The reconstruction of gene regulatory networks (GRNs) from mRNA expression data involves various approaches that have been developed to deduce gene regulatory networks from microarray expression data. Candidate transcription factor-target gene (TF-TG) relationships are assumed more likely if the expression of TF and TG are mutually dependent. Candidate transcription factor-target gene (TF-TG) relationships are assumed more likely if the expression of TF and TG are mutually dependent. Candidate transcription factor-target gene (TF-TG) relationships are assumed more likely if the expression of TF and TG are mutually dependent.

Results: We present a new score for network inference, $\eta^2$, that is derived from an analysis of variance. Candidate transcription factor-target gene (TF-TG) relationships are assumed more likely if the expression of TF and TG are mutually dependent in at least a subset of the examined experiments. We evaluate this dependency by $\eta^2$, a non-parametric, non-linear correlation coefficient. It is fast, easy to apply and does not require the discretization of the input data. In the recent DREAM5 blind assessment, the arguably most comprehensive evaluation of inference methods, our approach based on $\eta^2$ was rated the best performer on real expression compendia. It also performs better than methods tested in other recently published comparative assessments. About half of our predicted novel interactions are true interactions as estimated from qPCR experiments performed for DREAM5.

Conclusions: The score $\eta^2$ has a number of interesting features that enable the efficient detection of gene regulatory interactions. For most experimental setups, it is an interesting alternative to other measures of dependency such as Pearson’s correlation or mutual information.

Availability: See http://www2.bio.ifi.lmu.de/~kueffner/anova.tar.gz for code and example data.

Contact: kueffner@bio.ifi.lmu.de

Supplementary information: Supplementary data are available at Bioinformatics online.

Received on July 11, 2011; revised on March 18, 2012; accepted on March 22, 2012

1 INTRODUCTION

The reconstruction of gene regulatory networks (GRNs) from expression data can help to improve our understanding of molecular regulation events. A variety of algorithms have been devised to predict gene regulatory interactions, frequently based on mutual regulation events. A variety of algorithms have been devised to predict gene regulatory interactions, frequently based on mutual regulation events.

For most experimental setups, it is an interesting alternative to other measures of dependency such as Pearson’s correlation or mutual information. For most experimental setups, it is an interesting alternative to other measures of dependency such as Pearson’s correlation or mutual information.

One source of false positive predictions are indirect effects, i.e. in a cascade $A \rightarrow B \rightarrow C$ methods are likely to also predict the additional effect $A \rightarrow C$. Extensions like the data processing inequality [ARACNe, Margolin et al. (2004)] and gene dependent background distributions [CLR, Faith et al. (2007)] have been proposed to overcome this problem. The minimum redundancy, maximum relevance concept [Ding and Peng (2005); Meyer et al. (2006)] offers another way to select important edges. Indirect effects might also be identified and removed by partial correlations [Castelo and Roverato (2009)]. Elastic net or lasso, a L1-penalized estimation of the inverse covariance matrix [Friedman et al. (2008); Kabir et al. (2010)] offers another way to select important edges. Indirect effects might also be identified and removed by partial correlations [Castelo and Roverato (2009)].

A high proportion of our predicted novel interactions were confirmed by small-scale qPCR experiments performed by the DREAM5 organizers. In addition, our approach was evaluated as the best performer for the inference of real datasets in the recent DREAM5 blind assessment [http://wiki.c2b2.columbia.edu/dream/index.php; Marbach et al. (2010); Prill et al. (2011)]. Here, 29 participating teams applied a diverse set of inference methods to a variety of large real (Escherichia coli, Saccharomyces cerevisiae) and artificial expression compendia with thousands of genes from several hundreds of microarray measurements. The microarray experiments consisted of various gene, drug or environmental perturbations that were in some cases carried out as time courses.

After a brief summary of related work we describe the GRN inference setting, introduce our inference approach based on the score $\eta^2$ and evaluate its properties and performance.

1.1 Related work

The inference of large GRNs of 500+ nodes is frequently tackled by unsupervised, data-driven approaches that aim to resolve dependencies from expression data alone. We briefly review some commonly used techniques in the following and refer the reader to review papers [e.g. by Altay and Emmert-Streib (2010b), Lee and Tzou (2008) and Markowetz and Spang (2007)] for a more comprehensive overview of methods.

