is available, we also consider GRN refinements with semi-supervised adaptations. For most researchers, the primary interest in GRN reconstruction is discovering novel edges, or pairs of genes that are not previously known to interact, but their existence can be supported with evidence from transcriptomic data. On the other hand, prior information may also contain incorrect edges due to curatorial errors or differences between a canonical GRN forming the prior and that which exists in a particular phenotype. Thus, discovering both novel and anomalous connections is of interest.

We modify the de novo approach for semi-supervised reconstruction by embedding a prior GRN into the lasso as follows. Rather than use one general penalty parameter λ, we replace it with two parameters, λ_{edge} and $\lambda_{\text{non-edge}}$, to separately regularize the prior edge and non-edge coefficients, respectively. An expression model for gene k with maximum lag $l_{\max}$ is fit by solving the following problem with the time-lagged Ordered Lasso:

$$
\min_{\{w_{ji,k}\}} \quad \frac{1}{2} \sum_{t=1}^{T} \left(x_i(t) - \sum_{j=1}^{p} \sum_{k=1}^{l_{\max}} w_{ji,k} x_j(t-k\Delta t) \right)^2
$$

$$
+ \lambda_{\text{edge}} \sum_{j|(j,i)\in E} \sum_{k=1}^{l_{\max}} \left| w_{ji,k} \right|
$$

$$
+ \lambda_{\text{non-edge}} \sum_{j|(j,i)\notin E} \sum_{k=1}^{l_{\max}} \left| w_{ji,k} \right| t
$$

$$
\text{subject to} \quad \left| w_{ji,1} \right| \geq \left| w_{ji,2} \right| \geq \cdots \geq \left| w_{ji,l_{\max}} \right|,
$$

$$
\tag{2}
$$

where E denotes the set of edges in the prior GRN. If $\lambda_{\text{edge}} < \lambda_{\text{non-edge}}$, the magnitude of the coefficients of the prior edges will be penalized to a lesser extent than those of the prior non-edges, thereby allowing the former to account for most of the evolution in expression of the target gene. As a result, the prior edge coefficients are more likely to be non-zero, and the prior edges are more likely to be recovered as posterior edges.

Since the prior edges will not necessarily account for all of the output variance and the corresponding coefficients may still be fit with zero values (for $\lambda_{\text{edge}} > 0$), this approach allows us to predict novel and anomalous edges. For a gene j that is not known to regulate a target

gene i, we predict a novel edge if any of the coefficients $w_{ji,1}, \ldots, w_{ji,l_{\max}}$ are non-zero. On the other hand, if gene j is previously known to regulate gene i, we predict an anomalous edge if all of the coefficients $w_{ji,1}, \ldots, w_{ji,l_{\max}}$ are zero. As in the de novo case, both tests only require checking the first lagged variable because of the monotonicity constraint.

Related methods

We first compare our method to two baseline approaches. The first, Granger causality, is based on the notion that the utility of the information in one time series in forecasting another may be a potential indicator of causality [35]. A variable x is said to Granger-cause a variable y if the past values of x and y combined are more predictive of future values of y than just those of y alone are. For GRN reconstruction, since sample sizes tend to be much smaller than the number of genes, the most basic form of Granger causality is typically used [16]. Bivariate or pairwise Granger causality fits two autoregression models to predict y, one that includes the lagged values of x and the other without, and uses an F-test to assess the explanatory gain of using x in predicting y. A GRN is predicted by aggregating the F-test p-values across every ordered pair of genes and thresholding with false discovery rate correction procedures.

The second standard approach to which we compare our method is Lasso-Granger [36]. One of the major drawbacks of pairwise Granger causality is that it cannot be used with short time series; $l_{\max}$ must be sufficiently small relative to the number of sampled time points so that the models used to assess causality can be fit with non-zero residuals. In addition, since causality is only tested on a pairwise-basis, multiple regulators are not accounted for. Lastly, a dataset with n genes requires $O\left(n^2\right)$ tests for Granger causality, which may be computationally prohibitive when n is large. Lasso-Granger attempts to address these problems by solving Eq. 1, but without the monotonicity constraint.

We also compare our GRN predictions to those made by the truncating adaptive lasso [39], grouped graphical Granger modeling [40], and CNET [41]; details of these algorithms may be found in their respective publications. Like the time-lagged Ordered Lasso, Lasso-Granger, and Granger causality, the truncating adaptive lasso and grouped graphical Granger modeling assume that the expression of a gene linearly depends on the expression of its predictors at multiple preceding time points, but each of the methods applies different modeling constraints to fit the temporal model.

Datasets

A summary of the time-course datasets to which we apply the method is given in Table 1; we provide further details below.

Table 1 Evaluation datasets and information on the number of network genes (G), time points (TP), and time series (TS)

Dataset	Network	# G	# TP	# TS
Repressilator (sim)	Repressilator	3	5–2049	1
DREAM (sim)	d2c4-{1–2}	50	26	23
	d3c4-size-10-ecoli-{1–2}	10	21	4
	d3c4-size-10-yeast-{1–3}	10	21	4
	d3c4-size-100-ecoli-{1–2}	100	21	46
	d3c4-size-100-yeast-{1–3}	100	21	46
	d3c4-size-50-ecoli-{1–2}	50	21	23
	d3c4-size-50-yeast-{1–3}	50	21	23
	d4c2-size-10-network-{1–5}	10	21	5
	d4c2-size-100-network-{1–5}	100	21	10
HeLa (real)	Sambo et al. / BioGRID	9	47	1

The number of repressilator time points varies and depends on the time series length and sampling rate (see "Repressilator" section)
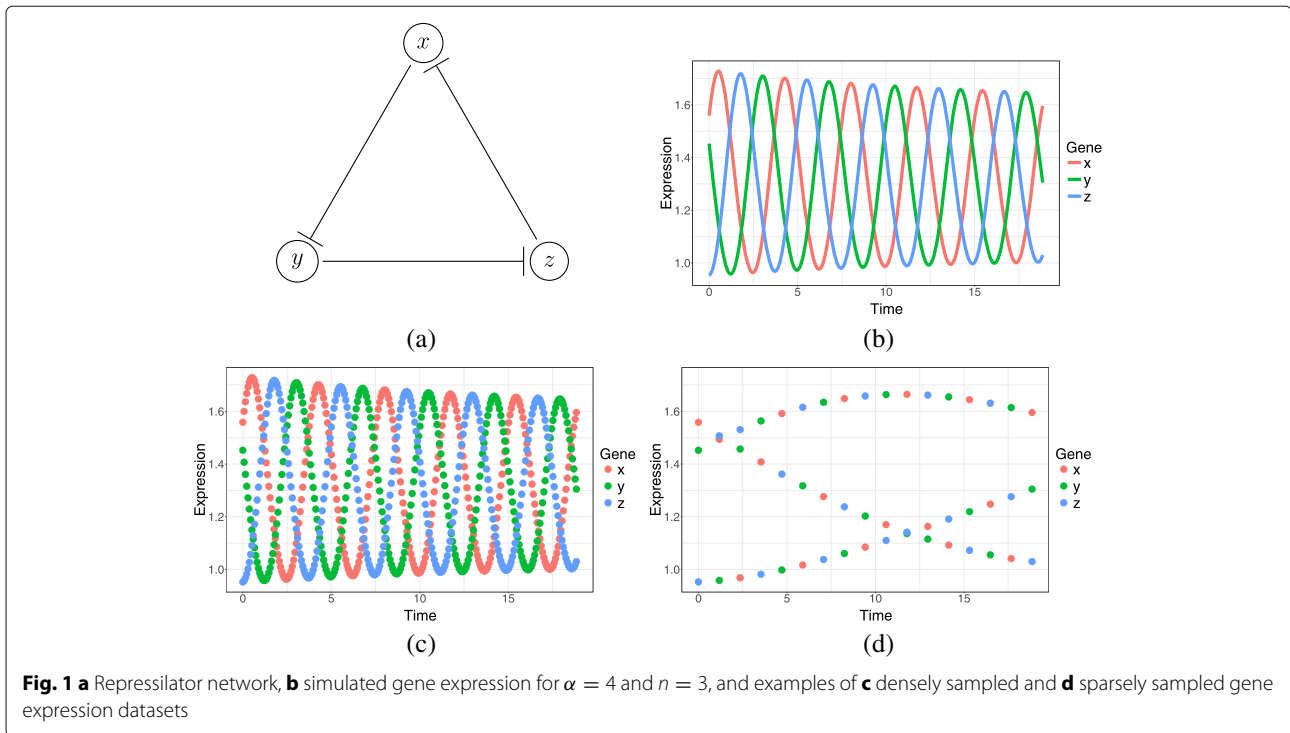
Repressilator

