Simulating gene silencing through intervention analysis

Vera Djordjilović,

University of Oslo, Norway

Monica Chiogna

University of Bologna, Italy

and Chiara Romualdi

University of Padova, Italy

Summary. We propose a novel method for simulating the effects of gene silencing. Our approach combines relevant subject matter information provided by biological pathways with gene expression levels measured in regular conditions to predict the behaviour of the system after one of the genes has been silenced. We achieve this by modelling gene silencing as an external intervention in a causal graphical model. To account for the uncertainty that is associated with the structure learning of the graphical model, we adopt a bootstrap approach. We illustrate our proposal on a *Drosophila melanogaster* gene silencing experiment.

Keywords: Biological pathways; Gene knock-down; Gene silencing; Graphical models; Intervention; Structure learning

1. Motivation

Motivation for this work comes from the field of ribonucleic acid (RNA) interference, which is a biological process in which RNA molecules suppress or inhibit gene expression, causing the targeted gene to be turned off. Discovery of this process and its regulatory potential by Fire *et al.* (1998) opened up previously inaccessible areas of research. RNA interference can be introduced experimentally to silence genes of interest: a process named *gene silencing*. Scientists routinely use this powerful method to study functions of specific genes by analysing phenotypic effects of their silencing, also in view of potential development of new therapeutic agents. Not sooner than in August 2018 did US regulators approve the first therapy based on RNA interference (Ledford, 2018). This approval is likely to jump-start further efforts in RNA-interference-based research.

Despite this potential, some questions regarding the design and application of RNA interference, as both a research tool and a therapeutic strategy, still remain. Experiments are expensive and time consuming and the design of silencing experiments, which are optimal with respect to specific targeting, might require complex adaptive procedures. Considerable savings could be

made if potential effects of silencing could be investigated before physically performing it, to engineer experiments in an optimal way. This is the pressing motivation for the current work.

We argue that predicting the effects of gene silencing is, in some circumstances, possible by using careful statistical modelling. In this work, we propose a novel combination of consolidated and new statistical tools aimed at anticipating the output of a gene silencing experiment. The ingredients that have been combined share some common elements with tools that are used in causal inference and with approaches that are used in the field of structure learning of biological networks.

The outline of the paper is as follows. In Section 2, we cover the rationale behind our proposal. In Section 3, we introduce some background material on the biological knowledge that is used in the modelling process. In Section 4, we describe the statistical framework for gene silencing, and in Section 5 we outline our proposed approach. In Section 6, we validate our proposal by analysing data from the *Drosophila melanogaster* experiment of gene silencing. In Section 7, we illustrate some of the practical difficulties that are encountered in the experimental process of gene silencing and we show how our method can be used to signal the presence of some of the undesired experimental artefacts. Some tentative conclusions, limitations and directions for future research are discussed in Section 8.

The data that are analysed in the paper and the programs that were used to analyse them can be obtained from

```
https://rss.onlinelibrary.wiley.com/hub/journal/14679876/series-
c-datasets.
```

2. The rationale

Gene silencing experiments are employed by researchers for studying gene function, and for the development of therapeutics for diseases such as cancer, infectious diseases and neurodegenerative disorders. Evidence deriving from gene silencing experiments, the so-called *intervention data*, offer causal explanations (Moffatt, 2016); in some cases, it is used along with *observational data* to infer gene regulatory processes (Hauser and Bühlmann, 2012; Rau *et al.*, 2013; Cho *et al.*, 2016).

Questions that motivate gene silencing experiments resemble questions that are asked in most studies in the health, social and behavioural sciences. For example, studying the effect of lowering the level of a gene on a given phenotype pursues aims that are similar to those related to exploring the efficacy of a given drug in a given population. In both cases, a *causal* question is seeking an answer. Differently from experimental research, in empirical research, answers to these questions are more complex to give: they cannot be based on data alone but require some knowledge of the data-generating process. Despite these difficulties, the conceptual framework and algorithmic tools that are needed for addressing such questions are well established (see, for example, Pearl (2000) for a comprehensive overview).

We argue that tools to address causal questions in empirical research could be put into practical use also in the context of gene silencing experiments with the aim of simulating, i.e. predicting at an observational level, the results of an experiment without physically performing it. Such predictions could find natural application in planning future gene silencing experiments. Moreover, they could provide challenging information in active learning of biological networks (Sverchkov and Craven, 2017), where additional data, usually coming from *ad hoc* experiments, are collected to improve models under construction. But many other uses of such information are easy to foresee.
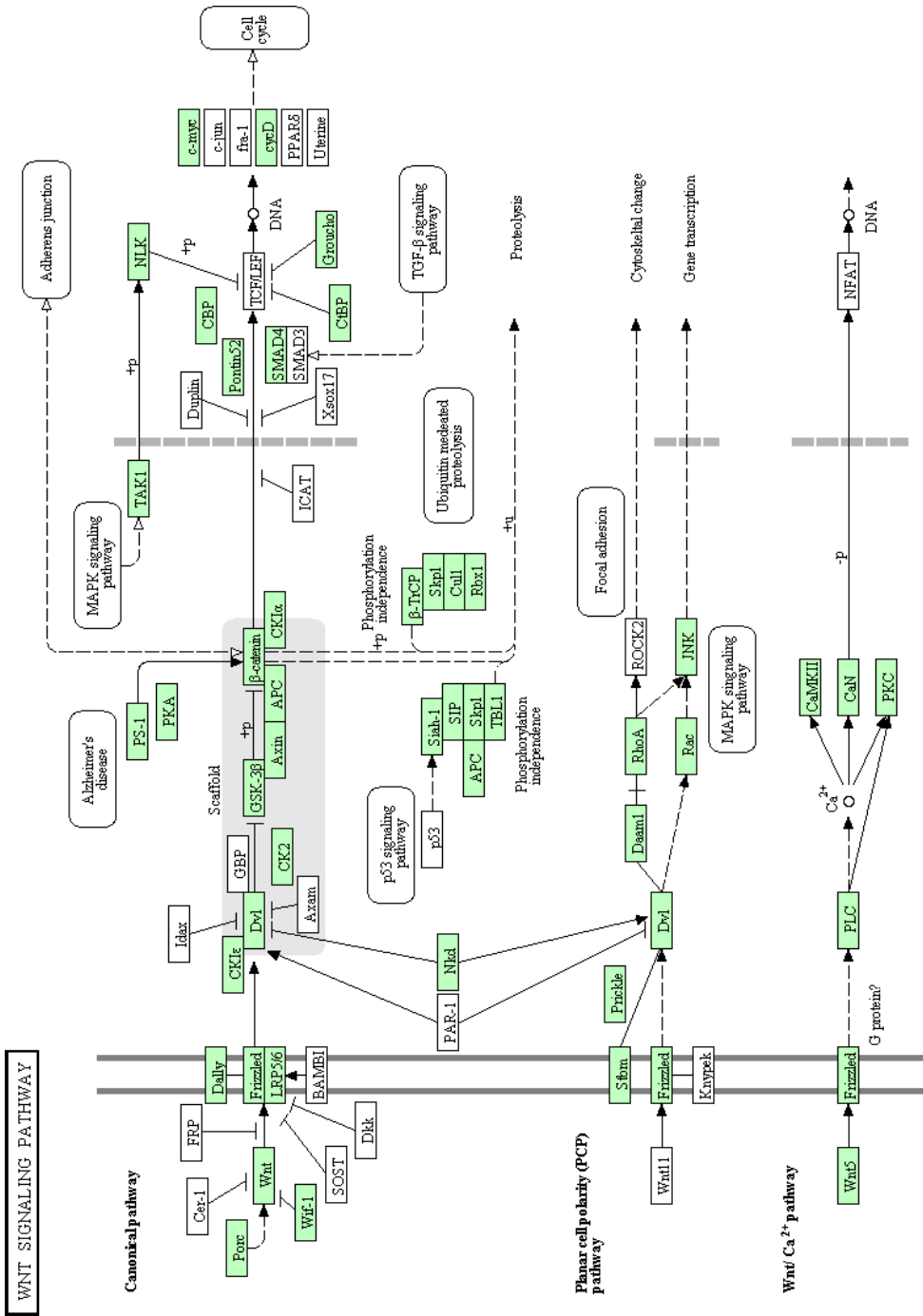
Fig. 1.   WNT signalling pathway taken from the *Kyoto Encyclopedia of Genes and Genomes* pathway repository

To predict the effects of silencing, we propose to combine relevant biological knowledge with the gene expression data that are measured in controls, which, in our case, are the observational data. In principle, biological knowledge comes in various forms, but in this work we focus on the information that is stored in biological pathways. Biological pathways are elaborate schematic representations of cellular processes (Fig. 1). We translate the structural information of the pathway into a directed acyclic graph (DAG), which depicts functional relationships between genes under study. Gene silencing can then be modelled as an external intervention in a graphical model (Pearl, 2000; Lauritzen, 2001). Nevertheless, numerous processes taking place in a cell at any given moment activate different pathways and, as a result, a DAG based on a single pathway is typically not a sufficiently good model for this purpose. For this reason, we use the observational expression data to find a DAG that better describes the gene set under study. In particular, we extract the information on the topological ordering of genes from the pathway DAG, and then we infer the graphical structure from the observational data under the imposed ordering. Finally, we use the bootstrap to assess the uncertainty arising from the fact that the intervention analysis is based on an estimated, rather than fixed, graphical model.