Unparameterized topologies can be approximated even for large networks by measures of pairwise gene dependencies, e.g. using Pearson’s linear correlation coefficient [Bunte and Kohand (1999)]. To take non-linear correlations into account, information theoretic approaches can be employed such as Bayes conditional probability tables or mutual information [Butte and Kohand (2004); Ding and Peng (2005); Faith et al. (2007); Margolin et al. (2006); Meyer et al. (2007); Zhang et al. (2007)]. The latter techniques require a very careful discretization of the expression data to avoid the loss of signal [Alay and Emmert-Streib (2010); Mukherjee and Spang (2008); Zhu et al. (2008)].

One source of false positive predictions are indirect effects, i.e. in a cascade $A \rightarrow B \rightarrow C$ methods are likely to also predict the additional effect $A \rightarrow C$. Extensions like the data processing inequality [ARACNe, Margolin et al. (2004)] and gene dependent background distributions [CLR, Faith et al. (2007)] have been proposed to overcome this problem. The minimum redundancy, maximum relevance concept [Ding and Peng (2005); Meyer et al. (2006)] offers another way to select important edges. Indirect effects might also be identified and removed by partial correlations [Castelo and Roverato (2009)]. Elastic net or lasso, a L1-penalized estimation of the inverse covariance matrix [Friedman et al. (2008); Kabir et al. (2010)] offers another way to select important edges. Indirect effects might also be identified and removed by partial correlations [Castelo and Roverato (2009)].

All of the mentioned approaches measure global dependencies, i.e. dependencies that are visible...
across the majority of measured experimental conditions. Local dependencies that are only apparent in a subset of the conditions [Kwon et al. 2003] might thus be missed.

Models like Boolean, probabilistic Bayesian networks, ordinary differential equations (ODE) or Petri Nets are generative i.e. they allow the generation of the original training datasets by simulation. Optimization approaches minimize the deviation from given data by parameterizing models [Gauthke et al. 2004; Kuffner et al. 2004; Wang et al. 2004]. Due to the huge parameter space these algorithms may not scale well to large networks.

The assessment of the multitude of reconstruction algorithms is quite difficult. Comparative studies [Iachuci et al. 2004; Michael et al. 2004; Narendra et al. 2009; Zou and Feng 2009] evaluate only subsets of approaches. More comprehensive assessments are facilitated through community-wide challenges conducted by the DREAM consortium.

2 METHODS

2.1 Inference setting, data sources and evaluation

Problem statement and evaluation. GRN inference aims at the detection of gene regulatory relationships from mRNA expression datasets. The task is to reverse engineer the directed topology of one network for each of the available expression datasets (Table 1). In the following, we describe a setup for the evaluation of inference methods that has been adopted by many comparative assessment studies including DREAM5 (http://wiki.c2b2.columbia.edu/dream/index.php/D5S4) and Narendra et al. [2011].

For each dataset, potential TEs are given. Only those TEs should be included as regulators in the network predictions as the used gold standards do not contain gene regulatory interactions for other regulators such as sigma factors or miRNAs. The list of TEs was available to all participants of the challenge. Approaches were then required to check and rank (TE x Genes) candidate relationships. Lists of ranked candidate interactions are evaluated against the true topology (in case of the artificial dataset) or against experimentally determined TF:TG interactions. Candidate lists are evaluated against gold standard networks (see below) based on the area under the precision-recall curve (AUPR) and the area under the receiver-operator characteristics curve [AUROC; see Prill et al. (2010)]. In DREAM5, only the top 100 000 interactions were considered for this analysis. The resulting AUPR and AUROC will be lower if only a subset of the interactions is considered. Although this difference has only little effect on the ranking of the approaches, we will report AUROC values for the top 100 000 predictions as well as for all predictions in order to enable the comparison to other studies [e.g. Narendra et al. 2011]. The performance evaluation in this article focuses on the AUROC, but additional evaluation and scores can be found in the Supplementary Material (part 4).

