

Systems biology

Inferring the experimental design for accurate gene regulatory network inference

Deniz Seçilmiş ^{1,†}, Thomas Hillerton^{1,†}, Sven Nelander² and Erik L. L. Sonnhammer^{1,*}

¹Department of Biochemistry and Biophysics, Science for Life Laboratory, Stockholm University, Solna 17121, Sweden and
²Department of Immunology, Genetics and Pathology and Science for Life Laboratory, Uppsala University, SE-75185 Uppsala, Sweden

*To whom correspondence should be addressed.

[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Associate Editor: Alfonso Valencia

Received on November 2, 2020; revised on March 29, 2021; editorial decision on May 6, 2021; accepted on May 11, 2021

Abstract

Motivation: Accurate inference of gene regulatory interactions is of importance for understanding the mechanisms of underlying biological processes. For gene expression data gathered from targeted perturbations, gene regulatory network (GRN) inference methods that use the perturbation design are the top performing methods. However, the connection between the perturbation design and gene expression can be obfuscated due to problems, such as experimental noise or off-target effects, limiting the methods' ability to reconstruct the true GRN.

Results: In this study, we propose an algorithm, IDEMAX, to infer the effective perturbation design from gene expression data in order to eliminate the potential risk of fitting a disconnected perturbation design to gene expression. We applied IDEMAX to synthetic data from two different data generation tools, GeneNetWeaver and GeneSPIDER, and assessed its effect on the experiment design matrix as well as the accuracy of the GRN inference, followed by application to a real dataset. The results show that our approach consistently improves the accuracy of GRN inference compared to using the intended perturbation design when much of the signal is hidden by noise, which is often the case for real data.

Availability and implementation: <https://bitbucket.org/sonnhammergri/ideMAX>.

Contact: erik.sonnhammer@dbb.su.se

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

1 Introduction

Gene regulatory interactions control many of the *in vivo* biochemical mechanisms, and thus play a key role in most processes in living organisms. Interruptions of these mechanisms can result in several diseases including cancers (Emmert-Streib *et al.*, 2014; Price *et al.*, 2010; Sonawane *et al.*, 2019). Therefore, identification of these regulatory interactions spearheads the way to understand the nature of the genetic diseases and ultimately cure them. The interactions between regulators and their targets form a system called a gene regulatory network (GRN), and the underlying mechanism of the system can be revealed by the accurate inference of these GRNs. For this reason, GRNs can be considered as a key factor in understanding and separating the pathological mechanisms from physiological. Through this understanding, GRNs can also be used to directly propose targets for potential treatments.

Inference of GRNs can be performed from gene expression data where each gene's expression is altered by knockdown or

overexpression experiments, such as via shRNAs, siRNAs or small molecules i.e. drugs. Such experiments are generally referred to as gene perturbations. There exists a variety of GRN inference methods, some of which require knowledge of the targets of the perturbations as input, and some that do not. The most popular examples of the first type, requiring known perturbations, include LASSO (Friedman *et al.*, 2010; Tibshirani, 1996), least squares with cut-off (Tjärnberg *et al.*, 2013) and ridge regression (Friedman *et al.*, 2010). Popular examples of the latter type, not needing known perturbations, include GENIE3 (Huynh-Thu *et al.*, 2010), ARACNe (Margolin *et al.*, 2006) and Context Likelihood of Relatedness (CLR) (Faith *et al.*, 2007). It has previously been shown that methods of the first type are able to achieve perfect GRN inference accuracy under good conditions (Tjärnberg *et al.*, 2017, 2015), which is an advantage over the latter type whose performance remained limited in previous benchmarks (Greenfield *et al.*, 2010; Guo *et al.*, 2016; Marbach *et al.*, 2012; Schaffter *et al.*, 2011). However, a

potential problem for the first type of methods is that the design of the perturbation, even if targeted at single genes and in principle known, may not be representative. This can be due to either random noise or biases from experimental error, such as off-target effects, both potentially confounding the true effect of the perturbation in the gene expression data. This can lead to suboptimal GRN inference when the method fails to fit the experimental perturbation design to the noisy gene expression. In more technical terms this problem can be attributed to regression dilution bias. A dilution bias is a common problem in linear regression when there is too much random noise between the observed effect and the predictor, here gene expression and the perturbation. The effect of this bias is that the fitted values are incorrectly forced toward zero, which for GRN inference means losing out on explanatory edges (Rennolls, 1990). Regression dilution is very common in biological fields that use linear regression to explain their observed effect, yet it is often not accounted for, which leads to poorer performance of the models (Hutcheon et al., 2010).

In this study, to overcome the aforementioned problem with design-utilizing GRN inference methods, we developed an algorithm that infers the perturbation from the gene expression data. It is called IDEMAX for ‘Infer DESign MATriX’. The method works by capturing the alteration of a gene’s expression in relation to its overall distribution when compared over multiple experiments. The inferred perturbation design matrix is then used as input to the GRN inference method, replacing the intended one. We show that the method can correctly predict the majority of the perturbations for cases with low noise. Although it tends to yield a rather different perturbation design matrix for cases with high noise, this still substantially improves accuracy of the GRN inference compared to when using the intended perturbation design matrix.