3.   The biological knowledge

Fig. 1 represents an example of a biological pathway, the WNT signalling pathway in *Drosophila melanogaster*, retrieved from the *Kyoto Encyclopedia of Genes and Genomes* (KEGG) pathway repository (Kanehisa and Goto, 2000). It is composed of edges and nodes with the following meaning. Rectangles represent gene products, mostly proteins, but also RNA and complexes. The edges between rectangles represent functional interactions. Circles represent other types of molecules, mostly chemical compounds, whereas the large white rectangles are links to other pathways. See `https://www.genome.jp/kegg/` for a more detailed description.

Pathways show sequences of interactions ensuring maintenance and control of the flow of information, energy and biochemical compounds in a cell. They are considered the best representation of the experimentally validated knowledge regarding cellular processes. In fact, the annotation of a biological pathway is the result of an extensive effort of hundreds of researchers who manually codify their experimental findings about a specific biological process into a graphical representation.

Incorporating structural information that is provided by pathways in the statistical analysis is common in the third generation of pathway analysis methods, namely, topology-based pathway enrichment analysis, and it has been considered by several popular approaches to detecting differentially expressed genes between two conditions (see for instance Shojaie and Michailidis (2009), Vaske *et al.* (2010) and Haynes *et al.* (2013)). The key hypothesis is that including such external information can improve the power and interpretability of an employed statistical procedure. Following the same line, we consider pathways as the source of subject matter information to be used together with the observational data when building a model for predicting the effects of gene silencing.

4.   Gene silencing from a statistical perspective

Let V be a set of p genes under study, and let X_V denote the associated expression levels. Let f denote the distribution of X_V and assume that there is a DAG $G = (V, E)$, $E \subset V \times V$, such that f is Markov with respect to G, i.e. f can be factorized as

$$f(x_V) = \prod_{v \in V} f(x_v | x_{\mathrm{pa}(v)}),$$

where pa(v) is a set of parents of v in G (we refer readers who are less familiar with graph terminology to Lauritzen (1996)).

Gene silencing can be perceived as an external intervention on X_V that forces the expression of a targeted gene towards a desired value. Intuitively, we expect this intervention to affect the distribution of X_V. The nature and the extent of this effect are in general impossible to predict when the causal relationships between variables are unknown. We can, however, say more about the distribution of the system after the intervention if we further assume that G is causal for V. The graph G is said to be causal for X_V, if X_V is Markov with respect ot G, and it further holds for any $A \subset V$ that

$$f(x_{V \setminus A} \| x_A^*) = \left. \prod_{v \in V \setminus A} f(x_v | x_{\mathrm{pa}(v)}) \right|_{x_A = x_A^*} = \left. \frac{f(x_V)}{\prod_{a \in A} f(x_a^* | x_{\mathrm{pa}(a)})} \right|_{x_A = x_A^*}, \tag{1}$$

where the notation $f(x \| x_A^*)$ denotes the post-interventional distribution, i.e. the distribution of X after manipulating X_A, forcing it to assume the value x_A^*. Formula (1) is usually referred to as the *intervention formula* and can be found in various forms in Pearl (2000) and Spirtes *et al.* (2000). This is one possible definition of a causal DAG; see Spirtes *et al.* (2000) for a definition in terms of direct causes.

It is important to stress the difference between two types of conditioning featured in expression (1). Conditioning by intervention, $f(x_u \| x_v^*)$, describes how the distribution of X_u should be modified if X_v has been modified by external forces and set to x_v^*. This is different from the conventional conditioning by observation, $f(x_u | x_v^*)$, which describes how the distribution of X_u should be modified if the value of X_v has been observed and equals x_v^*. The difference is best illustrated on a simple example; assume that the causal DAG for X_u and X_v is $u \to v$. Then an external intervention on X_v has no influence on X_u and thus $f(x_u \| x_v^*) = f(x_u)$; in contrast, when there is no external intervention, observing the value of X_v gives us information about X_u and thus modifies the distribution of X_u according to $f(x_u | x_v^*) = f(x_u, x_v^*) / f(x_v^*)$.

The value of the intervention formula becomes clear once we understand that it provides a recipe for determining the effects of interventions by expressing the post-intervention distribution of the system in terms of the preintervention distribution. In expression (1), the intervention manipulates a target variable and sets its value to a constant; the intervention formula can be easily adapted to more general types of interventions, including so-called *soft* interventions (Markowetz *et al.*, 2005).

5. Pipeline for simulating gene silencing

Within the framework that was described in the previous section, the ingredients that are needed for predicting the effects of gene silencing are two: qualitative information provided by G and quantitative information provided by f. We estimate these quantities by using pathway information together with the observational expression data and then perform intervention analysis. Our approach can be summarized in the following three steps:

step 1, retrieval of the pathway information;
step 2, guided structure learning and estimation of the statistical model;
step 3, intervention analysis.

In what follows we discuss a possible way of performing these steps in a real data analysis.

5.1. Retrieval of the pathway information

Predicting the effects of gene silencing when the underlying DAG is known is straightforward:

we can use the observational data to estimate the preintervention distribution of the set of genes and then obtain the post-intervention distribution from the intervention formula (see Section 5.3 for details). The underlying DAG is typically constructed from subject matter knowledge. In this case, we can exploit information that is stored in repositories such as the KEEG. Often, information that is stored in the KEEG does not allow immediate extraction of a directed graph, as annotation foresees, for example, different types of relationships between variables and even loops. We refer readers to Djordjilović *et al.* (2015), who discussed the most critical issues related to conversion of pathways into graphical models. Here, we limit ourselves to the observation that it is always possible to convert pathways into DAGs. For this, some simple steps can be taken: since the edges ending with a bar represent inhibition, they can be interpreted as arrows, whereas the arrows annotated with +p and −p can be considered as undirected edges. Then, Wikipathways (Slenter *et al.*, 2017) can be used to derive the orientation of all undirected edges. Most of this preliminary work can be performed by the R package `graphite` (Sales *et al.*, 2018), which transforms KEGG pathways into graph objects.

5.2.   Guided structure learning and the estimation of the statistical model

DAGs that are obtained by simple pathway conversion do not always fit the data particularly well. Indeed, biological links between gene products do not necessarily correspond to statistically significant associations between the corresponding variables, whereas latent factors might induce associations that are not depicted in the pathway. As an example of the latter instance, consider compound-mediated interactions, i.e. interactions between two nodes for which another node, a chemical compound, acts as a bridge. As data for chemical compounds are not usually available, they are typically removed from the network. Removal is performed to avoid interruptions of the biological signal passing through the nodes, but graphs resulting from these manipulations do not necessarily reflect statistical properties of the distribution of the observed variables. An illustration is given in Fig. 2.

Suppose that data are generated according to the DAG in Fig. 2(a) and that compounds are not observed, so that pathway conversion calls for elimination of compounds. Routine conversion would lead to the DAG in Fig. 2(b). Such DAGs, however, would not offer an accurate representation of the observed quantities. In fact, according to the DAG in Fig. 2(b), B is independent of C given A. However, the distribution of the observed variables (A, B, C) will, because of the unobserved compounds, show an association between B and C even after conditioning on A, i.e. the DAG representing the distribution of (A, B, C) should include a link between B and C, or between C and B.