When designing experiments, experimentalists need to decide for how long and often data should be collected while factoring in technical complexity, cost, and other considerations. Since time series expression data tends to be short or sparse, predicting accurate GRNs from these datasets may be difficult. Therefore, we first analyze and demonstrate the effect of using different time series sampling rates and lengths on accuracy. To do so, we simulate data for a repressilator [42], a synthetic network of three genes connected in a feedback loop in which each gene represses the next to induce oscillatory patterns of expression. The behavior of a basic repressilator may be described using the coupled differential equations

$$\dot{x} = \frac{\alpha}{1 + z^n} - x, \ \dot{y} = \frac{\alpha}{1 + x^n} - y, \ \dot{z} = \frac{\alpha}{1 + y^n} - z. \ (3)$$

For our simulations, we set $\alpha = 4$ and $n = 3$; examples of simulated time series data are shown in Fig. 1.

DREAM

To analyze the utility of the methods in recovering known GRNs, we apply them to synthetic time-course data from several DREAM challenges. In one of the DREAM2 challenges, 50-node networks were derived from Erdos-Renyi and scale-free topologies with Hill-type kinetics driving gene expression [43]. The DREAM3 in silico network challenge contained 10-, 50-, and 100-gene subnetworks extracted from *E. coli* and *S. cerevisiae* gene networks, and expression values were simulated with ordinary differential equations and added measurement noise using GeneNetWeaver [44–47]. Finally, in the DREAM4 in silico network challenge, GeneNetWeaver was used to simulate data from stochastic differential equations by applying perturbations to 10- and 100-gene networks.

Fig. 1 a Repressilator network, **b** simulated gene expression for $\alpha = 4$ and $n = 3$, and examples of **c** densely sampled and **d** sparsely sampled gene expression datasets

HeLa cell cycle subnetwork

While synthetic expression data can be used elucidate the GRN inference properties of a method in a controlled manner, models for generating these datasets do not fully capture all of the nuances of real data and GRNs. To assess empirical practicality, we consider applications to the HeLa cell cycle gene expression dataset by Whitfield et al. [48]. This dataset was previously used by Sambo et al. [41] to benchmark their algorithm and has since been used to benchmark other methods [16, 39, 40]. These methods focus on the third experiment of the dataset, which contains expression values at 47 hourly time points; we also use the same data for our applications.

The first reference subnetwork to which we compare our results, shown in Fig. 2a, consists of nine genes with interactions that were previously derived from BioGRID [49] and treated as ground truth by Sambo et al. [41]. However, since this network may be incomplete, any measures of performance that are interpreted with respect to it may not be indicative of a method's true predicative capability. For that matter, the known interactions in BioGRID have been updated since the analysis by Sambo et al. We therefore update the network using the interactions in the most recent release of BioGRID as of this writing (Release 3.4.160). We also compare our reconstructions to the updated network, shown in Fig. 2b, to evaluate our method's ability to make discoveries that were not known when the original network was curated.
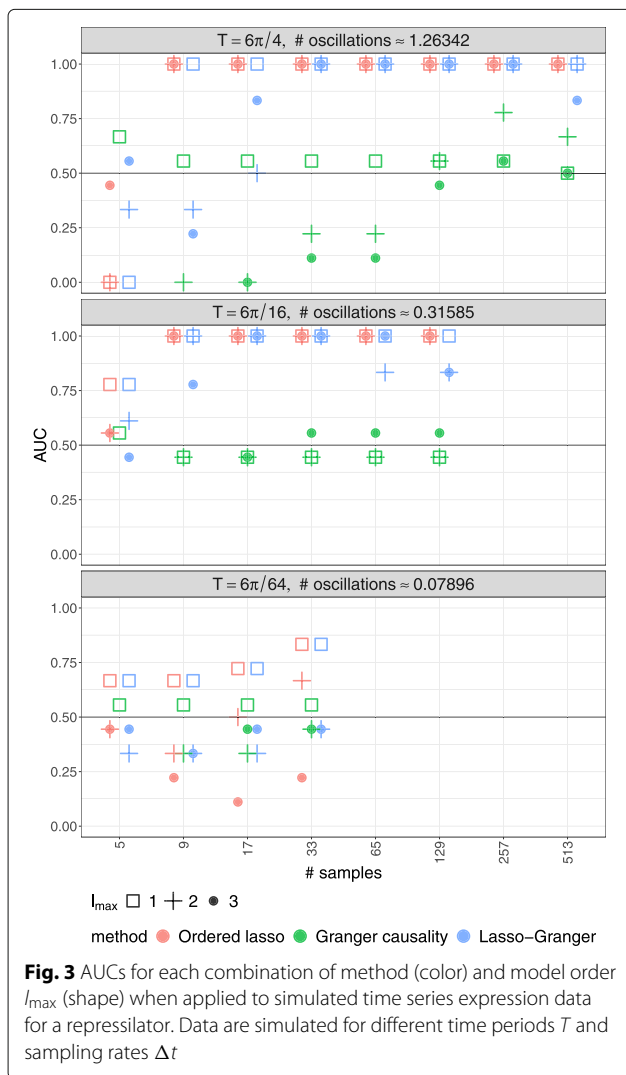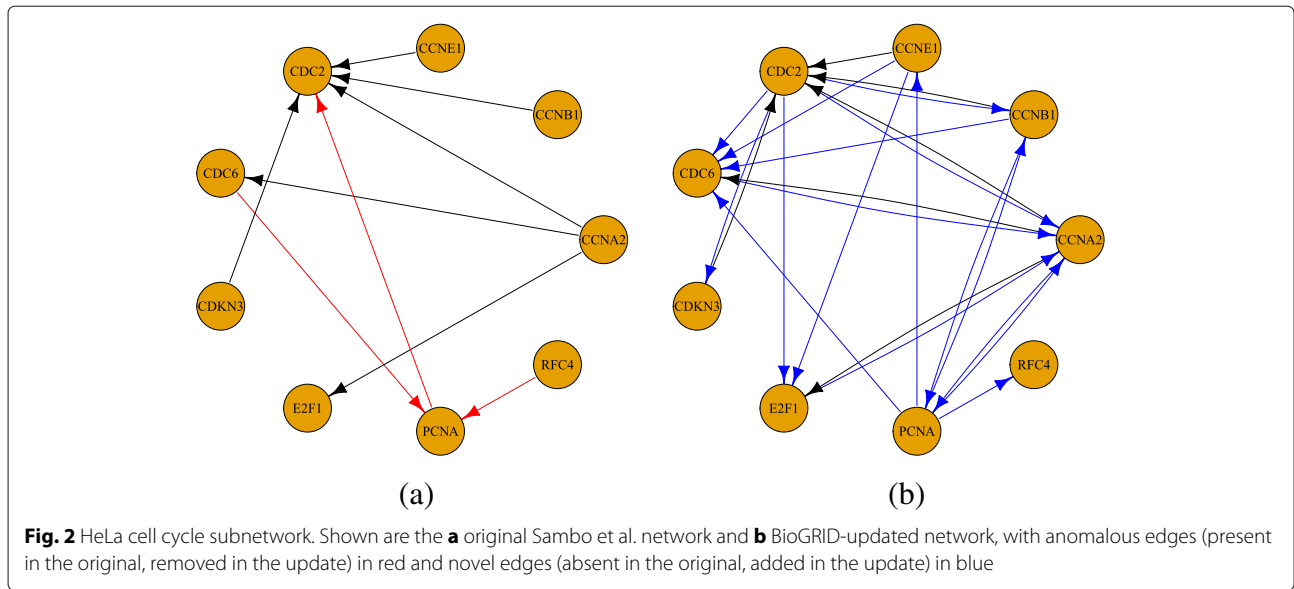
Results