2 Materials and methods

2.1 Algorithm

The IDEMAX perturbation design matrix inference method, as applied here, assumes that a known number of replicates of gene perturbations have been performed for each gene, as this is a common setup. The method finds the expression values that are the most different from the rest for a given gene, and considers these as the experiments where the gene was perturbed. It statistically tests each fold change expression value against the distribution of all the other expression values of the gene using a Z-score approach (Eq. 1). The approach is inspired by previous work that have shown that a Z-score approach excels at outlier detection in distributions (Cousineau and Chartier, 2010; Misra et al., 2020; Shiffler, 1988). The highest absolute Z-score value among all is considered to be the most different gene expression, suggesting that the gene was

perturbed in that position. Depending on the number of the replicates per gene, this many top absolute Z-scores are used to identify the experiments where a gene was perturbed. By doing this for each gene in the expression matrix, IDEMAX is able to identify the perturbation matrix P . P is a sparse matrix of the same size as the input expression data with n non-zero values in each row, where n is the requested number of replicates for each gene. To capture the effect of the perturbation the sign of the Z-score is assigned to the corresponding cell in the inferred P matrix, where -1 indicates knock-down/knockout and $+1$ overexpression perturbation.

$$Z_i^{(j)} = \frac{x_i^{(j)} - \mu_i^{(j)}}{\sigma_i^{(j)}}, i = 1, \dots, N, j = 1, \dots, M \quad (1)$$

In Equation 1, $Z_i^{(j)}$ refers to the Z-score value of the i^{th} gene for the j^{th} experiment. The Z-score is calculated over the mean (μ) and standard deviation (σ) of all the other measurements of the i^{th} gene excluding the j^{th} one, which is denoted as $(!j)$. Here N is the number of genes and M the number of experiments.

By assigning the specified number of perturbations to each gene, we make sure that this number is preserved within the gene. However, we introduce a potential problem where some experiments might be assigned to multiple target genes and some experiments may not be assigned to any target genes in a set of single gene perturbation experiments. While this does occur (Supplementary Figs. 1 and 7), the results show that this potential drawback does not have a negative effect on the GRN inference accuracy, as they are either the same or improved. Therefore, such situations are allowed in the pipeline, and no optimizations are done. Even though the inferred P matrix is not strictly one experiment—one target, it is meant to better correspond to the effective perturbations.

2.2 Synthetic network and data generation

Throughout the process of developing and maturing our approach, we used synthetic datasets that are connected to an intended design matrix and a true GRN, which is required to measure the accuracy of the predictions. For the generation of these synthetic networks and datasets, we used both the GeneNetWeaver (GNW) network and data generation tool (Schaffter et al., 2011), and the GeneSPIDER Matlab toolbox (Tjärnberg et al., 2017). We generated five different networks for each size of 100, 150 and 200 genes with GNW, and 100, 250 and 500 genes with GeneSPIDER. Note that the difference in sizes of the networks is due to using different network and data generation tools. This selection was made to ensure that at least 50% of all genes are regulators, which in GNW limits the size to around 200. As GeneSPIDER does not have this limitation, it allows benchmarking on larger datasets. For GNW the true networks come from the *Escherichia coli* network where at