As in this work we aim at generating predictions at a measurable level, DAGs that are obtained by pathway conversion need to undergo a refinement. A radical solution would be to discard the pathway DAG and to estimate the graphical structure from data (see, for example, Maathuis *et al.* (2009, 2010)). However, it is widely recognized that this estimation problem, which is also known as structure learning, represents an extremely challenging statistical task. In such situations, including even vague prior information can significantly improve the estimation accuracy (Ideker *et al.*, 2011). Ma *et al.* (2016) pursued constrained network estimation where data are combined with the information that is provided by pathways. Following a similar strategy, which was first outlined in Djordjilović *et al.* (2017), we restrict attention to a topological ordering of the variables that are retrieved from the pathway DAG and reconstruct the network from the data on the basis of only such information. We denote this topological ordering by a p-dimensional vector o, such that o_k gives the kth node in the chosen topological ordering.

In graphical terminology, a topological ordering is a linear ordering of the nodes of G such
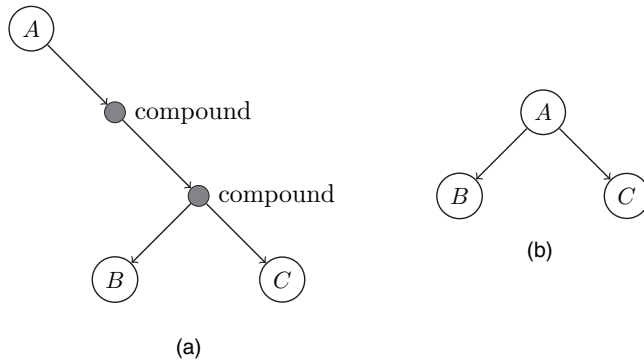
Fig. 2. Example of (a) compound-mediated interactions and (b) the DAG resulting from their elimination

that, for every directed edge $u \to v$, node u comes before v in the ordering. In general, topological ordering is not unique since a DAG induces only a partial ordering. When there are multiple orderings, none of which can be preferred on the basis of pathway or other relevant external information, we suggest considering different alternative orderings and simulate gene silencing under these alternative settings.

Given a single, fixed topological ordering, we propose to perform guided learning of the structure of the model, along the lines of the K2 algorithm (Cooper and Herskovits, 1992). By specifying the topological ordering, we respect the flow of relationships between genes, but we leave the data to highlight the significant associations.

The K2 algorithm belongs to the family of score-based structure learning algorithms, which search for a DAG that maximizes a given score function. The original K2 algorithm employs the greedy search strategy to find, among all DAGs with a given topological ordering, the DAG that maximizes the K2 score: a new scoring function proposed by Cooper and Herskovits (1992). The K2 score is defined for categorical variables, whereas gene expression is usually measured on a continuous scale. To avoid the loss of information that is associated with discretization of continuous measurements, we modified the K2 algorithm to allow for multivariate normal data. We refer to this proposal as CK2; see Appendix A for details.

After applying the CK2 algorithm, and obtaining an estimate of the underlying graph, $\hat{G}$, we pursue an estimate of the parameters of the joint distribution f. We exploit a definition of DAGs in terms of structural equations (Pearl, 2000):

$$X_v = \alpha_v + \beta_v^{\mathrm{T}} X_{\mathrm{pa}(v)} + \epsilon_v, \qquad v \in V, \qquad (2)$$

where $\epsilon_v \sim N(0, \sigma_v^2)$ is the random disturbance, β_v is the vector of regression coefficients, α_v is the intercept and pa(v) is the set of parents of v in $\hat{G}$. We estimate model parameters by estimating p nodewise linear models relating each node to the set of its parents. Typically, the size of the parent set is small, so β_v can be estimated by maximum likelihood. When the sample size is too small to allow for maximum likelihood estimation, the structure learning within CK2 and the estimation of the final DAG might call for some sort of regularization. In principle, various approaches can be employed for this. In Section 6, we describe a practical solution that we adopted in our study, which seems particularly convenient in the context at hand.

5.3. Intervention analysis

Let $X = X_o$ denote the vector of expression levels of the genes in V, where genes are ordered

according to the topological ordering o fixed in step 1. The matrix representation of model (2) is then

$$X = \alpha + BX + \epsilon,$$

where $\alpha = (\alpha_v)'_{v \in V}$ and $\epsilon = (\epsilon_v)'_{v \in V}$ are vectors that are ordered according to the same topological ordering, and each row of a $p \times p$ upper triangular matrix B contains a corresponding vector of parameters β_v from model (2). This model representation is particularly useful when investigating effects of interventions. The effect of silencing of gene v on gene u can be summarized by the causal effect δ_u (Maathuis *et al.*, 2009):

$$\delta_u = E(X_u \| X_v = \alpha + 1) - E(X_u \| X_v = \alpha)$$

comparing the expected level of X_u if the intervention had set the value of X_v to $\alpha + 1$ instead of α. Since we assume a multivariate normal distribution, δ_u is independent of α. If node v is kth in the given topological ordering, the vector of causal effects can be easily found through matrix algebra to be given by the kth row of the matrix $L = (I - B)^{-1}$. To obtain estimates of causal effects, it is sufficient to replace B by its estimate $\hat{B}$ that is obtained in step 2.

To assess the variability of the causal effect estimator, we resort to a bootstrap strategy. We sample with replacement a large number of samples of the size of the original sample and apply the CK2 algorithm to learn the graphical structure. We then estimate the matrix B conditionally on the learned structure. Finally, we compute the causal effect estimates at each bootstrap replication. As a result, we obtain the bootstrap distribution of the causal effect estimator that takes into account not only uncertainty related to parameter estimation—i.e. matrix B—but also the uncertainty that is related to the estimation of the graphical structure. The bootstrap distribution of silencing effects is a mixture of two components: one corresponding to a random variable degenerate at zero and the other to a random variable with a non-zero mean. The degenerate component of the distribution corresponds to the bootstrap replications in which no direct path between the silenced gene and the gene under consideration is estimated. In these cases, the silencing has no effect on the gene.

6. Biological validation: *Drosophila melanogaster* experiment

In this section, we apply the proposed approach for the prediction of effects of gene silencing to the data from the *Drosophila melanogaster* experiment. In this experiment run by the Department of Biology of the University of Padova (Italy), the naked cuticle gene *nkd* in a fruit fly (*Drosophila melanogaster*) was silenced. Data before and after silencing were measured on a number of statistical units with realtime polymerase chain reaction, which is a technology for detection and quantification of the messenger RNA expression that is typically considered a gold standard thanks to its control of the inherent noise. We refer the interested reader to Djordjilović (2015) for a description of the protocol of the experiment.

This experiment provides an ideal opportunity to access the performance of our approach, since it offers the possibility of comparing model-based predictions with observed effects of gene silencing. To do that, we build a statistical model using only observations from the control group and then compare our predictions with the actually observed changes that are seen in the knock-down group.

6.1. Data
The data consist of two sets of 14 observations of 12 genes participating in the WNT pathway:

Table 1.   Selected genes involved in the WNT pathway

Gene	Annotation	Description
dally	CG4974	Division abnormally delayed
por	CG6205	Porcupine
daam	CG14622	Dishevelled associated activator of morphogenesis
fz	CG16785	Frizzled 3
rho1	CG8416	Ras homologue gene family, member A
dco	CG2048	Discs overgrown or casein kinase 1, ε
rok	CG9774	ρ-associated kinase
sgg	CG2621	Shaggy or glycogen synthase kinase 3 β
arm	CG11579	Armadillo or catenin β 1
pont	CG4003	RuvB-like protein 1 (pontin 52)
nkd	CG11614	Naked cuticle

the first set corresponding to the treatment (silenced) group and the second set corresponding to the control (wild type) group. Genes that are involved in the pathway are listed in Table 1. The silenced gene is the naked cuticle gene *nkd*.

6.2.   Retrieval of the pathway information
We started from a DAG, representing a subset of the WNT pathway, manually curated by biologists on the basis of Fig. 1. We refer to this DAG as pathway DAG (Fig. 3(a)).

6.3.   Guided structural learning
We found a topological ordering of this DAG, using the `topological.sort` function of the `igraph` R package (Csardi and Nepusz, 2006). We pass the ordering obtained along with gene expression data of the control group to the CK2 algorithm.