Expression compendia. In this study, we used three datasets provided by DREAM5 and two additional datasets from M3D [Faith et al. 2005]. All datasets consisted of several thousand genes and several hundred microarray measurement time courses. In comparison to data repositories such as GEO [Barrett et al. 2011], DREAM5 and M3D provide fewer but uniformly preprocessed and normalized datasets. Measurements as well as annotations are rendered comparable across different experiments and are thus suited to automated network inference.

In case of the real DREAM5 datasets, organism, experiment and gene names are replaced by random IDs to enable the evaluation of the inferred networks against experimentally confirmed interactions unknown to the participants. Thus, no prior knowledge could be utilized for the inference.

Datasets in the expression compendia are subdivided into experiments that consist of all microarrays described in a single publication or conducted by the same experimenter. Besides wild-type measurements, experimental conditions represent (combinations of) drug, environmental and gene perturbations. Some of the drug or environmental perturbations are provided as time course measurements. We considered each time point as a separate condition. A condition may contain multiple replicates. In case of gene perturbations (deletion or over-expression), the annotations provide the IDs of the perturbed genes. The artificial dataset was generated by the tool GeneNetWeaver [Marbach et al 2012; see Section 6 of the challenge description at http://wiki.c2b2.columbia.edu/dream/index.php/D5S4] and mimicked the E. coli dataset in the composition of the perturbations and time courses.

For additional validation, we obtained E. coli and S. cerevisiae expression data from the M3D database [see Table 1 and Faith et al. 2005]. Similar to the chip annotations provided in DREAM5, M3D provided manually curated metadata for their chip measurements.

Gold standard networks. Predicted E. coli interactions were validated based on RegulonDB [6] models conducted by large-scale chromatin immunoprecipitation (ChIP) binding assays. Physical binding is not a sufficient evidence, as noted by Iachuci et al. 2004 and Boulesteix and Strimmer 2009. Thus, ChIP will lead to many false positive interactions. M3D datasets aimed to overcome this problem by complementing ChIP assays with conservation-based motif discovery algorithms. Due to the more reliable small-scale assays and the manual curation, the E. coli gold standard should be regarded as more reliable than the one for S. cerevisiae. This is supported by a recent review by Narendra et al. 2011, where even otherwise accurate methods fail to predict this gold standard. The true network topology is known for the artificial dataset and was used for evaluation.

Combined with the above mentioned M3D datasets, these gold standard networks have been used for evaluating inference methods by DREAM5 as well as Faith et al. [2005] and Narendra et al. [2011]. Because the same gold standards are used and because the majority of the experiments in the DREAM5 datasets on E. coli were taken from the M3D database (Daniel Marbach, personal communication) results of DREAM5 and Narendra et al. [2011] are approximately comparable. This also applies to assessments based on artificial datasets that have been generated by the tool GeneNetWeaver [Marbach et al 2012] in both studies.

2.2 Network inference

Fold changes. Basal gene levels can be very different between experiments. To compensate for this, we transformed the absolute expression values into expression fold changes (see also Supplementary Material, part 1). Fold changes are computed by mapping the measurements m, i.e. |m| of each condition to one of more valid control conditions (Fig 1). Each m is subject to a combination of gene, drug or environmental perturbations P. A condition met measured at time point t under the set of perturbations P=P’ represents the differential treatment between two conditions.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>[TF]</th>
<th>[Genes]</th>
<th>[TF] pert.</th>
<th>[Genes] pert.</th>
<th>[Chips]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Artificial</td>
<td>195</td>
<td>1643</td>
<td>38</td>
<td>38</td>
<td>805</td>
</tr>
<tr>
<td>E. coli</td>
<td>334</td>
<td>4511</td>
<td>20</td>
<td>43</td>
<td>805</td>
</tr>
<tr>
<td>E. coli</td>
<td>167</td>
<td>4297</td>
<td>17</td>
<td>67</td>
<td>907</td>
</tr>
<tr>
<td>S. cerevisiae</td>
<td>333</td>
<td>5950</td>
<td>5</td>
<td>14</td>
<td>536</td>
</tr>
<tr>
<td>S. cerevisiae</td>
<td>156</td>
<td>6572</td>
<td>11</td>
<td>37</td>
<td>904</td>
</tr>
</tbody>
</table>