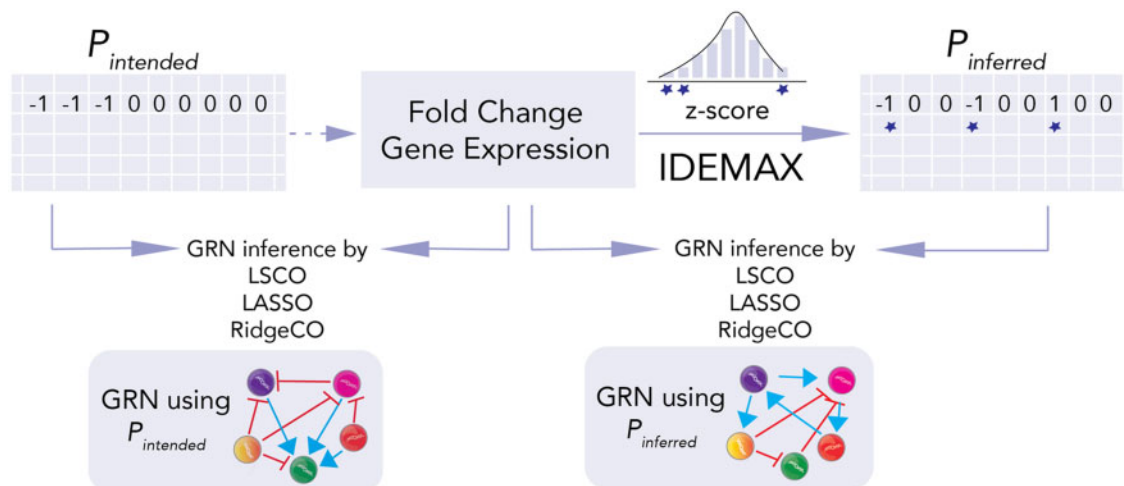


Fig. 1. The workflow of the IDEMAX P matrix inference algorithm

least half of the number of genes are set to be transcription factors, and the datasets are generated from a stochastic model followed by the addition of Gaussian noise. Due to the limitation of not having biological replicates through GNW, we generated three datasets from the same true network and merged those as if they were replicates of the same dataset. The true GeneSPIDER networks are generated from a scale-free topology with ~ 3 links per gene on average where unlike GNW each gene is considered a potential regulator, and the datasets are generated by applying a linear model where the negative pseudoinverse of the true network is multiplied by the P matrix with three replicates per gene, and Gaussian noise was added. Datasets from both GNW and GeneSPIDER were separated into two categories: data with higher variance and data with lower variance. For a detailed list of the parameters used in the network and data generation with both GNW and GeneSPIDER, see [Supplementary Notes S1 and S2](#).

2.3 Performance evaluation of the inference

After the P matrix was inferred, prediction accuracy was calculated in two ways: (i) the global true positive rate (TPR_{Global}) of the number of correctly predicted perturbations over the number of total perturbations in the intended design (Eq. 2a). Here this measurement also corresponds to precision and F1-score because the same amount of predictions is made as there are intended perturbations, causing the number of false positives to equal the number of false negatives. (ii) The one-replicate true positive rate (TPR_{Onerep}) of the number of cases where at least one perturbation of the gene's replicates is correctly predicted, averaged over the total number of genes (Eq. 2b). The inferred P matrices were compared to the intended ones one-by-one for each single cell in the matrix, in terms of both TPR_{Global} and TPR_{Onerep} .

$$TPR_{Global} = \frac{\sum_{j=1}^M TP_{(j)}}{\sum_{j=1}^M TP_{(j)} + \sum_{j=1}^M FN_{(j)}} \quad (2a)$$

$$TPR_{Onerep} = \frac{1}{N} \sum_{i=1}^N TP_{(i)}, TP_{(i)} = \begin{cases} 1, & \text{if } \sum_{w=1}^W TP_{(w)} \geq 1; \\ 0, & \text{else} \end{cases} \quad (2b)$$

In Equation 2a, M corresponds to the number of experiments, in other words number of total perturbations, and TP corresponds to the number of correctly predicted perturbations while FN refers to not predicting a perturbation when it exists in the intended P matrix. In Equation 2b, N is the number of genes; $TP_{(i)}$ is a binary answer that is '1' when at least one perturbation out of W replicates of one gene's perturbation is correctly predicted and '0' when no perturbation out of W is correctly predicted. Overall, TPR_{Global} corresponds to the total number of correctly predicted perturbations divided by the total number of experimental perturbations, and TPR_{Onerep} denotes the number of cases where at least one perturbation is correctly predicted out of the number of gene's replicates divided by the number of genes.

2.4 GRN inference methods

Five different GRN inference methods were used in this study, three design-utilizing methods: least squares with cut-off (LSCO), LASSO and ridge regression with cut-off (RidgeCO) (Tibshirani, 1996; Tjarnberg et al., 2013), and two testing methods: Genie3 and Context Likelihood of Relatedness (CLR) (Faith et al., 2007; Huynh-Thu et al., 2010). For all methods, the wrappers available in the GeneSPIDER Matlab toolbox were used. GRN inference for all methods was performed by inferring 20 networks to fulfill a set of different sparsities ranging from full-to-empty for each single run, and the inference accuracy was calculated by comparing these networks to the true GRN and representing each as a point in the ROC and precision-recall curves.

2.5 Performance evaluation of the GRN inference

To investigate the effect of the inferred P matrix on the accuracy of the GRN inference, GRN inference was performed on three different

versions of the generated synthetic data: (i) the data with the intended perturbation design, (ii) the data with a broken connection between the gene expression and its perturbation design, (iii) the data using the perturbation design matrix inferred by IDEMAX. The accuracy of all inferences with the two different perturbation designs and the random control were calculated in terms of the area under the receiver operating characteristics curve (AUROC) and the area under the precision-recall curve (AUPR) values in comparison to a known true network.

2.6 Significance of the GRN inference accuracy

In order to test whether the improvement in GRN inference accuracy achieved by using the perturbation design inferred by IDEMAX is significant, we performed an unpaired two-samples two-sided Wilcoxon test, and calculated the P -values in 95% confidence intervals. Selection of a non-parametric test was made due to the small sample size, that is 15 observation points for combinations of three GRN inference methods and five datasets. The significance testing was performed between intended and inferred, intended and random, and inferred and random perturbations for both AUROC and AUPR. The P -values are given in Table 1 for the high variance data and in Supplementary Table S1 for the low variance data.