As the sample size was limited, when estimating the parameters of the candidate models within CK2 we adopted the shrinking procedure of Schäfer and Strimmer (2005). We briefly review the procedure for the benefit of readers. Let R be the empirical correlation matrix for the set V of genes under study. We shrank R towards an identity matrix

$$R^* = \lambda_R I + (1 - \lambda_R) R,$$

where I is the $p \times p$ diagonal matrix and λ_R is the regularization parameter whose estimated optimal value is available analytically. We also shrank the p-dimensional vector of gene-specific variances towards the median empirical variance, so that the regularized estimate of the variance of gene v becomes

$$s_v^* = \lambda_v \tilde{s} + (1 - \lambda_v) s_v,$$

where $\tilde{s} = \mathrm{median}\{s_z\}_{z \in V}$ is the median empirical variance across p genes and s_v is the empirical variance of gene v. The optimal shrinkage intensity is also in this case estimated analytically (Opgen-Rhein and Strimmer, 2007). Estimates of the variance vector and the correlation matrix were combined to provide a regularized estimate of the variance matrix

$$V^* = \mathrm{diag}\{\sqrt{s_v^*}\}_{v \in V} R^* \mathrm{diag}\{\sqrt{s_v^*}\}_{v \in V}.$$

The procedures applied are available in the `corpcor` R package (Schafer *et al.*, 2017).
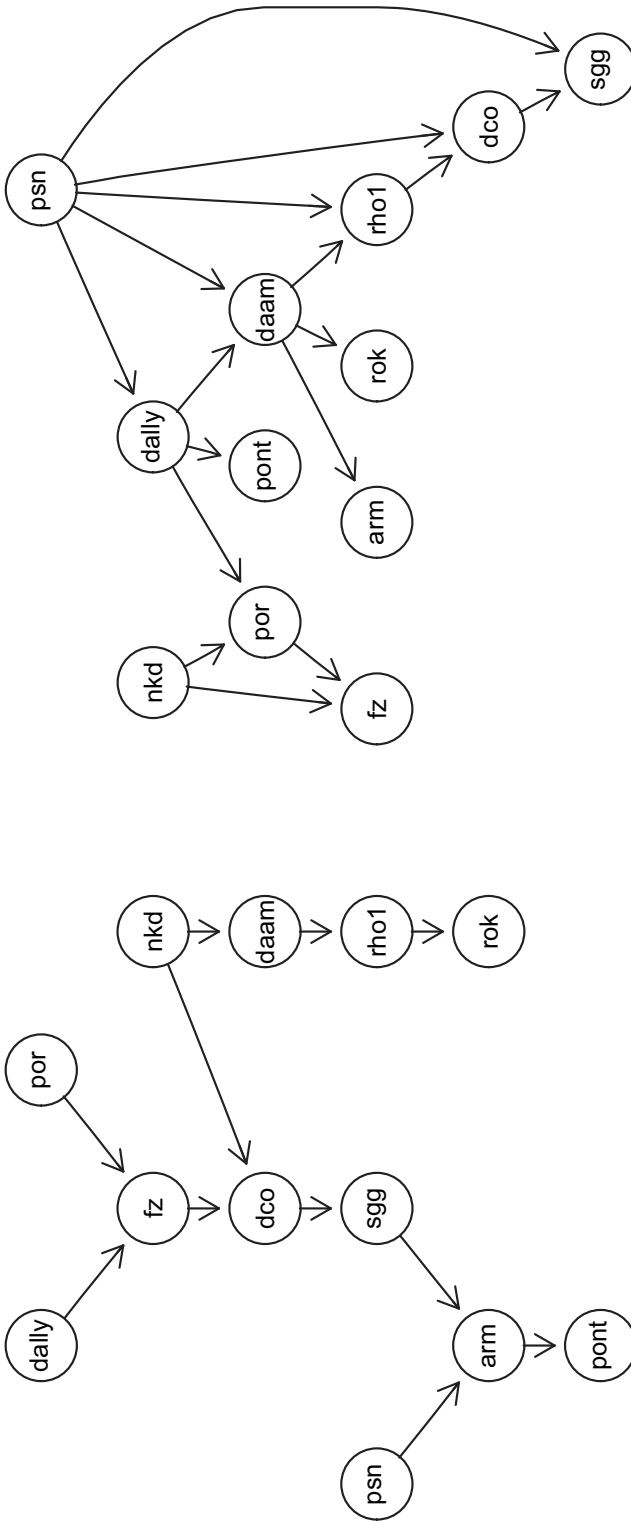
Fig. 3. (a) The pathway DAG and (b) the estimated DAG obtained by the CK2 algorithm

We used the regularized estimate V^* to estimate the parameters of the candidate models within CK2. When considering a parent set S for the node v, to estimate the parameter $\beta_{v,S}$ in

$$X_v = \alpha_{v,S} + \beta_{v,S}^{\mathrm{T}} X_S + \epsilon_{v,S},$$

we replaced the empirical covariance with its regularized version V^* to obtain

$$\hat{\beta}_{v,S} = (V_{S,S}^*)^{-1} V_{S,v}^*,$$

where $V_{A,B}^*$ denotes a submatrix of V^* given by subsets $A, B \subset V$. This estimate is then used to compute the fitted values and the scoring criterion.

We opted for this shrinkage procedure as opposed to regularization strategies based on penalized regression for two reasons. Firstly, with regard to the tuning of the regularization parameter, shrinking the covariance matrix is computationally fast and inexpensive because of the closed form expression for the shrinkage intensity. Secondly, shrinking the covariance matrix globally, as opposed to applying penalization in the local nodewise models, means that all the variables are treated on an equal footing and subjected to the same amount of penalization.

The resulting network is shown in Fig. 3(b). It is worth noting that the estimated graph appears quite different from the pathway DAG. This output, although compatible with a learning process in a noisy context with limited data, might also signal possible inaccuracies in the representation of molecular pathways. In this study, for example, a participant of the WNT pathway, *dally*, is itself regulated by the WNT pathway, so there is a feedback loop that is not depicted in the pathway DAG and not even admissible by our modelling framework. Such a violation is solved by the learning strategy by acting on the outward and inward edges of the node involved.

6.4. Intervention analysis

We estimated causal effects as described in Section 5.3. To obtain the distribution of the estimator, we considered 2000 bootstrap samples: to each of them we applied the CK2 algorithm, followed by parameter estimation and intervention calculus. The results are reported in Table 2. Here, we summarize the main results with respect to the goodness of predictions that were obtained by our strategy.

(a) According to our procedure, the percentage of samples in which the estimated causal effect was zero indicates that there are two genes with a non-zero estimated effect, i.e. *fz* and *por*, of which only the former is stably deemed in the bootstrap analysis. The sign and the size of the effects, if compared with that observed in the silenced group, appear to have been predicted well.

(b) Results of a two-sample t-test comparing the mean of each gene in the control condition with its mean after silencing (the last column of Table 2) show that *dally, fz* and *pont* were most affected by the intervention. Out of these three genes, model predictions capture well the effect on *fz*. For *dally* and *pont*, predictions of silencing effects did not correspond to that actually observed in the experiment: the size of the zero component of the bootstrap distribution, $\%(\delta_v = 0)$, is high; moreover, the actually observed effect d_v is outside the interval that is given by the quantiles of the non-zero component of the bootstrap distribution. This might reflect the feedback mechanisms that were previously discussed.

Table 2. Results of the procedure for simulating the effects of silencing the *nkd* gene†

| Gene	$\%(\delta_v = 0)$	$q_{0.025}$	$q_{0.975}$	d_v	p-value
dally	81.15	−2.09	−0.69	−10.27	1.07×10^{-11}
por	26.90	−0.98	−0.15	−0.02	9.10×10^{-1}
daam	68.90	−6.43	−1.83	1.21	3.13×10^{-1}
fz	4.30	−0.98	−0.14	−0.37	5.88×10^{-4}
rho1	79.10	−71.36	15.83	−3.45	8.16×10^{-1}
dco	79.80	−5.36	1.57	0.48	7.25×10^{-1}
rok	68.95	−4.92	−1.04	−2.13	2.51×10^{-2}
sgg	77.40	−6.03	−0.26	2.07	2.70×10^{-1}
arm	69.00	−45.56	−7.05	−0.48	9.58×10^{-1}
pont	32.50	−1.46	−0.09	−2.89	2.25×10^{-8}

†The name of the gene is given in the first column, the percentage of bootstrap samples in which the estimated causal effect was zero in the second column, the 2.5% and the 97.5% quantile of the non-zero component of the bootstrap distribution of the causal effect estimate in the third and fourth columns, the actually observed causal effect in the fifth column and the p-value for the test of equality of the mean expression in the silenced and the control condition in the last column.