Shown is the size of the examined datasets as well as the number of measurements subject to gene specific perturbations (gene over-expressions and deletions).
We compute log-fold changes by mapping each measured condition \( m \) to one or more control conditions (replicated measurements) from the same experiment. A control may have fewer drug or gene perturbations than the corresponding measurements, but not more. In the example shown, both conditions \( a \) and \( b \) have fewer perturbations than \( m \) and are valid control conditions. Then, the replicates of the controls are averaged and (here: a total of 6) fold changes are computed for the replicates in \( m \) against the means of the selected controls, log-fold changes are computed as differences as measurements are already log-transformed. For the given application, the resulting geometric mean performs similar to an average (not shown).

For instance, the DREAM5 E. coli dataset consisted of 805 chip measurements of 487 different experimental conditions. Among the 805 chips we selected controls for 599 chips (corresponding to 379 conditions). Due to the multiplicity of measurement-control combinations, 935-fold changes were computed from the 599 chips.

Relevance networks. Our network inference approach is based on the estimation of the relevance of candidate interactions ([34, 35]). Candidate interactions, i.e. pairs of a TF and a TG, are ranked by a score \( s \). The score \( s \) can be any measure of dependency between the expression of the TF and its TG. Frequently used measures of dependency are based on Pearson’s or Spearman’s correlation coefficients or mutual information. In this article, we propose to use the score \( s^2 \) that is introduced below. The application of \( s^2 \) in the relevance network framework will be referred to as the \( s^2 \) approach. In addition to relevance networks, we also evaluate \( s^2 \) in the context of the C3NET [36, 37] frameworks (Supplementary Material, part 4).

A measure of association derived from two-way ANOVA. Our inference approach is based on a two-way ANOVA. A two-way ANOVA can be used to model experimental observations \( Y_{ijk} \) as responses to two factors \( C \) and \( G \) as well as the measurement error, \( \epsilon_{ijk} \):

\[
Y_{ijk} = \mu + C_i + G_j + \epsilon_{ijk}
\]  

where \( \mu \) is the average response, \( \gamma \) is the effect of the \( i \)-th level of the factor \( C \), \( \beta_j \) is the effect from the \( j \)-th level of factor \( \gamma \), and \( \epsilon_{ijk} \) represents the remaining unexplained error in replicate \( k \). In our application of ANOVA, \( C \) models the effect of differential expression across \( j \in \{1, \ldots, c\} \) different experimental conditions and \( G \) models whether the expression profiles of the genes \( j \in \{1, \ldots, g\} \) (as we consider exactly one TF and one TG) differ. Thus, we apply ANOVA to a matrix of conditions, genes and replicates as depicted in Figure 1. A two-way ANOVA tests three null hypotheses: (i) no differences in means of factor \( C \), (ii) no differences in means of factor \( G \), and (iii) no interaction between \( C \) and \( G \), by partitioning the total sum of squares \( SS_{Y} \) into four components (Fig. 1).

\[
SS_Y = SS_C + SS_G + SS_{CG} + SS_{err}
\]  

A sum of squares (SS) is a sum of squared deviations from a mean \( \bar{Y} \) and can be regarded as an unadjusted measure of dispersion.

Fig. 1. Transformation of absolute expression values into fold changes. We compute log-fold changes by mapping each measured condition \( m \) to one or more control conditions (replicated measurements) from the same experiment. A control may have fewer drug or gene perturbations than the corresponding measurements, but not more. In the example shown, both conditions \( a \) and \( b \) have fewer perturbations than \( m \) and are valid control conditions. Then, the replicates of the controls are averaged and (here: a total of 6) fold changes are computed for the replicates in \( m \) against the means of the selected controls, log-fold changes are computed as differences as measurements are already log-transformed. For the given application, the resulting geometric mean performs similar to an average (not shown).