2.7 Perturbation inference on real data

IDEMAX was applied to a biological dataset for a nine-gene subnetwork of the SOS pathway in *E.coli* (Gardner et al., 2003), followed by GRN inference and accuracy calculation. The data were collected through overexpression experiments, and an experimentally supported GRN for the dataset is available, allowing us both to assess the overlap between the intended and inferred perturbations and to calculate the accuracy of GRN inference. The same measurements as above, AUROC and AUPR, were used for the accuracy evaluation, and the comparison between the intended and inferred perturbations was made using the heatmap visualization allowed by the small data sizes.

2.8 Perturbation inference on data from DREAM5

IDEMAX was applied to subsets of DREAM5 *in silico* and *E.coli* datasets consisting of knockout and overexpression perturbations on single and multiple target experiments. The accuracy of the perturbation design inference was measured with the same global and one-replicate true positive rates as before, and accuracy of GRN inference was evaluated in terms of AUPR in order to allow for a comparison to the original challenge publication (Marbach et al., 2012). Subset data collection pipeline and analysis results are given in Supplementary Section S4.

3 Results

The performance of the method for inferring the design matrix P was benchmarked in two ways: first by evaluating how well it can reconstruct the intended P matrix, and second by measuring its effect on accuracy of GRN inference. Both benchmarks were performed for two data categories, namely with higher and lower variance.

The first benchmark uses synthetic data and calculates the true positive rates (TPRs) of links in the inferred P matrix relative to the intended P as specified in Equations 2a and 2b. Datasets of three different sizes for GeneNetWeaver (100, 150 and 200 genes) and three different sizes for GeneSPIDER (100, 250 and 500 genes) were benchmarked.

For the lower variance data, the overlap between the intended and inferred P matrices is high, reaching a TPR near 1, indicating that the intended P matrix is well represented by the data and that the inferred P successfully captures this when the noise level is low (Supplementary Fig. S8). For the higher variance data however, the global true positive rates for all sizes in the GNW and GeneSPIDER data (Fig. 2) are below 0.07, meaning that most perturbations are inferred differently than in the intended P . The one-replicate TPRs are higher, yet stay below 0.09 for GNW data and 0.19 for GeneSPIDER data. This still means that more than 80% of the gene expression did not correspond to the intended perturbations,

Table 1. Significance of difference in AUROC and AUPR between the inferred, intended and random GRN predictions for high variance data

Data generation tool	Number of genes	AUROC			AUPR		
		Intended versus inferred	Intended versus random	Inferred versus random	Intended versus inferred	Intended versus random	Inferred versus random
GNW	100	0.00*	0.27	0.00*	0.00*	0.16	0.00*
	150	0.00*	0.00*	0.00*	0.00*	0.16	0.00*
	200	0.00*	0.05	0.00*	0.00*	0.41	0.00*
GeneSPIDER	100	0.00*	0.19	0.00*	0.00*	0.20	0.00*
	250	0.00*	0.00*	0.00*	0.00*	0.00*	0.00*
	500	0.00*	0.00*	0.00*	0.00*	0.00*	0.00*

Note: *P*-values were obtained from the two-tailed unpaired Wilcoxon test with $\alpha = 0.05$.

*Statistical significance.