Repressilator

We first evaluate our method using simulated data for a repressilator. We primarily investigate the effect of using different sampling intervals $\Delta t = \frac{6\pi}{2^j}$, $j \in \{2, 3, \ldots, 11\}$ and time series lengths $T = \frac{6\pi}{2^i}$, $i \in \{0, 1, \ldots, 9\}$ by simulating data using Eq. 3. In addition, we analyze the effect of the model order $l_{\max} \in \{1, 2, 3\}$ when fitting Eq. 1. In each case, AUCs are computed to analyze the prediction accuracies.

In Fig. 3, the repressilator AUCs are shown for the time-lagged Ordered Lasso at a subset of the aforementioned parameter values (the remaining values may be found in Additional file 1: Figure S-1). When T is large, many of these AUCs are 1, indicating that the time-lagged Ordered Lasso can correctly infer the network when an adequate amount of regularization is used to learn the expression models. As T decreases, the AUCs remain constant until much less than a period of oscillatory behavior is sampled. However, the AUCs remain above 0.5, so the method still does better than chance at identifying the true edges. When the time series is too short to observe any relevant dynamics, the method effectively does no better than chance. Therefore, using a time series that covers a sufficiently long period of time is necessary to ensure that a reliably accurate GRN is inferred.

However, low sampling rates can be detrimental. When the time series are extremely sparse because Δt is large relative to T, the AUCs degrade considerably, in some

Fig. 2 HeLa cell cycle subnetwork. Shown are the **a** original Sambo et al. network and **b** BioGRID-updated network, with anomalous edges (present in the original, removed in the update) in red and novel edges (absent in the original, added in the update) in blue

Fig. 3 AUCs for each combination of method (color) and model order l_{max} (shape) when applied to simulated time series expression data for a repressilator. Data are simulated for different time periods T and sampling rates Δt

cases to 0. However, when the time series are dense, the time-lagged Ordered Lasso produces high AUCs. Moreover, beyond a sampling rate when the surplus of sampled points do not provide any additional detail about the relevant dynamics, the AUCs do not change, therefore becoming robust to changes in Δt. Accordingly, Δt does not have to be extremely small to infer an accurate GRN, but the resulting time series should not be excessively sparse. (We note that Δt should not be an integer multiple of the system's oscillatory period; otherwise, the sampled data will be constant, and edges will be predicted at chance.)

Lastly, the effect of l_{max} on the AUCs appears to be negligible for large T and small Δt. This suggests that the time-lagged Ordered Lasso can accurately describe the repressilator's behavior with $l_{max} = 1$. Moreover, for $l_{max} > 1$, the time-lagged Ordered Lasso is able to suppress the effect of the additional lagged variables by enforcing the monotonicity constraint. However, when T is small or Δt is large relative to T, the AUCs appear to be sensitive to the choice of l_{max}. Since increasing l_{max} results in fewer samples to learn from, the accuracy of the time-lagged Ordered Lasso is expected to be robust to changes in l_{max} when it is small relative to the number of time points.

For comparison, Granger causality and Lasso-Granger AUCs are also shown in Fig. 3. Granger causality generally predicts edges at chance or worse, but performs comparably to the time-lagged Ordered Lasso when T is small. In addition, its AUCs are sensitive to changes in l_{max} and vary unpredictably with changes in T and Δt, making it difficult to suggest experimental designs for other GRNs. In contrast, Lasso-Granger tends to be on par with the time-lagged Ordered Lasso. For a one-lag model, there is no monotonicity constraint, so their AUCs match for

$l_{max} = 1$. For $l_{max} > 1$, the AUCs deviate when Δt has sufficiently increased; when Δt is large, the Lasso-Granger AUCs tend to decrease with increasing l_{max}, while the time-lagged Ordered Lasso AUCs are more robust, remaining at 1 in some cases.

Based on these results, the time-lagged Ordered Lasso has the potential to outperform other methods. Unlike Granger causality, it can handle short time series and still produce reasonably accurate networks. In addition, while Granger causality and Lasso-Granger allow higher lags to flexibly explain the repressilator's expression dynamics, they may correspond to false edges; in contrast, the time-lagged Ordered Lasso enforces a reasonable assumption about the diminishing strength of higher lags to mitigate their presence. Therefore, the repressilator is an example in which a better regression fit does not imply a more accurate GRN. Lastly, using time series that cover long periods of time can improve the time-lagged Ordered Lasso's ability to articulate the true edges, provided that the sampling rate is not extremely low. However, sampling over shorter periods with relatively high sampling rates to observe sufficient changes in expression can still produce fairly accurate networks. Therefore, the total number of observations, rather than frequency or length alone, is a major factor in inference accuracy.

DREAM

We next apply our method to the DREAM challenge datasets. Since the networks are fully known, biologically plausible, and endowed with detailed dynamical models of gene expression, these challenges serve as a testbed for benchmarking methods across different network sizes, topologies, sample sizes, and stochasticity conditions. As with the repressilator, we compute AUCs at different model orders l_{max}.

We study the overall performance of each method across the DREAM networks by considering the distribution of AUCs for each combination of method and l_{max}. In Fig. 4, densities are fit to the AUCs for each combination. Unlike with the repressilator, the three methods perform similarly across the DREAM datasets; their densities largely overlap, so the time-lagged Ordered Lasso is competitive with other methods on many of the datasets. In addition, the densities concentrate around moderately high AUC values, so the methods are capable of inferring true edges at rates better than chance. However, the median AUCs for Granger causality tend to be slightly higher than those of the other methods, attributable to the method having slightly better performance on a subset of the networks.