7. The ideal experiment

Provided that the topological ordering correctly grasps the hierarchy of relationships between the variables of interest, a major strength of our approach is its ability to reproduce results that would be observed if an ideal silencing were performed. By ideal silencing we intend an experimental silencing that affects directly only its target and leaves the mechanism governing the distribution of the remaining genes unaltered. In practice, many technical artefacts might influence the specificity of the silencing experiment. Most notably, short interfering RNA designated to match and destroy the target gene transcript might match more than one gene and, as a consequence, induce undesired off-target effects. Several studies (see, for example, Jackson and Linsley (2004)) have reported experimental findings of non-specificity of RNA interference gene silencing. Qiu *et al.* (2005) reported that the probability of off-target effects can be considerable (from 5% to 80% across organisms), which prompted researchers to invest further efforts into increasing RNA interference specificity (Pei and Tuschl, 2006; Cullen, 2006).

When gene silencing is weakly specific and acts off target, the evidence resulting from the experiment is affected by spurious effects. We illustrate this point through a worked example focusing on the cAMP response element binding protein (CREB) knock-down in myeloid leukaemia cell lines.

Data from this study (which are available from the *Gene Expression Omnibus* portal as GS12056) contain measurements of 22410 genes in 10 samples from the control condition and 10 samples from the knock-down condition. For illustration, we selected a subset of genes pertaining to the KEGG prostate cancer pathway. Pathway DAG, which is shown in Fig. 4, contains 85 genes. CREB is represented by a dark grey node labelled 1385 (in this study Entrez identifiers are used instead of gene symbols). Fig. 4 suggests that, in the CREB experiment, specificity of the knock-down is in question. In Fig. 4, genes with a Bonferroni adjusted p-value for the two-sample Wilcoxon test (control *versus* knock-down) below $\alpha = 0.05$ are highlighted. These can be considered to be significantly affected by the knock-down. In this example, four of
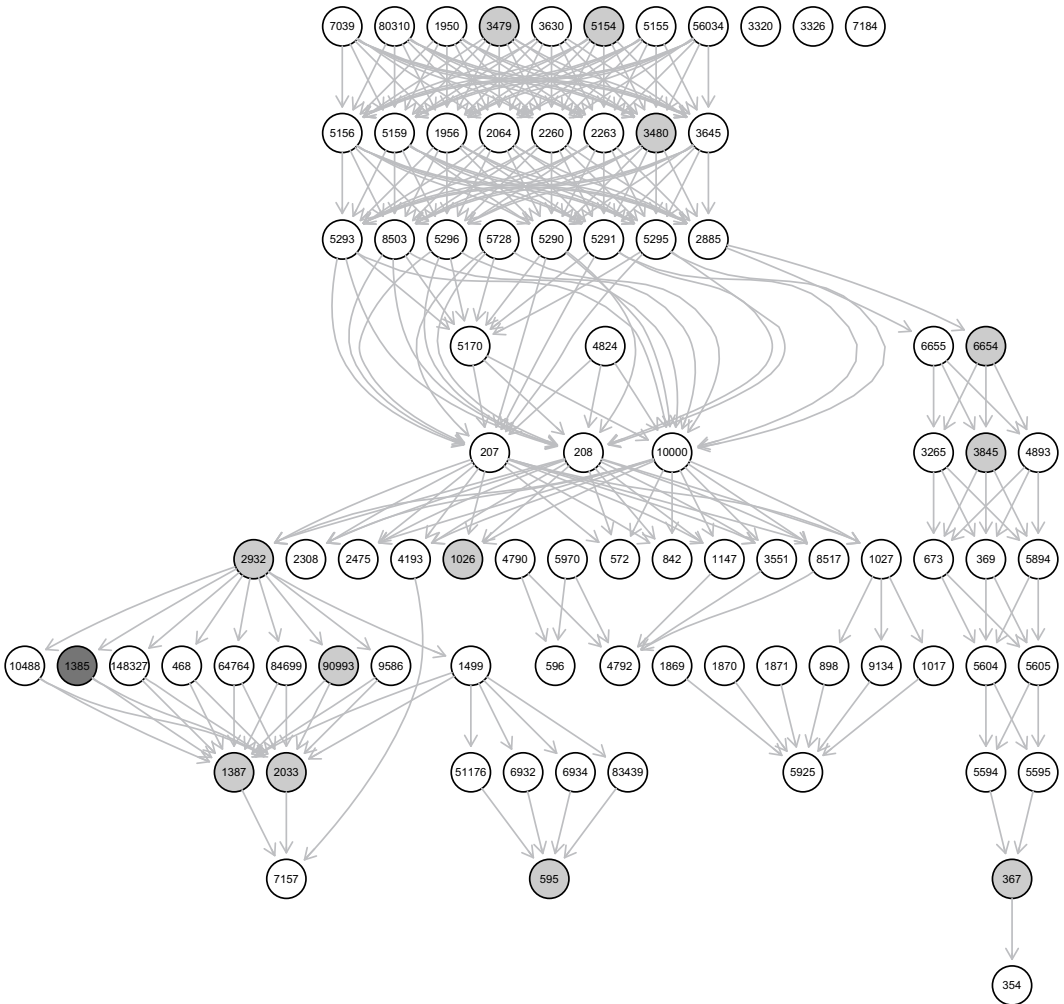
Fig. 4. DAG representing a prostate cancer pathway: ●, knock-down target; ●, genes affected by the knock-down according to the two-sample Wilcoxon test comparing the control with the knock-down condition (Bonferroni adjusted p-value below 0.05)

the affected genes are ancestors of CREB (2932, 3480, 3479 and 5154). Nonetheless, according to the principles of intervention analysis, genes that precede the target in the pathway DAG cannot be affected by the intervention. This effect may be brought about by either feedback loops or off-target silencing effects.

To explore specificity of this knock-down further, we considered the expression of the target, i.e. CREB, and of one of the genes that is associated with it, namely gene 2033. In Fig. 5(a), grey points represent the expression levels of CREB and of gene 2033 in the control condition. Although CREB should, according to the pathway annotation, inhibit gene 2033, the association between the two is positive. Disregarding the pathway annotation and looking only at their relationship in the control condition, it appears that an ideal knock-down of CREB should lower the mean expression of gene 2033. Instead, the mean of gene 2033 after knock-down has indeed increased, as shown by the cloud of black points, picturing the expression of CREB and

(a)

(b)
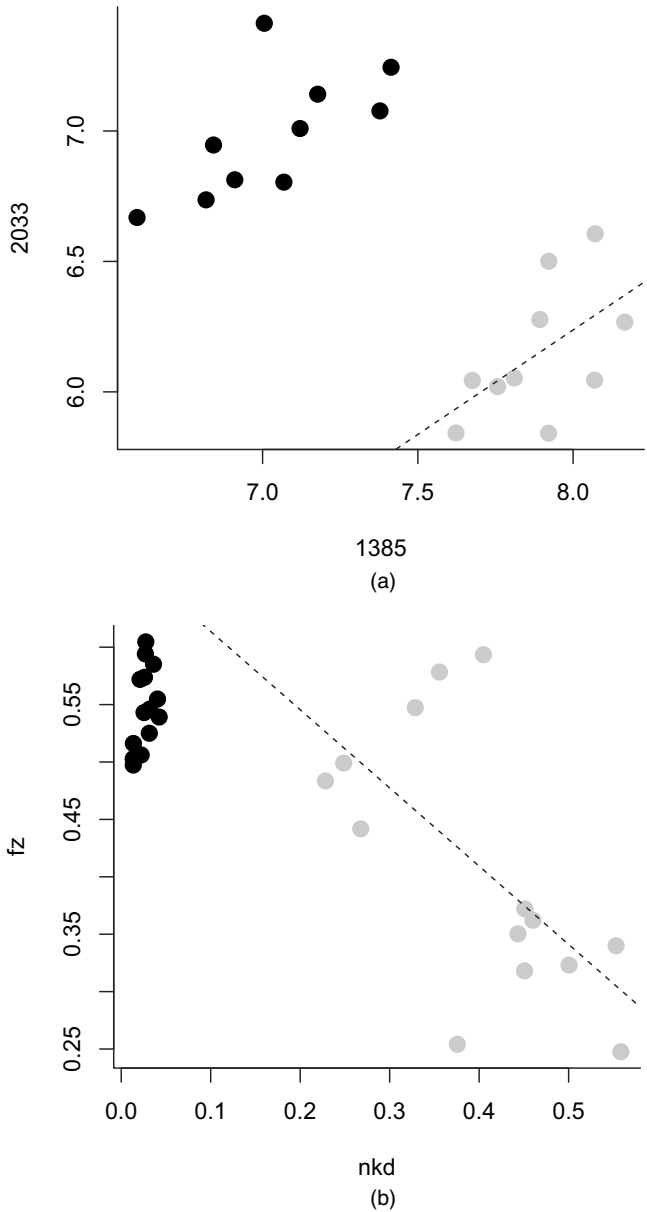
Fig. 5. Expression of knock-down targets and selected genes in (a) the CREB and in (b) the *nkd* knock-down: ⟋, ⟍, linear fit added for reference: ●, control; ●, knock-down