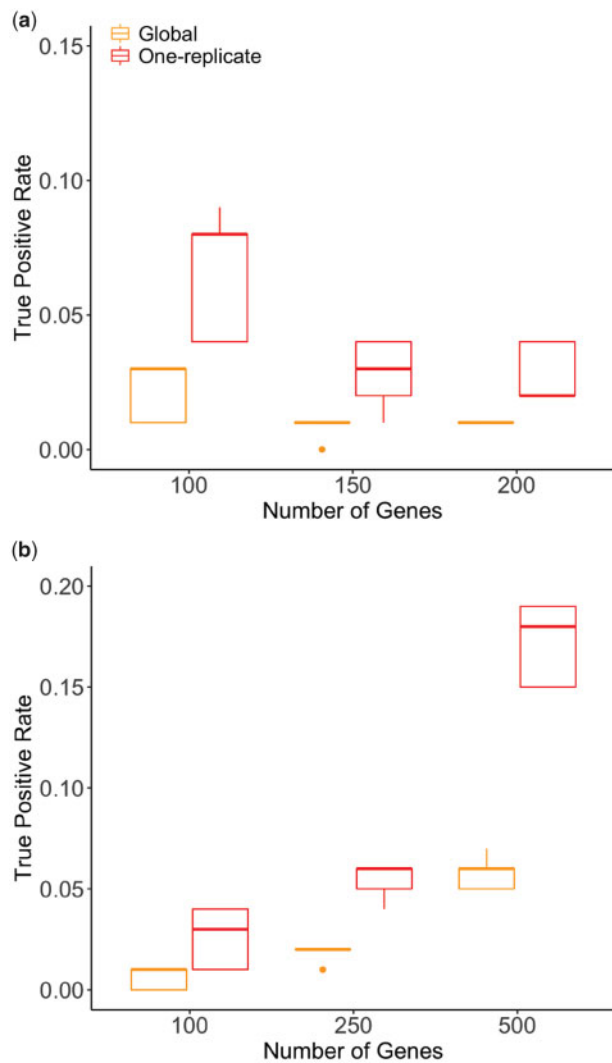


Fig. 2. Comparison between the intended and inferred perturbations. True positive rates using data from (a) GeneNetWeaver and (b) GeneSPIDER datasets were calculated either globally for all replicates or when one correctly predicted replicate was considered sufficient

indicating that the intended *P* matrix is poorly represented by the data and that the inferred *P* captures other information.

The second benchmark measured the effect on accuracy of GRN inference that inferring the *P* matrix has, compared to the intended *P* matrix. For this purpose the same simulated datasets from GNW

and GeneSPIDER were used. In order to provide a control, we also inferred GRNs from datasets where the connection between the gene expression and perturbation design was broken by random shuffling, and calculated the inference accuracy. GRN inference accuracy was measured in terms of the area under the receiver-operating-characteristic and precision-recall curves (AUROC and AUPR, respectively).

The results show that despite a very low overlap between intended and inferred *P* matrices, the inferred *P* matrix gives substantially higher accuracy of GRN inference for all datasets (Fig. 3). The largest improvement is seen for the 100-gene GeneSPIDER data where the median AUROC increased from 0.50 to 0.64 and the median AUPR from 0.05 to 0.33. The AUROC and AUPR values differ significantly between GRN inference using intended and inferred *P* matrices, as well as between inferred and random *P* matrices in all cases (Table 1). Between intended and random *P* matrices, AUROC was not significant for 100- and 200-gene datasets from GNW and for the 100-gene datasets from GeneSPIDER data while significant for other sizes, and AUPR values were only significant for the 250- and 500-gene GeneSPIDER datasets.

GNW and GeneSPIDER are fundamentally different from each other in their data generation approach, resulting in different data properties, such as signal-to-noise ratio, variance and condition number (Fig. 4 and Supplementary Fig. S6 for the higher and lower variance datasets, respectively). Despite these differences, our method identified a *P* matrix that yielded more accurate GRNs than the intended *P* for data from both generators. As expected, for low variance data where the inferred and intended *P* are similar, the GRN inference accuracy was also similar (Supplementary Fig. S9). The accuracy levels are approximately the same for low and high variance with IDEMAX, whereas they drop substantially when using the intended *P* for the high variance data. Therefore, the variance appears to be the main determining factor for the observed improvement by IDEMAX as it can improve the accuracy of GRN inference from data with high variance up to the accuracy level seen for low variance data.

We further applied IDEMAX to a public experimental dataset for the nine-gene subnetwork of the SOS pathway in *E.coli* (Gardner et al., 2003). Here it identified almost the same *P* matrix as the intended one (Supplementary Fig. S10). Noteworthy, a single change in the position of a perturbation in the *P* matrix caused more true links to be captured, resulting in a clear improvement in the accuracy of the inferred GRN in terms of AUROC and AUPR (Fig. 5). The results indicate that even a slight alteration in the perturbation design matrix can lead to a considerable improvement in the following GRN inference.

3.1 Application to DREAM5 data

The analyses on the DREAM5 subsets supported the results presented here as for the *in silico* subset data a similar perturbation matrix was inferred by IDEMAX which led to a similar GRN inference accuracy, and for the *E.coli* subset data IDEMAX inferred a rather