Fig. 2. Sum of squares and the two-way ANOVA. A two-way ANOVA analyzes two dimensions or effects (here: \( C \) for conditions and \( G \) for genes) by partitioning the SS into four components: \( SS_T = SS_C + SS_G + SS_{CG} + SS_{err} \). The first example (panel B) exhibits strong associations between TF and TG. Here, \( SS_C \) is high as there is strong differential expression between the conditions. In panel C, the genes exhibit strong differences so \( SS_G \) will be high. If the two effects are linked (panel D), i.e. differential expression across conditions occurs only if strong differences are exhibited between both genes, \( SS_{CG} \) will be high. A high replicate variance leads to a high \( SS_{err} \) (panel E).

A variance \( \nu_2 \) is computed by adjusting the \( SS_C \) for the degree of freedom \( df_2 \), where \( df_2 \) is the number of data points under consideration minus 1, and \( x \in \{C, G, CG, err\} \). An \( F \)-value is computed by weighting the effect variance against the error variance [Equation (3)]. \( F \)-values follow the \( F \)-statistic, which can be used to derive the statistical significance of the involved factors as \( p \)-values. For instance, to estimate the significance of differential expression across conditions we compute \( F_C \) by:

\[
F_C = \frac{SS_C}{SS_{err}}
\]  

Effects so far describe differences, but ANOVA can also be used to detect specific similarities or associations between TF and TG. Phrased in terms of the two-way ANOVA, the strength of an association is proportional to the fraction of \( SS_C \) in the total sum of squares \( SS_Y \). Hence, \( \eta^2 = \frac{SS_{CG}}{SS_T} \) refers to \( \eta^2 \) as the non-parametric non-linear correlation coefficient. Its statistical significance can be estimated via \( F_{eta} \). A more detailed description of the algorithm including pseudo code can be found in the Supplementary Material (parts 2 and 3).

Adjusting for negative correlation. In contrast to Pearson’s \( r^2 \), \( \eta^2 \) does not directly test for negative correlations. We therefore propose to reverse the signs of the TF-fold changes to compute an additional \( \eta^2 \). The final ranking of candidate interactions is performed using \( \eta^2 = \max(\eta^2, \eta^2) \).
Incorporation of gene perturbation experiments. We extend the basic approach to incorporate measurements on gene specific perturbations. A candidate interaction between a TF and a putative TG should be considered of high interest, as it involves interactions that are specific to the condition of interest. The datasets (Table 1) have been estimated based on M3D data prior to our participation in DREAM. The weight of such an interaction is increased by inserting $w_2 = 1$ additional copies into the ANOVA matrix (Fig 2A). We also noticed that ~50% of the regulation in the artificial data is due to inhibition. Inhibition, i.e. negative correlations between TF and TG is comparatively rare in the real datasets (Fig 3). Run time complexity. The run time complexity of our network inference approach is $O(|TF| \times |genes| \times |chips|)$ (see Supplementary Material, part 3), i.e. the complexity of the ANOVA estimator is linear in the number of chips. The datasets (Table 4) required between 280 k and 2 M evaluations of $\eta^2$ across 160 to 907 chips. The largest datasets required a run time of 2 min on a single processor core. The complexity of other inference approaches has been discussed previously [Narendra et al. 2011].

Performance evaluation. Table 4 and Figure 4 show the performance of our $\eta^2$ method in comparison to other approaches. For a broad comparison of methods we combined our own evaluation results for some of the publicly available methods (for additional evaluation and scores see Supplementary Material, part 4) with the results of the DREAM5 network inference challenge as well as the large comparative assessment study of Narendra et al. 2011. We selected methods that performed best on one of the three DREAM5 datasets (methods 1–3) and the best performing methods (with respect to AUROC) as determined by Narendra et al. 2011 (methods 6–9). For comparison, we also applied the methods 2 as well as 5–7 as end-user ready tools were available.

In order to render the DREAM evaluation (that considered only the top 100 k predictions) comparable to the evaluation by Narendra et al. 2011, we re-computed the performance based on all predictions (i.e. not only the top 100 k) and re-applied publicly available methods. Considering all predictions usually increases the resulting AUROC by up to a few percentage points, which usually has only little effect on the performance ranking of the methods.