gene 2033 in the knock-down condition. Note that the association between the two remains positive also in the knock-down condition. This behaviour can be contrasted with the *nkd* knock-down that was considered earlier. Fig. 5(b) shows the expression of *nkd* and *fz* in the *Drosophila melanogaster* experiment. The knock-down seems to have been more successful in this case—consistently lower expression of *nkd* in the knock-down condition—and, although the mean of *fz* does not lie on the line representing the linear relationship in the control condition, it seems to be quite close to it.

Table 3. Results of the procedure for simulating the effects of CREB silencing†

Gene	Gene symbol	%($\delta_v = 0$)	$q_{0.025}$	$q_{0.975}$	d_v	p-value
148327	CREB3L4	91.1	−4.39	3.19	−0.20	3.15×10^{-1}
1499	CTNNB1	30.4	0.13	0.65	0.11	2.32×10^{-2}
2033	EP300	15.6	0.27	2.01	−0.97	1.08×10^{-5}
369	ARAF	67.8	−0.35	0.53	−0.12	1.65×10^{-1}
51176	LEF1	50.4	−0.55	0.78	0.12	1.90×10^{-1}
5604	MAP2K1	22.2	−1.38	0.10	−0.06	5.79×10^{-1}
5605	MAP2K2	27	−0.44	0.38	0.11	3.55×10^{-2}
6932	TCF7	11.4	−0.13	1.46	−0.39	1.50×10^{-3}
6934	TCF7L2	19.3	−0.34	0.76	−0.27	1.50×10^{-3}
83439	TCF7L1	8.2	−0.36	1.23	0.07	6.84×10^{-1}
898	CCNE1	20.8	−0.53	0.32	−0.14	2.32×10^{-2}
9134	CCNE2	10.8	−0.05	3.30	−0.27	6.31×10^{-1}
4792	NFKBIA	18.9	−0.46	0.99	−0.01	3.93×10^{-1}
595	CCND1	8.5	−0.21	1.11	−0.72	4.33×10^{-5}
1017	CDK2	18	−0.72	0.11	0.13	2.32×10^{-2}
1387	CREBBP	8.1	−0.39	0.59	−0.31	4.33×10^{-5}
5594	MAPK1	11.6	−0.38	0.13	−0.13	1.50×10^{-3}
5595	MAPK3	5.2	−0.48	−0.01	0.04	7.53×10^{-2}
5925	RB1	5.4	−0.18	0.92	−0.27	5.20×10^{-3}
7157	TP53	9.9	−0.07	0.68	0.01	7.39×10^{-1}
367	AR	16.3	−0.93	0.76	−0.66	1.08×10^{-5}
354	KLK3	4.5	−0.60	0.71	0.15	6.30×10^{-2}

†The gene identifier is given in the first column, the percentage of bootstrap samples in which the estimated causal effect was zero in the second column, the 2.5% and the 97.5% quantile of the non-zero component of the bootstrap distribution of the causal effect estimate in the third and fourth columns, the actually observed causal effect in the fifth column and the *p*-value for the test of equality of the mean expression in the silenced and the control condition in the last column.

To evaluate the effects of the CREB knock-down in an ideal experimental setting, we applied our approach to the data from the control condition. In step 1, we obtained a topological ordering of the pathway DAG. The target gene occupied position 63. We restricted our attention to the portion of the pathway DAG containing the target gene, its only parent (2932) and all the genes following the target gene in the topological ordering (genes ranked from 64 to 85). The results are shown in Table 3.

Although there are some genes for which the zero component in the bootstrap distribution was low, e.g. 83438, 595 and 1387, for most of them the associated bootstrap interval for the causal effect contained zero, i.e. the change after CREB silencing was not significant. In two cases— genes 2033 and 5595—the associated bootstrap intervals did not contain zero. Remarkably, in these cases, the predicted sign of the silencing effect was different from the effects that were actually observed in the knock-down group, but coherent with the evidence derived by the analysis of the control condition. For example, going back to gene 2033 that was considered earlier, a lowering of its mean level is predicted by our approach, demonstrating its ability to indicate the presence of an off-target effect.

8. Discussion

We have described a model-based technique that can be used as a simulation tool for predicting the effects of gene silencing at any node in a DAG. It provides a rapid and inexpensive way to

foresee effects of gene silencing that can be usefully employed for constructing appropriately tuned designs of gene silencing experiments.

Our approach is different from an apparently similar approach of Maathuis *et al.* (2009) called 'IDA' (intervention calculus when a DAG is absent), which has recently been generalized by Nandy *et al.* (2017). The aim of IDA is to identify the most promising targets to be silenced to produce the largest effect on the phenotype of interest; the aim of our approach is to identify the genes that are mostly affected by an intervention on a given target. In the former case, therefore, the focus is on intervention development; in the latter on intervention evaluation. Although slight, this distinction justifies different methodological solutions chosen in the two cases, most notably regarding the causal effect estimation. In IDA, the estimated DAG is used locally, since the causal effect of gene X_v on gene X_u is taken to be the coefficient of X_v in the linear model for X_u adjusted for the set of parents of X_v, i.e. the coefficient of X_v in the model $X_u \sim X_v + X_{\text{pa}(v)}$ in the formula syntax that is used by the statistical software R. By contrast, in our approach the graphical structure is used globally when estimating the matrix B from which the causal effect estimates are obtained. In practice, this means that the causal effect of X_v on X_u will be zero in our approach whenever there is no direct path between the associated nodes, whereas in IDA this will occur only if X_u is a parent of X_v.

Our approach relies on a guided structure learning of the DAG and intervention calculus, coupled with a bootstrap analysis. Guiding the learning of the DAG by defining a topological ordering and letting the data identify a structure that is compatible with such a flow of relationships is, in our view, a fair compromise

'between the ability to describe phenomena at the conceptual level and the ability to generate predictions at a measurable level'

(Shmueli, 2010). A side effect of this strategy is that possible conflicts between biological knowledge and data-driven information can serve as pointers to where in a pathway and/or in a DAG further analysis is required. The next step is to understand, within the context of the specific problem under analysis, the reasons for the observed inconsistencies.

As far as the learning strategy is concerned, two aspects are crucial when considering a learning strategy: the computational cost and the quality of the structure that it returns. The search strategy on which our proposal is based, the so-called greedy search, is particularly attractive for three main reasons. First, being heuristic, it enables dealing with the enormous size of the space of the DAGs (Robinson, 1977). Secondly, it is easy to implement and the evaluation of single-edge additions is computationally inexpensive. As an illustration, in the *Drosophila melanogaster* experiment, the complete analysis including estimation for 2000 bootstrap samples took just under 1 min on a personal computer. Similarly, the analysis for the CREB knock-down took under 2 min. Finally, it is justified by the asymptotic theory (Chickering and Meek, 2002; Chickering, 2003) and numerous empirical studies (see, for example, Aliferis and Cooper (1994)).