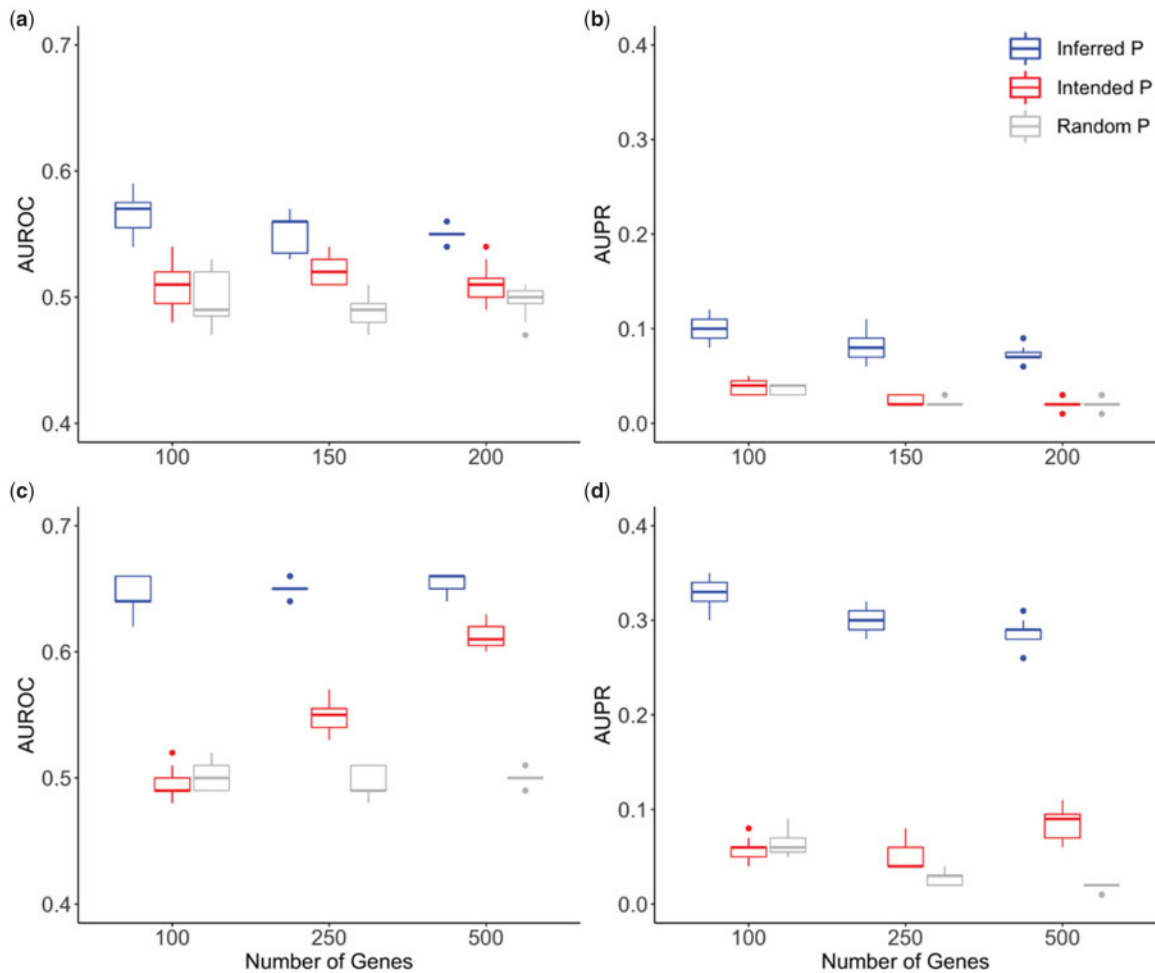


Fig. 3. Accuracy of GRN inference. (a) AUROC and (b) AUPR on GeneNetWeaver data and (c) AUROC and (d) AUPR on GeneSPIDER data. Each box contains combined values from three inference methods: least squares with cut-off (LSCO), LASSO and RidgeCO, and 5 different datasets for a total of 15 observations. Individual performances of the inference methods are given in [Supplementary Figures S2–S4](#)

different perturbation design resulting in improved GRN inference accuracy ([Supplementary Fig. S11](#)).

We investigated potential biological reasons behind the differences between the intended and inferred perturbation matrices on the DREAM5 *E.coli* subset. Of the 66 gene pairs where the intended and inferred perturbation designs had different target genes for an experiment, 8% were in the DREAM5 gold standard ($P < 2.2 \times 10^{-16}$), suggesting that some perturbations may bleed over to coupled genes in the system that display a stronger effect at steady state than the intended target. This could happen if the intended target is under stronger homeostatic control via feedback mechanisms ([Supplementary Note S4.2](#)).

3.2 Application of GRN inference methods not requiring a P matrix

To test whether a GRN inference method that does not require knowledge of the perturbation can result in higher accuracy than was achieved by methods that do, we inferred GRNs using Genie3 and Context Likelihood of Relatedness (CLR), using the high variance GeneNetWeaver and GeneSPIDER datasets. The results in terms of AUROC and AUPR are given in [Supplementary Figure S5](#). It can be seen that these methods are outperformed by IDEMAX. The low performance can in part be explained by the existence of

selfloops in the true GRNs, while selfloops are not predicted by these methods, causing false negatives.

4 Discussion

The perturbation design information is used in many GRN inference methods, some of which have been previously shown to perform highly accurately ([Tjärnberg et al., 2017](#)).