Other crucial differences between the methods can be identified. In particular, certain values of l_{max} can be used to obtain slight improvements in the overall accuracy of one method over another; based on the AUC density curves and medians at the considered l_{max} values (Fig. 4), the accuracy of the time-lagged Ordered Lasso appears to improve as l_{max} increases, while the Granger causality and Lasso-Granger AUCs peak at intermediate values of l_{max}. This suggests that inferring the most accurate GRNs possible for Granger causality and Lasso-Granger may require
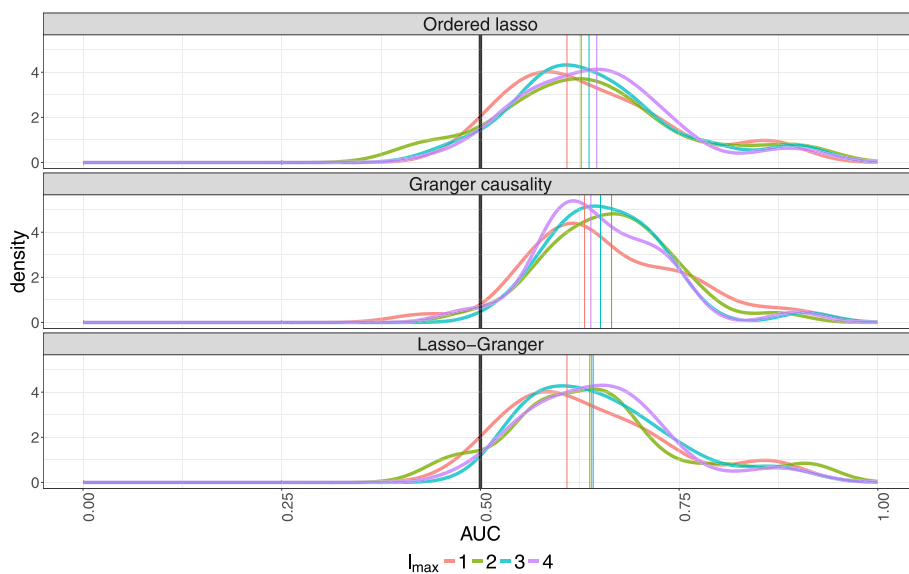

Fig. 4 Densities fit to the DREAM AUCs for each combination of method and model order l_{max}. Vertical lines indicate empirical medians. Note that the AUCs for the time-lagged Ordered Lasso increase monotonically with l_{max}, in contrast to the other methods

optimizing l_{max}. However, since the GRNs are generally not fully known beforehand, devising heuristics or methods to select l_{max} and maximize the prediction accuracy may be difficult. In contrast, the time-lagged Ordered Lasso results suggest that large values of l_{max} are preferable to take advantage of automatic maximum effective lag selection through the monotonicity constraints. That is, we do not need to optimally select l_{max}.

HeLa cell cycle subnetwork

Lastly, to evaluate its performance on real datasets, we apply our method to the HeLa cell cycle gene expression dataset by Whitfield et al. [48]. To compute AUCs, we first consider the subnetwork curated by Sambo et al. [41] using then-known interactions from BioGRID as the ground truth network. However, this network has since been updated to include additional discoveries. Consequently, AUCs computed with respect to the original network are not indicative of a method's true performance, but they can be useful to illustrate the effects of treating partially known networks as the gold standard and the cautionary measures that are required. Therefore, we also compute AUCs based on an updated network that consists of interactions among the same genes from a recent

release of BioGRID (Release 3.4.160). Although this network may still only be considered "partially" known as there may still be edges among these genes that have yet to be discovered, treating it as the "truth" will provide a more reliable estimate of a method's prediction accuracy than the older version will.

In Fig. 5, AUCs computed with respect to the updated network are shown for each method and model order $l_{max} \in \{1, \ldots, 6\}$; in the inset, AUCs computed with respect to the original Sambo et al.-network are shown. With the updated network, the time-lagged Ordered Lasso AUCs tend to increase as l_{max} increases, eventually attaining the highest values across all methods and at rates better than chance. In contrast, when they are computed based on the original network, the AUCs suggest that the time-lagged Ordered Lasso does no better than chance at predicting the true network. Therefore, the original network AUCs may be inaccurate and misleading indicators of accuracy by virtue of the original network's relative incompleteness, and by considering the updated network, our outlook on the utility and comparability of the methods readjusts considerably. Likely due to the high-ranking novel edges that were previously considered false positives with respect to the original network,
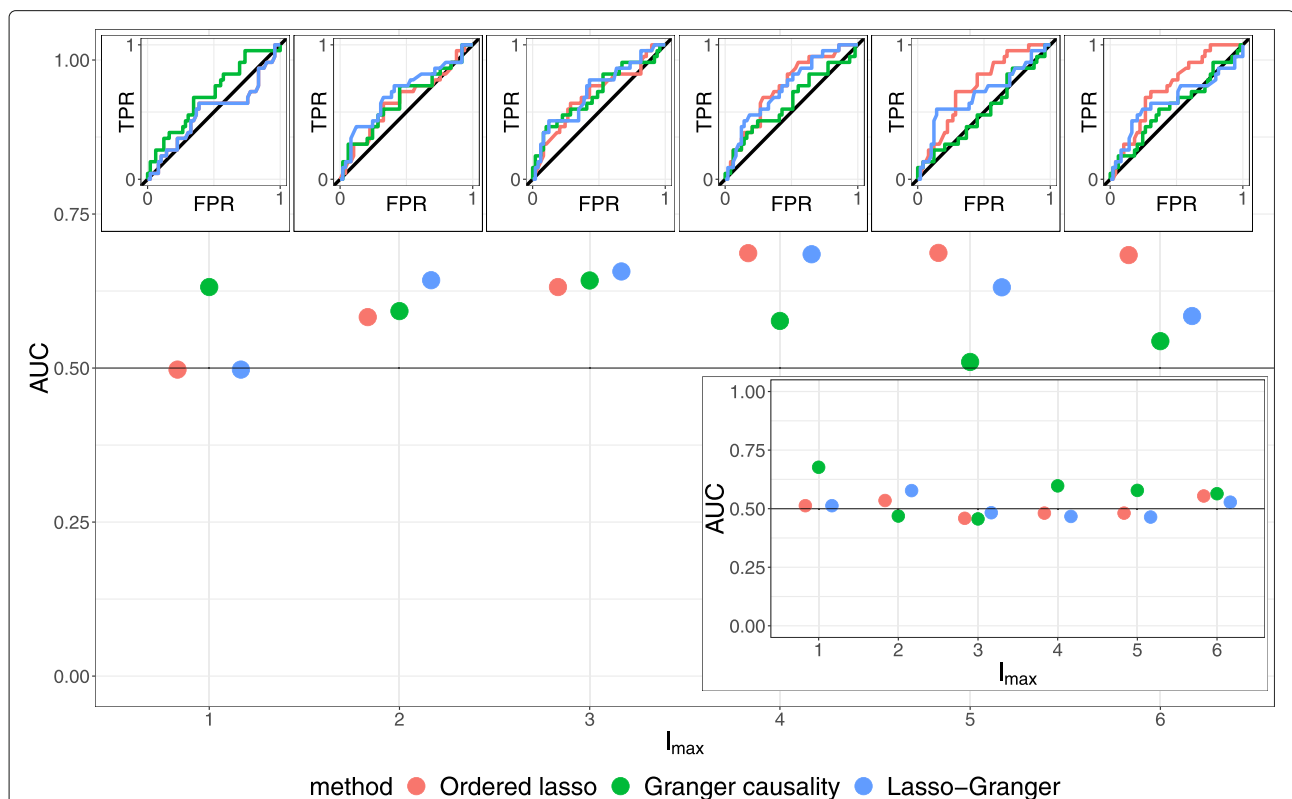