Interestingly, some of the participants in the DREAM5 challenge outperformed existing inference methods significantly, particularly in case of the artificial and E. coli datasets (Fig 3). For the inference of yeast interactions such a clear statement was not possible as all methods performed rather poorly. The presented $\eta^2$ method was the best performer for inferring E. coli interactions and was also competitive for the artificial dataset, outperforming the previously published methods analyzed in this article or by Narendra et al. 2011. The performance on artificial and E. coli interactions is depicted in Figure 3 for all methods that participated in the DREAM5 network inference challenge.

The prediction performance apparently also depends on the number of chip experiments. All methods yielded a better prediction performance for the M3D E. coli dataset (Table 4), which contains more measurements than the DREAM5 E. coli dataset. The same reasons also contribute to the lower performance observed in the S. cerevisiae dataset.

Properties of $\eta^2$ exemplified via selected interactions. We analyzed properties of $\eta^2$ on the E. coli dataset from M3D. A strong linear correlation is exhibited for instance by the fix:dusB interaction (fix: organization and maintenance of nucleoid structure; dusB: tRNA-dihydrouridine synthase B; Fig 5A). The observed linear correlation might be due to the fact that both genes are part of the same operon. Non-linear correlations such as gadE:hdeA are also detected by $\eta^2$ (gadE: acid-induced positive regulator of glutamate-dependent acid resistance; hdeA: stress response acid resistance protein). hdeA is already activated by low gadE concentrations (Fig 5B). In contrast, mdtE (multidrug transporter component) is activated only at high concentrations of gadX (regulator of acid resistance) resulting in an upwardly-curved scatterplot (not shown).

$\eta^2$ also enables the detection of correlations that are only apparent in a subset of the measured conditions, i.e. it detects local correlations. This increased sensitivity is due to the
Table 2. Performance of selected methods on the DREAM5 and M3D datasets.

<table>
<thead>
<tr>
<th>Methods</th>
<th>References (abbrev.)</th>
<th>Artificial</th>
<th>E. coli</th>
<th>S. cerevisiae</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANOVA</td>
<td>This article</td>
<td>DS:100k 74.6</td>
<td>D5:100k 79.8</td>
<td>Nar2011 53.9</td>
</tr>
<tr>
<td>Genie3</td>
<td>Hoyhul Tha et al. (2010)</td>
<td>81.5</td>
<td>67.4</td>
<td>75.7</td>
</tr>
<tr>
<td>Team 395</td>
<td>unpublished</td>
<td>78.6</td>
<td>60.2</td>
<td>61.7</td>
</tr>
<tr>
<td>Pearson’s r²</td>
<td>Butte and Kohane (1999)</td>
<td>75.3</td>
<td>61.2</td>
<td>51.6</td>
</tr>
<tr>
<td>CLR</td>
<td>Faith et al. (2007)</td>
<td>76.2</td>
<td>66.1</td>
<td>50.9</td>
</tr>
<tr>
<td>ARACNe</td>
<td>Margolin et al. (2006)</td>
<td>76.3</td>
<td>64.2</td>
<td>50.4</td>
</tr>
<tr>
<td>qgr graphs</td>
<td>Castello and Roverato (2009)</td>
<td>69.6</td>
<td>63.5</td>
<td>45.4</td>
</tr>
<tr>
<td>GeneNet</td>
<td>Oppen-Rhein and Steimner (2007)</td>
<td>52.4</td>
<td>65.5</td>
<td>55.2</td>
</tr>
</tbody>
</table>

The area under the ROC curve (AUROC) is used for the evaluation of inference methods performed by the DREAM5 organizers. The best predictions are shown in bold. See Supplementary Material (part 4) for additional scores. All methods were invoked with the designated options to utilize the preselected different. Shown are histograms of the correlation of non-interacting and interacting gene pairs as well as gene pairs regulated by the same set of TFs.

In vivo confirmation of novel interactions. Novel candidate interactions in E. coli were preselected by applying a 50% precision cutoff to the predictions, i.e. we stop iterating over the list of predictions from most to least confident when the precision evaluated against RegulonDB drops <50%.

Predicted TF:TG interactions were tested by quantifying the presence of the TG mRNA through qPCR amplification in E. coli gene knockouts of the corresponding TF. The mRNA levels for the same TG were quantified in non-mutant, wild-type E. coli to measure gene expression differences. Expression differences >2-fold for TGS are considered evidence for a true regulatory relationship between the predicted TF:TG pair.