For real biological data, the perturbation design is often thought to be known, especially if the perturbations are performed through knockdown/knockout or overexpression experiments. However, experimental noise that masks the perturbation effect as well as off-target effects of the perturbations can break the connection between the intended perturbation design and the measured gene expression, which might introduce an obstacle in inferring the underlying GRNs. Another reason that the inferred design can be different from the intended is that the perturbation may bleed over to coupled genes in the system that display a stronger effect than the intended target, for instance if the intended target is under stronger homeostatic control. We have found evidence that this happens, see [Supplementary Note S4.2](#). In the presence of high noise, other GRN inference methods which do not require knowledge of the perturbation design may also fail to identify any accurate GRNs. To alleviate this situation, we developed a P matrix inference method,

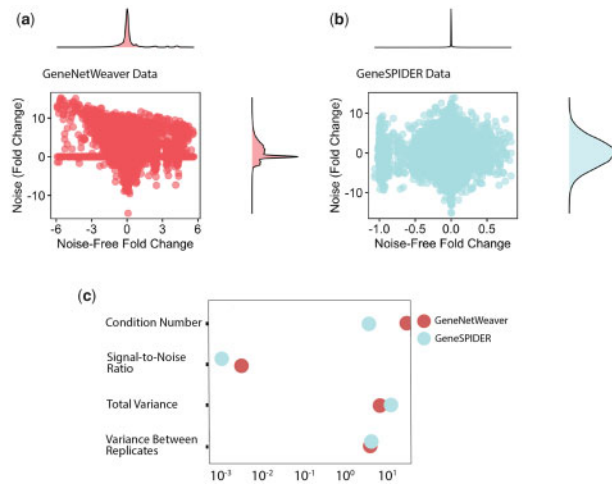


Fig. 4. Distributional characteristics of the 100-gene GeneNetWeaver (GNW) and GeneSPIDER datasets with higher variance in the main article file in comparison. (a) Noise-free gene expression versus noise both in base 2 logarithm of the fold change for the GeneNetWeaver data and (b) noise-free gene expression versus noise both in base 2 logarithm of the fold change for the GeneSPIDER data. (c) Properties of the GNW and GeneSPIDER datasets (fold change gene expression data including noise). Signal-to-Noise Ratio was calculated according to (Tjärnberg et al., 2017). Total variance refers to the variance of the gene expression matrix as a whole, whereas variance between replicates is the mean value of all intra-replicate variances

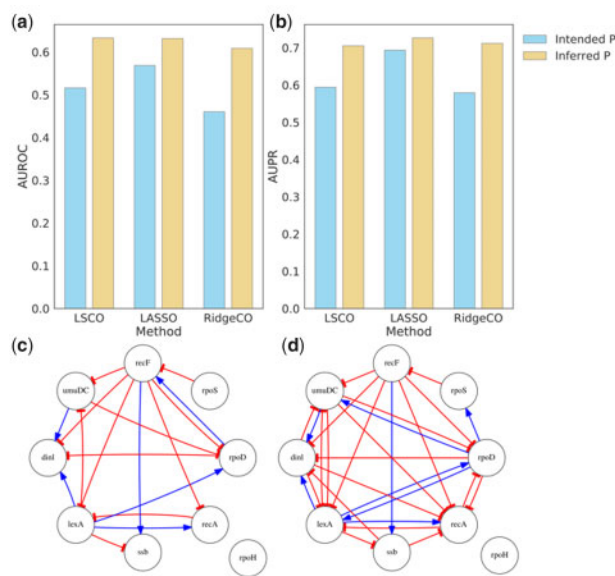


Fig. 5. Application to biological data. GRN inference accuracy in terms of (a) AUROC and (b) AUPR using intended and inferred P matrices and gene expression data for the nine-gene subnetwork of the SOS pathway in *E.coli*. To visualize the improvement by IDEMAX, true positives in GRNs inferred by RidgeCO are shown using the (c) intended and (d) inferred P matrices

IDEMAX, based on a Z-score approach, detecting a predefined number of perturbations for each gene in the columns where its fold change diverges the most from the distribution of the remaining genes in the system.

The inferred design matrix P is meant to replace the intended P matrix as an input to design-utilizing GRN inference methods, such as least squares, LASSO and ridge regression. The IDEMAX design matrix bypasses potential shortcomings of the intended perturbation design which may not be appropriate for noisy or biased gene expression data, and allows the GRN inference methods to more accurately reconstruct the underlying GRNs. IDEMAX can be used by any design-utilizing

GRN inference method, but unfortunately not by methods, such as Genie3 that do not take the design matrix into account.

We note that IDEMAX does not assume normality when applying Z-scores, but the Z-scores are only used for ranking and not for any statistical significance analysis. Potential alternatives to our approach include using median instead of the mean, or the absolute distance to the median under jackknifing. However, preliminary testing suggested that the taken approach was most favorable.

Also note that IDEMAX may result in multiple targets being assigned to one experiment while some experiments are not assigned to any targets. Despite this possibility, the assigned number of targets per experiment is typically low, with less than 3% of the experiments in the used data assigned to more than 3 target genes. As this means that the P matrices inferred by IDEMAX are largely realistic, we did not introduce any restrictions to force the inferred P matrix to be one target per experiment, especially given that drug perturbations rarely target single genes.