As suggested by one of the reviewers, one could, instead of relying on a greedy search, learn the graphical structure by solving a series of independent, possibly regularized, regression problems. Indeed, the problem of finding the optimal parent set for a given node could be seen as the problem of identifying variables with non-zero regression coefficients in a linear model containing all potential parents. If the number of potential parents is small, all possible subsets of parents for each node could be ranked according to some scoring criterion. When the number of potential parents is medium to large, such a search rapidly becomes computationally prohibitive. A simple strategy would consist of considering the restricted set of models that is obtained as solutions of a lasso problem (Tibshirani, 1996) for different values of the tuning parameter. To choose the optimal model among them (and thus the optimal value of the tuning

parameter), one could employ the extended Bayesian information criterion (Chen and Chen, 2008), which accounts for both the number of unknown parameters and the complexity of the model space. It is, however, unclear to what extent this approach can be useful in settings with a very low sample size, like those typically considered in the applications of interest for this paper. We leave this issue for future research.

The bootstrap strategy enables accounting for uncertainty in the predictions. Constructing the bootstrap distribution of silencing effects is, in our view, particularly informative. Indeed, it offers two distinct and equally interesting pieces of information: the percentage of cases when no effect was found, and the interval for the predicted mean when the effect was observed. As we employ simple non-parametric resampling, in small sample size settings some problems might occur if many resamples have sample values that are repeated many times. In that case it is advisable to control the composition of the resamples. It is also worth mentioning that in the case of very small sample sizes one might resort to algorithms for exact computation of non-parametric bootstrap estimates (Fisher and Hall, 1991).

Instead of constructing the bootstrap distribution of silencing effects, an alternative strategy could be employed. Specifically, one could first learn an ensemble of DAGs based on bootstrap resamples and then derive an aggregate DAG by some consensus measure. For example, Wang and Peng (2014) minimized the overall distance to the entire ensemble measured in terms of the structural Hamming distance. The overall distance to the pathway DAG could also be considered. In our study, such a strategy led to the same substantive conclusions as were reached by our approach.

There are still practical and theoretical considerations that require further investigation. The multiplicity of topological orderings and its effect on predictions is one such issue. Different topological orderings might lead to different estimated DAGs, i.e. to alternative tools to generate predictions of gene silencing. We do not perceive this as a limitation of our method. Indeed, in a predictive perspective—that does not need to exploit the exact role of each variable in terms of an underlying causal structure—different models might provide equally valid predictions. Our feeling was supported by findings in the *Drosophila melanogaster* experiment, where different topological orderings led to estimated DAGs with the same descendants (*fz* and *por*) for the intervened gene (*nkd*) and comparable predictions of silencing effects (the results are reported in Appendix A). An alternative solution to the problem of multiple orderings would be to pursue a generalization of the method proposed that takes as an input a partial, rather than a complete, topological ordering.

Another issue of concern arising when sets of genes are modelled by DAGs is the dynamic nature of gene networks characterized by feedback loops. Obviously, feedback loops correspond to cycles and cannot be modelled by DAGs. A possible solution to this problem is offered by dynamic DAGs that unfold loops in time, but methodological research in this direction has been hindered by both technological and biological issues. Experiments with repeated measurements of gene expression are still an exception, rather than a rule and, moreover, since events within a cell exhibit great variation in terms of the amount of time that they require, determining an appropriate distribution of points in time when gene expression is to be measured is a challenging task. With reference to the method that is proposed here, these limitations imply that the predicted effects of gene silencing refer to an instantaneous reaction of the system to an external intervention, which might be modified with time through various feedback mechanisms.

graphical models, and Jiři Vomlel for generously sharing his expertise in the area of structure learning algorithms.

Appendix A

A.1. CK2 algorithm

The CK2 algorithm is an adaptation of a score-based greedy search algorithm for structure learning: the K2 algorithm (Cooper and Herskovits, 1992). In the K2 algorithm, variables are ordered according to the prespecified topological ordering and parent sets for all nodes are initially empty. For each node, the algorithm then searches, among potential parents, i.e. variables preceding the given node in the topological ordering, for the variable that increases the chosen score the most and adds it to the parent set. This procedure is repeated until either the score cannot be improved any further or the maximum prespecified number of parents has been reached. The pseudocode is given in Table 4.

To apply the K2 algorithm to continuous data we modified the scoring function by considering criteria that are applicable to continuous data. We opted for a simple solution that is offered by the Bayesian information criterion (BIC). If we assume that n independent realizations of X_V have been observed, then the BIC score of a DAG G is

$$\mathrm{BIC}(G) = n\{\log|\hat{K}| - \mathrm{tr}(S\hat{K})\} - \log(n) \sum_{v \in V} \mathrm{card}\{\mathrm{pa}(v)\},$$

where S is the sample covariance matrix, $\hat{K}$ is the maximum likelihood estimate of the concentration matrix (constrained by the conditional independence relationships that are encoded in G), '$|\cdot|$' denotes a matrix determinant, $\mathrm{tr}(\cdot)$ denotes a trace and $\mathrm{card}(A)$ denotes the cardinality of the set A.

The BIC belongs to the Bayesian scoring metrics family and can be seen as an asymptotic approximation of the posterior probability of the structure, i.e. the approximation of the full posterior probability integrated over all possible parameterizations of the conditional densities for the given structure. Furthermore, Haughton (1988) showed that the BIC is a consistent scoring criterion. We recall the definition of consistency of a scoring criterion.

Definition 1 (consistency of a scoring criterion).   Assume that data are generated by some distribution f^* whose underlying DAG is G^* (in other words, the set of conditional independence relationships that hold in f^* coincides with the set of conditional independence relationships that are implied by G^*). We say that a scoring function is consistent if the following properties hold as the number of observations goes to ∞, with probability that approaches 1:

(a) the structure G^* will maximize the score;
(b) all structures that are not equivalent to G^* will have strictly lower score.

Table 4.   Algorithm 1: the K2 algorithm (Cooper and Herskovits, 1992)

```
Input: n observations of p variables X₁,...,Xₚ, a topological ordering, an upper bound
     for the number of parents u
for i = 0 to p do
    paᵢ ← ∅
    P_old ← g(i, paᵢ)          g is the chosen scoring criterion
    OKToProceed ← true
    while OKToProceed and |paᵢ| < u do
        let z be the node in Pred(xᵢ) \ paᵢ that maximizes g(i, paᵢ ∪ {z})
        P_new ← g(i, paᵢ ∪ {z})
        if P_new > P_old then
            P_new ← P_old
            paᵢ ← paᵢ ∪ {z}
        else OKToProceed ← false
        end if
    end while
    print 'parents of the node:', xᵢ, 'are', paᵢ
end for
```

Table 5. Results of the procedure for simulating the effects of silencing the *nkd* gene for eight alternative topological orderings compatible with the pathway DAG†

Ordering	Results for por			Results for fz		
	$\%(\delta_v = 0)$	$q_{0.025}$	$q_{0.975}$	$\%(\delta_v = 0)$	$q_{0.025}$	$q_{0.975}$
1	36.20	−2.16	−0.10	4.85	−1.07	−0.13
2	73.10	−3.34	−0.55	13.60	−1.14	−0.26
3	57.80	−2.97	0.40	11.00	−1.16	−0.11
4	36.20	−2.16	−0.10	4.85	−1.07	−0.13
5	26.90	−0.98	−0.15	4.30	−0.98	−0.14
6	26.90	−0.98	−0.15	4.30	−0.98	−0.14
7	36.20	−2.16	−0.10	4.85	−1.07	−0.13
8	26.90	−0.98	−0.15	4.30	−0.98	−0.14

†Only the effect on the two most affected genes, *por* and *fz*, are reported. For each gene, $\%(\delta_v = 0)$ is the frequency with which the estimated causal effect was zero in the bootstrap distribution, whereas $q_{0.025}$ and $q_{0.975}$ denote the 2.5% and 97.5% quantile of the non-zero component of the bootstrap distribution of the causal effect.

Chickering and Meek (2002) and Chickering (2003) derived optimality results stating that the greedy search used in conjunction with any consistent scoring criterion will, as the number of observations goes to ∞, identify the true structure up to an equivalence class. This, together with the fact that CK2 uses the specified topological ordering, ensures its consistency.