Fig. 5 Time-lagged Ordered Lasso, Granger-causality, and Lasso-Granger AUCs at different model orders l_{max} when applied to the HeLa cell cycle expression dataset with the BioGRID-updated network and (lower inset) Sambo et al.-network treated as ground truth networks. ROC curves with respect to the updated network are shown in the upper inset plots; the black diagonal line corresponds to predicting edges by pure chance

incorporating the novel and anomalous edges generally leads to higher AUCs that are respectable in light of the low time resolution of the data, which consequently suggests that the time-lagged Ordered Lasso can predict true edges from sparsely sampled data at rates much better than chance and certainly better than suggested by the Sambo et al.-network. Importantly, this demonstrates that the time-lagged Ordered Lasso can make discoveries that were not known at the time that the original network was curated.

While Granger causality and Lasso-Granger can outperform the time-lagged Ordered Lasso, they only do so at particular values of l_{max} and, even then, do not achieve the highest overall AUCs. In addition, at the larger values of l_{max}, the Ordered Lasso is subject to the most restrictive regression constraints of the three methods, but still achieves the highest AUCs, so we again see that a better regression fit does not imply a more accurate GRN. Furthermore, the AUCs of the competing methods may vary unpredictably with l_{max}, making it difficult to optimize when constructing GRNs. For example, with multiple local minima and maxima in the Granger causality AUCs, an arbitrary choice of l_{max} may not produce the best possible Granger causality-based network. The Lasso-Granger AUCs trend somewhat more predictably, increasing as l_{max} increases to 4 and decreasing afterwards, but it is not apparent how l_{max} may be optimized to maximize the AUC when the network is not known beforehand. In contrast, the time-lagged Ordered Lasso AUCs appear to increase monotonically with l_{max} and stabilize when $l_{max} \geq 4$, suggesting that the predicted networks barely change beyond a certain l_{max} for a sufficiently long time series; this is likely due to the monotonicity constraint taking full effect and ignoring the additional lagged variables that are introduced. This attribute and the results suggest that the time-lagged Ordered Lasso can optimally recover the true GRN from an expression dataset without a complicated heuristic to select l_{max}.
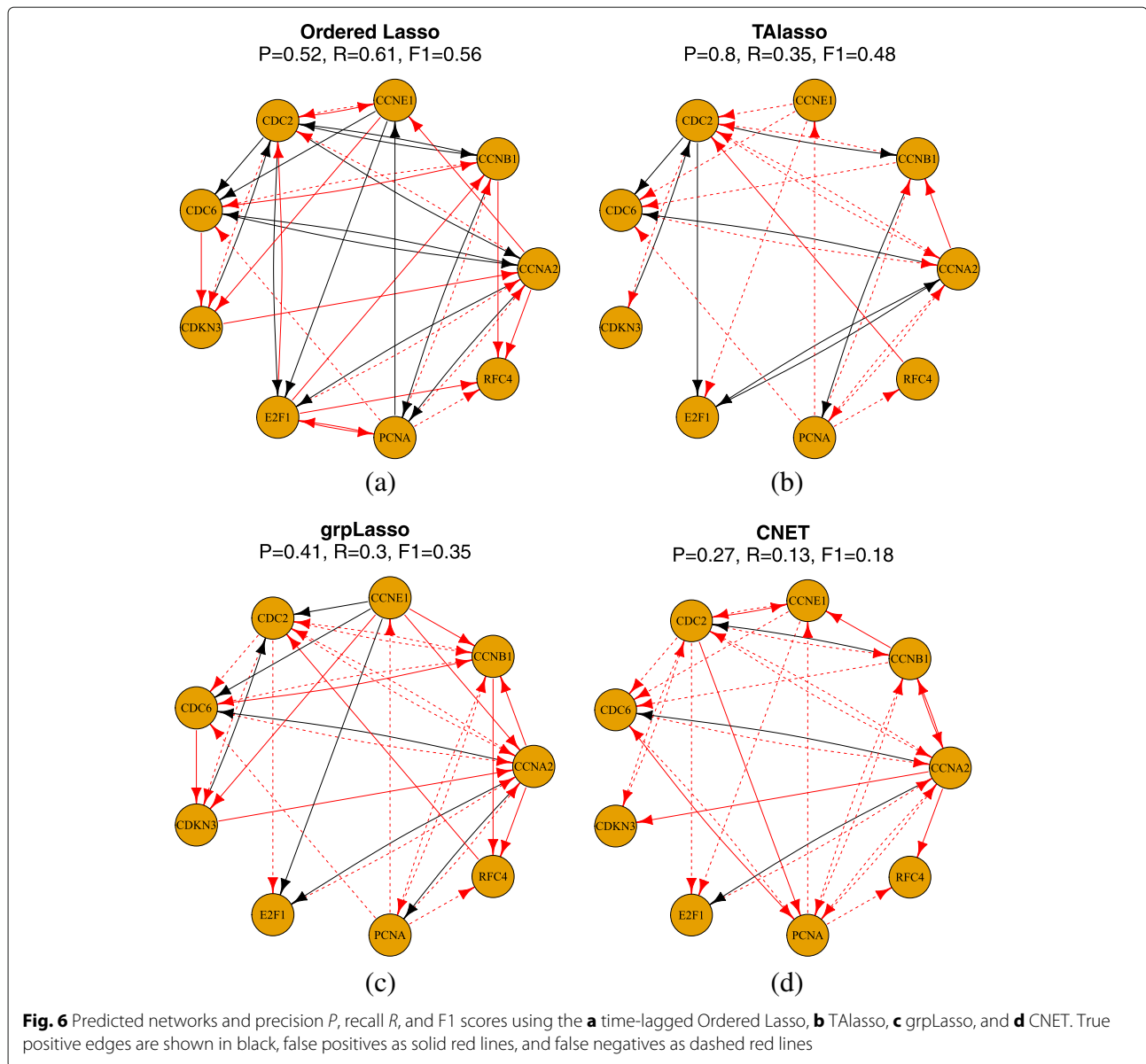
In summary, these results demonstrate several important properties of the time-lagged Ordered Lasso's GRN inference capabilities. The AUCs computed based on the updated subnetwork suggest that our method is able to derive accurate GRNs from time series gene expression data, even when it is sparsely sampled in time. In contrast, the AUCs computed against the original, incomplete Sambo et al.-network are lower and more volatile, suggesting that our method is able to discover relationships that were not known when the original network was curated. Furthermore, despite enforcing the most restrictive regression constraints of the three methods, the time-lagged Ordered Lasso is able to utilize the monotonicity constraint to outperform other methods. In particular, the inferred networks and AUCs are robust to the model order when it is sufficiently large, and these AUCs are the

highest across all methods. This suggests that an accurate GRN may be efficiently inferred with the time-lagged Ordered Lasso by simply choosing a sufficiently large model order that is permissable given the length of a time series to allow the constraint to optimize the maximum effective lag; other methods may require intricate or computationally intensive approaches to choose the model order and may still not predict the most accurate GRN. These features therefore make the time-lagged Ordered Lasso a viable mainstay for additional reconstruction analyses and approaches, and modifications such as an adaptive lasso [50] step to introduce specific source-target penalties may further improve prediction accuracy.