In the presence of high variance, IDEMAX yields P matrices with low overlap to the intended P matrix. Even though this could suggest a potential failure of the proposed algorithm, the situation is in fact the opposite, as a low overlap between the inferred and intended perturbations is needed for improving the GRN inference. Due to the high noise level in the fold change gene expression that corresponds to the dependent variable in the regression model, the intended perturbation design does not constitute a well enough fit to the data, resulting in suboptimal GRN inference. Therefore, in the presence of high noise levels, finding a different P matrix relates to the success of the method rather than a potential failure. When the variance was lower and the regression dilution situation is no longer relevant for such data, IDEMAX had much higher overlap with the intended perturbations, resulting in similar GRN inference accuracies with no statistically significant difference. For high variance data, IDEMAX yields a perturbation design better fitting the underlying gene expression data than the intended design, meaning that low overlap between the two is welcome and beneficial for overcoming any potential drawback that the regression dilution may introduce, as for improved GRN accuracy.

A key finding here was that the accuracy of perturbation inference remained very similar between the two data generation tools despite the differences in the data properties used for this study. This supports the generality of IDEMAX as it was capable of inferring significantly more accurate GRNs regardless of the source and properties of the data. The only data property that caused substantial differences across datasets was the variance.

The application of IDEMAX to experimental data for the SOS pathway added further support to the algorithm as a slightly different P matrix was inferred and a clear improvement was observed in the inferred GRN's accuracy in terms of both AUROC and AUPR. Even though the dataset was much smaller than the synthetic datasets, it illustrates what a large effect a single perturbation design change can cause. It should be highlighted that in a gene regulatory network, genes and their interactions are dependent on each other, allowing a single change in the position of a perturbation to recover many more true positives than before.

The biases introduced by the experimental noise or any other kind of experimental artefacts can be overcome by IDEMAX, and the inferred P matrix can improve the accuracy of the GRN inference significantly when used as an input to a design-utilizing GRN inference method instead of the intended P matrix. In conclusion, given that high noise levels is one of the biggest obstacles in the inference of accurate GRNs and that real data usually comes with a high level of experimental noise, IDEMAX introduces a welcome advance to the field by ameliorating this situation.

Funding

This work was supported by the Swedish Foundation for Strategic Research.

Conflict of Interest: none declared.

References

- Cousineau,D. and Chartier,S. (2010) Outliers detection and treatment: a review. *Int. J. Psychol. Res.*, **3**, 58–67.
- Emmert-Streib,F. *et al.* (2014) Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks. *Front. Cell Dev. Biol.*, **2**, 38.
- Faith,J.J. *et al.* (2007) Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol.*, **5**, e8.
- Friedman,J. *et al.* (2010) Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.*, **33**, 1–22.
- Gardner,T.S. *et al.* (2003) Inferring genetic networks and identifying compound mode of action via expression profiling. *Science*, **301**, 102–105.
- Greenfield,A. *et al.* (2010) DREAM4: combining genetic and dynamic information to identify biological networks and dynamical models. *PLoS One*, **5**, e13397.
- Guo,S. *et al.* (2016) Gene regulatory network inference using PLS-based methods. *BMC Bioinformatics*, **17**, 545.
- Hutcheon,J.A. *et al.* (2010) Random measurement error and regression dilution bias. *BMJ*, **340**, c2289.
- Huynh-Thu,V.A. *et al.* (2010) Inferring regulatory networks from expression data using tree-based methods. *PLoS One*, **5**, e12776.
- Marbach,D., DREAM5 Consortium. *et al.* (2012) Wisdom of crowds for robust gene network inference. *Nat. Methods*, **9**, 796–804.
- Margolin,A.A. *et al.* (2006) ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics*, **7**, S7.
- Misra,S. *et al.* (2019) Unsupervised outlier detection techniques for well logs and geophysical data. *Mach. Learn. Subsurface Charact.*, **1**.
- Price,N.D. *et al.* (2010) Systems biology and systems medicine. *Essent. Genomic Person. Med.*, 131–141. Academic Press.
- Rennolls,K. (1990) Correction for regression dilution bias. *Lancet*, **335**, 1534.
- Schaffter,T. *et al.* (2011) GeneNetWeaver: in silico benchmark generation and performance profiling of network inference methods. *Bioinformatics*, **27**, 2263–2270.
- Shiffler,R.E. (1988) Maximum Z scores and outliers. *Am. Stat.*, **42**, 79.
- Sonawane,A.R. *et al.* (2019) Network medicine in the age of biomedical big data. *Front. Genet.*, **10**, 294.
- Tibshirani,R. (1996) Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. Series B Stat. Methodol.*, **58**, 267–288.
- Tjärnberg,A. *et al.* (2015) Avoiding pitfalls in L1-regularised inference of gene networks. *Mol. Biosyst.*, **11**, 287–296.
- Tjärnberg,A. *et al.* (2017) GeneSPIDER – gene regulatory network inference benchmarking with controlled network and data properties. *Mol. Biosyst.*, **13**, 1304–1312.
- Tjärnberg,A. *et al.* (2013) Optimal sparsity criteria for network inference. *J. Comput. Biol.*, **20**, 398–408.