In total, 5 TFS were sampled and 53 interactions not contained in the gold standard were tested. Here, 21 TF:TG pairs showed greater than a 2-fold change corresponding to a confirmation rate of 39.6%. Relaxing the fold change cutoff to 1.8, 26 pairs are reported (precision of 49.1%). This approximately confirms the 50% precision cutoff from the computational analysis. At a precision cutoff of 50% we predict 1995 novel interactions thus expecting ~1000 (1995 × 0.97) additional true interactions not contained in RegulonDB. The qPCR experiments were performed by the lab of James J. Collins at the Boston University in the context of the DREAM5 challenge. The full description and analysis of these interactions as well as the participating inference approaches will be the subject of a future paper (Marbach, D., Costello, J. and Küffner, R., et al. The wisdom of crowds for gene network inference, submitted).

Fig. 4. Differences between artificial and real datasets. The correlation distributions of artificial, E. coli and S. cerevisiae data expression data are quite different. Shown are histograms of the correlation of non-interacting and interacting gene pairs as well as gene pairs regulated by the same set of TFs. In contrast to artificial and E. coli data, correlation between a TF and a TG is not a good indicator of a true regulatory relationship in S. cerevisiae.
4 DISCUSSION

To infer gene regulatory networks (GRNs), we rank the relevance of candidate relationships consisting of a TF and a TG by measuring the dependency between their respective expression profiles.

For the detection of dependencies we proposed the measure \( \eta^2 \) that is derived from an analysis of variance (ANOVA). To our knowledge, \( \eta^2 \) has not been applied to network inference or to other problems in bioinformatics although it has a number of interesting features (Fig. 6). Like Pearson’s \( \rho^2 \), but in contrast to Bayes conditional probability tables or mutual information, \( \eta^2 \) does not require the discretization of the input data. This increases the robustness of our method as inappropriate discretization might lead to loss of signal. In contrast to Pearson’s linear correlation coefficient, \( \eta^2 \) is a non-parametric, non-linear correlation coefficient. It also detects local correlations that are only apparent in a subset of the measured conditions. This increased sensitivity is due to the effective utilization of replicated measurements to model the measurement error.

The recent DREAM5 blind assessment solicited the prediction of GRNs with thousands of genes from two real datasets (E. coli and S. cerevisiae) and one artificial dataset. The 29 participating teams employed a variety of methods based on regression (Lasso, random forests), Bayesian networks, mutual information and correlation. In DREAM5, our approach was rated the best performer on the inference of real networks and the second best performer on real and artificial networks combined. Especially for the inference of E. coli interactions, our approach performed significantly better than the methods evaluated in DREAM5 (Fig. 5) as well as in the large assessment of Narendra et al. (2011) (Table 2).

In contrast to E. coli, predictions for S. cerevisiae received significantly lower scores because the yeast gold standard network is less reliable. Compared with E. coli and artificial networks, inference is substantially more difficult in S. cerevisiae as here the expression of TF and their regulated genes is hardly correlated (Fig. 4). Indeed, with an AUC between 0.49 and 0.54, predictions were hardly better than guessing. The difficulty of network inference in S. cerevisiae has also been recognized by Hu et al. (2007) and Narendra et al. (2011). Many publications on network inference approaches solely focus on E. coli (Faith et al. 2007; Mordelet and Vert 2008).

Some of the known E. coli interactions identified by our approach were quite interesting biologically. For instance, an interaction between multiple antibiotic resistance genes was active after antibiotic treatment (local correlation) but not in growth phase experiments. According to qPCR experiments that were performed as part of the DREAM5 conference >50% of our novel predictions represent true interactions. At a precision of 50% we thus expect that our predictions contain 1000 previously unobserved true interactions.
R. Küffner et al.

Funding: P.T. and T.P. are partially funded by the DFG (RTG 1563/1 RECESS and Z.I. 6163 (CLA, respectively). L.W. is partially funded by the Helmholtz Alliance on Systems Biology, Project CoReNe.

Conflict of Interest: none declared.

REFERENCES