Predicted network comparisons

We next compare our method to the truncating adaptive lasso (TAlasso) [39], grouped graphical Granger modeling (grpLasso) [40], and CNET [41], other algorithms that have been applied the HeLa expression dataset. Since these methods have no notion of an AUC (CNET) or have been designed to select particular parameters (TAlasso, grpLasso), we compare their predicted networks with those of the time-lagged Ordered Lasso at a fixed value of λ. To select λ, we use the same heuristic used by TAlasso. We note that the guarantees provided by this heuristic do not necessarily apply to our approach and that the additional adaptive lasso weights of the TAlasso result in larger effective penalties than are actually suggested by the heuristic. In addition, other suitable heuristics could have been chosen to select λ that may result in better true and false positive rates, so this choice of heuristic is only for comparative purposes. Since the parameter α for this heuristic was not specified by the authors, we select the customary $\alpha = 0.01$ and compute λ following Eq. 9 of [39]. Since we showed that the time-lagged Ordered Lasso AUCs increased and stabilized with increasing l_{max}, we set $l_{max} = 6$.

In Fig. 6, the predicted networks of the time-lagged Ordered Lasso and the three reference methods are shown. The networks of the reference methods have been reconstructed from the results presented by the authors [39] and updated to reflect the changes in BioGRID. Based on the updated network, the time-lagged Ordered Lasso is second to TAlasso in terms of precision, but achieves the highest recall and F1 score amongst the four methods. Since the F1 score is the harmonic mean of the precision and recall, the time-lagged Ordered Lasso is able to best balance the ability to recover many true edges while ensuring that many of the predicted edges are indeed true edges. In contrast, while TAlasso has very high precision, it recovers half as many true edges as the time-lagged Ordered Lasso does, resulting in a lower F1 score and overall weaker performance. Other methods are substantially less accurate than both TAlasso and time-lagged

Fig. 6 Predicted networks and precision *P*, recall *R*, and F1 scores using the **a** time-lagged Ordered Lasso, **b** TAlasso, **c** grpLasso, and **d** CNET. True positive edges are shown in black, false positives as solid red lines, and false negatives as dashed red lines

Ordered Lasso. Even though the TAlasso heuristic may not be optimal with respect to the time-lagged Ordered Lasso, our method still produces reasonably accurate networks, and further modifications such as choosing λ on a per-gene basis or different heuristics that are more specific to the time-lagged Ordered Lasso may improve its network predictions. In addition, these results can be used to guide a choice between TAlasso and the time-lagged Ordered Lasso, depending on the importance of specificity versus sensitivity as well as predicting a sparse network versus the potential to discover more novel edges that may be verified with follow-up experiments, especially when the reference networks may only be partially known.

Semi-supervised application

The availability of the original and updated networks also presents an opportunity to analyze the semi-supervised time-lagged Ordered Lasso adaptation. For illustrative purposes, we evaluate the method's ability to predict novel edges by treating the original Sambo et al.-network as the input prior network and setting λ_{edge} to 0. We again compute AUCs, this time by tracking the prior non-edges that enter an expression model as $\lambda_{non\text{-}edge}$ decreases from a sufficiently large value (corresponding to no prior non-edges predicted as posterior edges). This AUC may be interpreted as the probability that a randomly chosen true novel edge is ranked higher or enters a model earlier than a randomly chosen true non-edge.

The novel edge prediction AUCs for model orders $l_{max} \in \{1, \ldots, 6\}$ are shown in Fig. 7. Similar to the de novo case, the AUCs tend to increase and level off as l_{max} increases. More importantly, the AUCs at the larger values of l_{max} are well above 0.5, indicating that the semi-supervised method can predict novel edges at rates better than chance using the described parameter settings. Since all prior edges were unpenalized in these results, possible improvements in accuracy can be made by choosing positive values of λ_{edge}, which can also facilitate anomalous edge detection. Nevertheless, the time-lagged Ordered Lasso already displays a strong potential for reliable novel edge detection; even without these adjustments, the current semi-supervised adaptation is still able to synthesize a partially known GRN with an expression dataset to resolve the inconsistencies between both inputs and accurately identify the missing edges in the GRN.

Discussion

The time-lagged Ordered Lasso imposes a monotonicity constraint based on temporal distance that is adequate for many time series applications, performs model regularization, and has a canonical feature selection mechanism, making it well-suited for GRN reconstruction. We have

presented adaptations of the method for de novo and semi-supervised reconstruction from time-course gene expression data. To do so, we assumed that the expression of a gene depended linearly on the expression of its regulators at multiple preceding time points and that the regulatory strength of a predictor decreased for increasing lags. A local model of gene expression is then learned for each gene using the time-lagged Ordered Lasso, and a GRN is predicted by applying the feature selection mechanism on each gene's model to determine the predicted regulators. To modify the de novo method for semi-supervised reconstruction, we introduced a second regularization parameter that allows us embed a prior GRN into the model fitting procedure in order to predict novel and anomalous edges.

In our applications, we showed that the time-lagged Ordered Lasso enforces the monotonicity constraint to accurately predict a variety of networks. In most cases, the time-lagged Ordered Lasso performed on par with or better than competing methods. Most importantly, we showed that it can accurately discover novel network connections and anomalous links using real data, as demonstrated by the improved performance when compared to the updated HeLa network. Specifically, the time-lagged
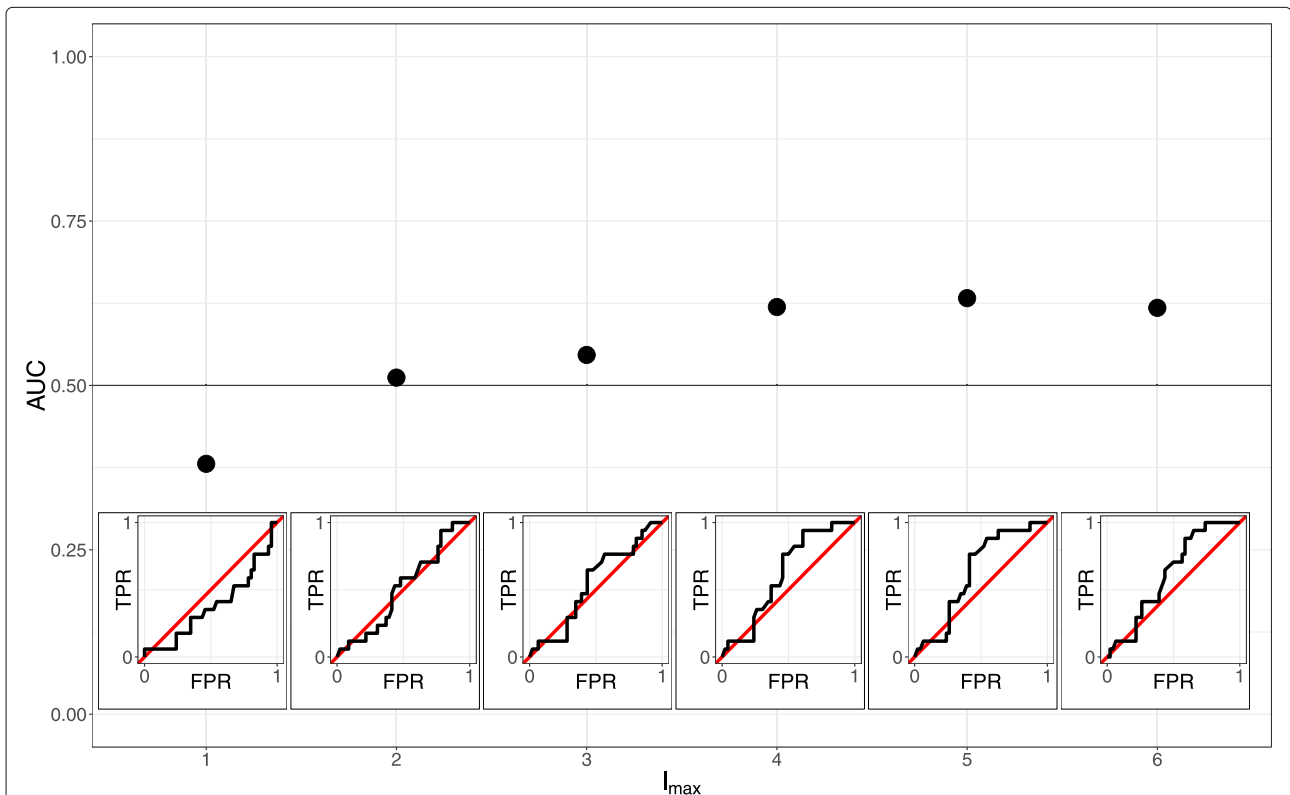