A.2. Alternative topological orderings

Our proposed approach relies on a topological ordering that is obtained from the pathway DAG. In general, there are many such orderings and predictions that are obtained might depend on a particular choice. To investigate this issue in our experiment, we considered an additional eight topological orderings in line with the pathway DAG. We predicted the effect of silencing under each of them, and the results for the two most affected genes are reported in Table 5. In all of the orderings considered, the silenced gene was the first, so all remaining genes could be potentially affected. For each ordering, the same 2000 bootstrapped data sets were generated, so that the observed variation can be assigned only to the topological ordering. Indeed, it can be seen that, for instance, orderings 5, 6 and 8 lead to identical results. More importantly, results under different orderings seem to be consistent, and in this case lead to the conclusion that *fz* gene is the only gene that can be reliably expected to be affected by the silencing of *nkd*.

References

Aliferis, C. F. and Cooper, G. F. (1994) An evaluation of an algorithm for inductive learning of Bayesian belief networks using simulated data sets. In *Proc. 10th Conf. Uncertainty in Artificial Intelligence* (eds R. L. De Mantanos and D. Poole), pp. 8–14. San Francisco: Morgan Kaufmann.

Chen, J. and Chen, Z. (2008) Extended Bayesian information criteria for model selection with large model spaces. *Biometrika*, **95**, 759–771.

Chickering, D. M. (2003) Optimal structure identification with greedy search. *J. Mach. Learn. Res.*, **3**, 507–554.

Chickering, D. M. and Meek, C. (2002) Finding optimal Bayesian networks. In *Proc. 18th Conf. Uncertainty in Artificial Intelligence*, pp. 94–102. San Francisco: Morgan Kaufmann.

Cho, H., Berger, B. and Peng, J. (2016) Reconstructing causal biological networks through active learning. *PLOS One*, **11**, article e0150611.

Cooper, G. F. and Herskovits, E. (1992) A Bayesian method for the induction of probabilistic networks from data. *Mach. Learn.*, **9**, 309–347.

Csardi, G. and Nepusz, T. (2006) The igraph software package for complex network research. *InterJournal*, article 1695.

Cullen, B. R. (2006) Enhancing and confirming the specificity of RNAi experiments. *Nat. Meth.*, **3**, 677–681.

Djordjilović, V. (2015) Graphical modelling of biological pathways. *PhD Thesis*. University of Padova, Padova.

Djordjilović, V., Chiogna, M., Massa, M. S. and Romualdi, C. (2015) Graphical modeling for gene set analysis: a critical appraisal. *Biometr. J.*, **57**, 852–866.

Djordjilović, V., Chiogna, M. and Vomlel, J. (2017) An empirical comparison of popular structure learning algorithms with a view to gene network inference. *Int. J. Approx. Reason.*, **88**, 602–613.

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E. and Mello, C. C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, **391**, 806–811.

Fisher, N. I. and Hall, P. (1991) Bootstrap algorithms for small samples. *J. Statist. Planng Inf.*, **27**, 157–169.

Haughton, D. M. (1988) On the choice of a model to fit data from an exponential family. *Ann. Statist.*, **16**, 342–355.

Hauser, A. and Bühlmann, P. (2012) Characterization and greedy learning of interventional Markov equivalence classes of directed acyclic graphs. *J. Mach. Learn. Res.*, **13**, 2409–2464.

Haynes, W. A., Higdon, R., Stanberry, L., Collins, D. and Kolker, E. (2013) Differential expression analysis for pathways. *PLOS Computnl Biol.*, **9**, article e1002967.

Ideker, T., Dutkowski, J. and Hood, L. (2011) Boosting signal-to-noise in complex biology: prior knowledge is power. *Cell*, **144**, 860–863.

Jackson, A. L. and Linsley, P. S. (2004) Noise amidst the silence: off-target effects of siRNAs? *Trends Genet.*, **20**, 521–524.

Kanehisa, M. and Goto, S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.*, **28**, 27–30.

Lauritzen, S. L. (1996) *Graphical Models*. Oxford: Clarendon.

Lauritzen, S. L. (2001) Causal inference from graphical models. In *Complex Stochastic Systems* (eds O. E. Barndorff-Nielsen, D. R. Cox and C. Klüppelberg), pp. 63–107. London: Chapman and Hall.

Ledford, H. (2018) Gene-silencing drug approved. *Nature*, **560**, 291–292.

Ma, J., Shojaie, A. and Michailidis, G. (2016) Network-based pathway enrichment analysis with incomplete network information. *Bioinformatics*, **32**, 3165–3174.

Maathuis, M. H., Colombo, D., Kalisch, M. and Bühlmann, P. (2010) Predicting causal effects in large-scale systems from observational data. *Nat. Meth.*, **7**, 247–248.

Maathuis, M. H., Kalisch, M. and Bühlmann, P. (2009) Estimating high-dimensional intervention effects from observational data. *Ann. Statist.*, **37**, 3133–3164.

Markowetz, F., Grossmann, S. and Spang, R. (2005) Probabilistic soft interventions in conditional Gaussian networks. In *Proc. 10th Int. Wrkshp Artificial Intelligence Statistics* (eds R. G. Cowell and Z. Gharamani), pp. 214–221.

Moffatt, S. (2016) siRNA-based nanoparticles for cancer therapy: hurdles and hopes. *MedCrave Online J. Proteom. Bioinform.*, **4**, 345–347.

Nandy, P., Maathuis, M. H. and Richardson, T. S. (2017) Estimating the effect of joint interventions from observational data in sparse high-dimensional settings. *Ann. Statist.*, **45**, 647–674.

Opgen-Rhein, R. and Strimmer, K. (2007) Accurate ranking of differentially expressed genes by a distribution-free shrinkage approach. *Statist. Appl. Genet. Molec. Biol.*, **6**, article 9.

Pearl, J. (2000) *Causality: Models, Reasoning and Inference*. Cambridge: Cambridge University Press.

Pei, Y. and Tuschl, T. (2006) On the art of identifying effective and specific siRNAs. *Nat. Meth.*, **3**, 670–676.

Qiu, S., Adema, C. M. and Lane, T. (2005) A computational study of off-target effects of RNA interference. *Nucleic Acids Res.*, **33**, 1834–1847.

Rau, A., Jaffrézic, F. and Nuel, G. (2013) Joint estimation of causal effects from observational and intervention gene expression data. *BMC Syst. Biol.*, **7**, 111.

Robinson, R. W. (1977) Counting unlabeled acyclic digraphs. In *Combinatorial Mathematics V* (ed. C. H. C. Little), pp. 28–43. Berlin: Springer.

Sales, G., Calura, E. and Romualdi, C. (2018) metaGraphite—a new layer of pathway annotation to get metabolite networks. *Bioinformatics*, **35**, 1258–1260.

Schafer, J., Opgen-Rhein, R., Zuber, V., Ahdesmaki, M., Silva, A. P. D. and Strimmer, K. (2017) corpcor: efficient estimation of covariance and (partial) correlation. *R Package Version 1.6.9*.

Schäfer, J. and Strimmer, K. (2005) A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statist. Appl. Genet. Molec. Biol.*, **4**, article 30.

Shmueli, G. (2010) To explain or to predict? *Statist. Sci.*, **25**, 289–310.

Shojaie, A. and Michailidis, G. (2009) Analysis of gene sets based on the underlying regulatory network. *J. Computnl Biol.*, **16**, 407–426.

Slenter, D. N., Kutmon, M., Hanspers, K., Riutta, A., Windsor, J., Nunes, N., Mélius, J., Cirillo, E., Coort, S. L., Digles, D., Ehrhart, F., Giesbertz, P., Kalafati, M., Martens, M., Miller, R., Nishida, K., Rieswijk, L., Waagmeester, A., Eijssen, L. M. T., Evelo, C. T., Pico, A. R. and Willihagen, E. L. (2017) Wikipathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res.*, **46**, D661–D667.

Spirtes, P., Glymour, C. N. and Scheines, R. (2000) *Causation, Prediction, and Search*. Cambridge: MIT Press.

Sverchkov, Y. and Craven, M. (2017) A review of active learning approaches to experimental design for uncovering biological networks. *PLOS Computnl Biol.*, **13**, article e1005466.

Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. *J. R. Statist. Soc.* B, **58**, 267–288.

Vaske, C. J., Benz, S. C., Sanborn, J. Z., Earl, D., Szeto, C., Zhu, J., Haussler, D. and Stuart, J. M. (2010) Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. *Bioinformatics*, **26**, i237–i245.

Wang, R. and Peng, J. (2014) Learning directed acyclic graphs via bootstrap aggregating. *Preprint arXiv:1406.2098*. University of California at Davis, Davis.