Fig. 7 Novel edge prediction AUCs for the time-lagged Ordered Lasso at different model orders l_{max} when applied to the HeLa cell cycle expression dataset with the Sambo et al.-network as the prior network and BioGRID-updated network as the ground truth network. The red diagonal line corresponds to predicting edges by pure chance

Ordered Lasso predicted edges that were not known at the time that the HeLa data was published and would have been erroneously considered false positives with the respect to the Sambo et al.-network, but were later confirmed by further experiments. This is an important validation of the time-lagged Ordered Lasso's capabilities.

Our results illustrated several important properties of the time-lagged Ordered Lasso adaptations. For instance, provided that a time series covers a sufficiently long period of time and is not extremely sparse, our method was able to accurately recover GRNs from the data, whereas other methods had more difficulty doing so under the same conditions. In addition, predicting a GRN from a fitted model only required checking the first lagged variable of each predictor. However, because the additional lagged variables of one gene may better explain a target gene's evolution in expression than the lagged variables of multiple other genes in a lower-lag model will, the higher order lags will still be important to the model and reduce false positive edge predictions at adequately chosen penalty parameters. Lastly, because of the monotonicity constraint, the time-lagged Ordered Lasso can automatically select the maximum effective lag of influence for each gene-gene pair, so the predicted GRNs are expected to be robust to the model order if a time series is sufficiently long and the model order is sufficiently large. As a result, the monotonicity constraint precludes the need for any complicated heuristics to choose the model order that other approaches may require to optimally reconstruct a GRN.

Our algorithms can be modified in several ways. Here, we assumed that the expression of a gene depended linearly on the lagged expression of its predictors. However, we included the lagged expression of the gene itself as covariates, even if self-regulation was not evident; one modification is removing them. Another common modeling approach is using differential equations. Details and results for these changes may be found in Additional file 1: Sections S-1 and S-6. As with multiple linear regression, the addition of non-linear and interaction terms can improve the fit of a model and allow for more complex, realistic dependencies. However, we observed in our applications that an improved fit does not necessarily imply a more accurate GRN. In addition, while this extension only requires a straightforward specification of new variables to include, having too many terms may be computationally restrictive, so some knowledge of which non-linear terms and interactions may be useful, in light of the sparsity of the data, is required. Thus, using linearity as a simplifying assumption serves to prevent overfitting and reduces computational overhead while remaining adequate for most applications, especially when detailed dynamics are difficult to observe due to the short time coverage and sparse sampling of a

dataset. By imposing monotonicity constraints in Eqs. 1 and 2, we also implicitly assumed that the influence of a predictor on a target always began with the immediately preceding time point. Therefore, the expression models can also be modified to account for larger delays of dependence, but this may require new approaches or substantial changes to the underlying time-lagged Ordered Lasso method to automatically select the delay. Alternatively, one may choose to measure expression data at sparser rates or subsample an existing dataset, but these approaches will require some knowledge of an appropriate delay. In some cases, a monotonicity constraint may inaccurately explain the expression dynamics and eliminate true regulatory genes from consideration, such as when there is a large delay of dependence. In fact, the relaxation to Lasso-Granger improved the AUCs at certain lags for some of the DREAM networks. Thus, the time-lagged Ordered Lasso may not always be appropriate, so other modifications may involve deciding when to relax the constraint. Furthermore, we have not investigated the impact that the level of noise has on the accuracy of our method, particularly when the expression data is derived from low molecule number measurements. The implicit assumption that the data is collected using high molecule numbers is currently a limitation of the method, so the stochasticity that is incurred in the low copy number case may be investigated in further detail, including the tolerance to noise and what additional modifications and parameter choices should be made, if any, to effectively deal with considerable amounts of noise. Lastly, when comparing the HeLa-predicted networks, we applied a heuristic used by another method to choose the lasso penalty that may not have resulted in optimal predictions for our approach. Another avenue for extensions may therefore involve designing new heuristics or employing commonly used heuristics such as BIC optimization to improve network predictions. Additional extensions include adding different regularization parameters for different genes, algorithms to automatically choose those parameters, and other feature selection procedures to infer edges.

Conclusion

While GRN inference remains challenging, our approach provides several advances. First, to infer GRNs, our approach uses a time-ordered constraint on regulatory influence, which we showed can accurately predict a variety of networks. Our approach can also accommodate prior knowledge for semi-supervised GRN inference. In addition, the performance of our methods increases monotonically with the maximum lag of an expression model, obviating the need to optimize that parameter. Lastly, our methods also have the ability to make accurate

novel discoveries, as demonstrated with the BioGRID example.

Even without extensive modifications, our current algorithm is still able to predict fairly accurate GRNs with reasonable, basic assumptions for dynamic gene expression modeling. Thus, the GRNs that are inferred using the time-lagged Ordered Lasso can be used as starting points for further analyses and network refinements, and the time-lagged Ordered Lasso can serve as a backbone for additional GRN reconstruction algorithms.

Additional file

Additional file 1: Supplementary information. (PDF 289 kb)

Authors' contributions
PN and RB designed the study. PN developed the methods and analyzed the data. All authors participated in writing the manuscript. All authors have 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.

References
1. Brazhnik P, de la Fuente A, Mendes P. Gene networks: how to put the function in genomics. Trends Biotechnol. 2002;20(11):467–72.
2. Karlebach G, Shamir R. Modelling and analysis of gene regulatory networks. Nat Rev Mol Cell Biol. 2008;9:770–80.
3. MacNeil LT, Walhout AJM. Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. Genome Res. 2011;21:645–57.
4. Thompson D, Regev A, Roy S. Comparative Analysis of Gene Regulatory Networks: From Network Reconstruction to Evolution. Annu Rev Cell Dev Biol. 2015;31(1):399–428.
5. Iyer AS, Osmanbeyoglu HU, Leslie CS. Computational methods to dissect gene regulatory networks in cancer. Curr Opin Syst Biol. 2017;2:115–22.
6. de Jong H. Modeling and simulation of genetic regulatory systems: a literature review. J Comput Biol. 2002;9:67–103.
7. Wang YXR, Huang H. Review on statistical methods for gene network reconstruction using expression data. J Theor Biol. 2014;362:53–61.
8. Schnoerr D, Sanguinetti G, Grima R. Approximation and inference methods for stochastic biochemical kinetics—a tutorial review. J Phys A Math Theor. 2017;50(9):093001.
9. Komorowski M, Finkenstädt B, Harper CV, Rand DA. Bayesian inference of biochemical kinetic parameters using the linear noise approximation. BMC Bioinformatics. 2009;10(1):343.
10. Fröhlich F, Thomas P, Kazeroonian A, Theis FJ, Grima R, Hasenauer J. Inference for Stochastic Chemical Kinetics Using Moment Equations and System Size Expansion. PLoS Comput Biol. 2016;12(7):1–28.
11. Boys RJ, Wilkinson DJ, Kirkwood TBL. Bayesian inference for a discretely observed stochastic kinetic model. Stat Comput. 2008;18(2):125–35.
12. Tiberi S, Walsh M, Cavallaro M, Hebenstreit D, Finkenstädt B. Bayesian inference on stochastic gene transcription from flow cytometry data. Bioinformatics. 2018;34(17):i647–55.
13. Rice JJ, Tu Y, Stolovitsky G. Reconstructing biological networks using conditional correlation analysis. Bioinformatics. 2005;21(6):765–73.
14. Butte AJ, Kohane IS. Unsupervised knowledge discovery in medical databases using relevance networks. In: Proceedings of the AMIA Symposium. America Medical Informatics Association. 1999. p. 711–5.
15. Butte AJ, Kohane IS. Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. Pac Symp Biocomput. 2000;5:415–26.
16. Tam GHF, Chang C, Hung YS. Gene regulatory network discovery using pairwise Granger causality. IET Syst Biol. 2013;7(5):195–204.
17. Haury AC, Mordelet F, Vera-Licona P, Vert JP. TIGRESS: Trustful Inference of Gene REgulation using Stability Selection. BMC Syst Biol. 2012;6(1):145.
18. van Someren E, Wessels L, Reinders M, Backer E. Regularization and Noise Injection for Improving Genetic Network Models. In: Zhang W, Shmulevich I, editors. Computational and Statistical Approaches to Genomics. Boston: Springer US; 2006. p. 279–95.
19. Friedman N. Inferring cellular networks using probabilistic graphical models. Science. 2004;303(5659):799–805.
20. Kauffman S. Homeostasis and differentiation in random genetic control networks. Nature. 1969;224:177–8.
21. Kauffman SA. Metabolic stability and epigenesis in randomly constructed genetic nets. J Theor Biol. 1969;22(3):437–67.
22. Qi H, Cheng D. Analysis and control of Boolean networks: A semi-tensor product approach. In: 2009 7th Asian Control Conference; 2009. p. 1352–6.
23. Xiao Y. A tutorial on analysis and simulation of boolean gene regulatory network models. Curr Genomics. 2009;10(7):511–25.
24. Shmulevich I, Dougherty ER, Kim S, Zhang W. Probabilistic Boolean Networks: a rule-based uncertainty model for gene regulatory networks. Bioinformatics. 2002;18(2):261–74.
25. Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P. Inferring regulatory networks from expression data using tree-based methods. PLoS ONE. 2010;5(9):e12776.
26. Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, Cottarel G, et al. Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. PLoS Biol. 2007;5(1):e8.
27. Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitsky G, Favera RD, et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics. 2006;7:S7.
28. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 2000;28:27–30.
29. Kanehisa M, Goto S, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. 2014;42:D199–205.
30. Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, et al. The Reactome Pathway Knowledgebase. Nucleic Acids Res. 2018;46(D1):D649–55.
31. Nguyen P, Braun R. Semi-supervised network inference using simulated gene expression dynamics. Bioinformatics. 2018;34(7):1148–56.
32. Li Y, Jackson SA. Gene Network Reconstruction by Integration of Prior Biological Knowledge. G3 Genes Genome Genet. 2015;5(6):1075–9.
33. Imani M, Braga-Neto UM. Maximum-Likelihood Adaptive Filter for Partially Observed Boolean Dynamical Systems. IEEE Trans Signal Process. 2017;65(2):359–71.

34. Mcclenny LD, Imani M, Braga-Neto UM. BoolFilter: an R package for estimation and identification of partially-observed Boolean dynamical systems. BMC Bioinformatics. 2017;18(1):519.
35. Granger CWJ. Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. Econometrica. 1969;37(3):424–38.
36. Arnold A, Liu Y, Abe N. Temporal Causal Modeling with Graphical Granger Methods. In: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '07. New York,: ACM; 2007. p. 66–75.
37. Tibshirani R, Suo X. An Ordered Lasso and Sparse Time-Lagged Regression. Technometrics. 2016;58(4):415–23.
38. Tibshirani R. Regression Shrinkage and Selection Via the Lasso. J R Stat Soc Ser B. 1996;58:267–88.
39. Shojaie A, Michailidis G. Discovering graphical Granger causality using the truncating lasso penalty. Bioinformatics. 2010;26(18):i517–23.
40. Lozano AC, Abe N, Liu Y, Rosset S. Grouped graphical Granger modeling for gene expression regulatory networks discovery. Bioinformatics. 2009;25(12):i110.
41. Sambo F, Di Camillo B, Toffolo G. CNET: an algorithm for reverse engineering of causal gene networks. Varenna: NETTAB2008; 2008.
42. Elowitz MB, Leibler S. A synthetic oscillatory network of transcriptional regulators. Nature. 1999;403(6767):335–8.
43. Stolovitsky G, Prill RJ, Califano A. Lessons from the DREAM2 Challenges. Ann N Y Acad Sci. 2009;1158:159–95.
44. Prill RJ, Marbach D, Saez-Rodriguez J, Sorger PK, Alexopoulos LG, Xue X, et al. Towards a rigorous assessment of systems biology models: the DREAM3 challenges. PLoS ONE. 2010;5:e9202.
45. Marbach D, Schafter T, Mattiussi C, Floreano D. Generating realistic in silico gene networks for performance assessment of reverse engineering methods. J Comput Biol. 2009;16(2):229–39.
46. Marbach D, Prill RJ, Schafter T, Mattiussi C, Floreano D, Stolovitsky G. Revealing strengths and weaknesses of methods for gene network inference. Proc Natl Acad Sci. 2010;107:6286–91.
47. Schaffter T, Marbach D, Floreano D. GeneNetWeaver: in silico benchmark generation and performance profiling of network inference methods. Bioinformatics. 2011;27(16):2263–70.
48. Whitfield ML, Sherlock G, Saldanha AJ, Murray JI, Ball CA, Alexander KE, et al. Identification of Genes Periodically Expressed in the Human Cell Cycle and Their Expression in Tumors. Mol Biol Cell. 2002;13(6):1977–2000.
49. Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M. BioGRID: a general repository for interaction datasets. Nucleic Acids Res. 2006;34(suppl_1):D535–9.
50. Zou H. The Adaptive Lasso and Its Oracle Properties. J Am Stat Assoc. 2006;101(476):1418–29.