



Context-Specific Nested Effects Models

Yuriy Sverchkov¹(✉)() , Yi-Hsuan Ho²() , Audrey Gasch²() ,
and Mark Craven¹()

¹ Department of Biostatistics and Medical Informatics,
University of Wisconsin–Madison, Madison, WI, USA
yuriy.sverchkov@wisc.edu

² Department of Genetics, University of Wisconsin–Madison, Madison, WI, USA

Abstract. Advances in systems biology have made clear the importance of network models for capturing knowledge about complex relationships in gene regulation, metabolism, and cellular signaling. A common approach to uncovering biological networks involves performing perturbations on elements of the network, such as gene knockdown experiments, and measuring how the perturbation affects some reporter of the process under study. In this paper, we develop context-specific nested effects models (CSNEMs), an approach to inferring such networks that generalizes nested effect models (NEMs). The main contribution of this work is that CSNEMs explicitly model the participation of a gene in multiple *contexts*, meaning that a gene can appear in multiple places in the network. Biologically, the representation of regulators in multiple contexts may indicate that these regulators have distinct roles in different cellular compartments or cell cycle phases. We present an evaluation of the method on simulated data as well as on data from a study of the sodium chloride stress response in *Saccharomyces cerevisiae*.

1 Introduction

Cellular processes such as gene regulation, metabolism, and signaling form complex interplay of molecular interactions. A primary means of uncovering the details of these processes is through the analysis of measured responses of cells to perturbation experiments. We present Context-Specific Nested Effect Models (CSNEMs), which are graphical models for analyzing screens of high-dimensional phenotypes from gene perturbations. In this setting, the perturbation consists of knocking out, knocking down, or otherwise disabling the activity of a gene, via the use of deletion mutants, RNA interference, CRISPR/Cas9, or other techniques. The high-dimensional phenotype may be a transcriptomic, proteomic, metabolomic or similar multidimensional profile of measurements. Such profiles provide indirect information about the pathways that connect the gene that is perturbed in an experiment to the effects observed in a phenotype. This poses a challenge for determining functional relationships, since the precise mechanisms by which the perturbation relates to the phenotype must be inferred using computational and statistical methods, expert knowledge, or a combination of both.

Related work on inferring networks from gene expression data includes methods based on statistical dependencies between expression measurements [4, 7], which are used to construct networks of probable interactions between the genes measured in the expression profile. Other work on using phenotypic data uses clustering of phenotypic profiles, or the similarity between profiles, to construct networks among the perturbation genes [17, 19]. The rationale behind these approaches is that genes that produce similar phenotypes when perturbed are likely to be functionally related [13].

The CSNEM approach is a generalization of the Nested Effect Model (NEM) [11]. In the NEM approach, a network structure among the perturbed elements of the cell is inferred from the nested structure of phenotypic profiles. The general idea is that perturbation of a gene that is further upstream in a signaling pathway would affect more elements than perturbation of a gene further downstream. For example, Fig. 1(a) shows an NEM in which Hog1 is upstream of Cka2. The table underneath the graph represents the differential expressions of the high-dimensional phenotypes observed in the screen, with rows corresponding to single-gene knockouts and each column corresponding to an *effect*: one dimension of a phenotype, such as a particular transcript in a transcriptomic phenotype. In the table of effect measurements in the figure, a ‘1’ indicates that a perturbation changed the response of the effect, and a ‘0’ indicates that it did not. The deletion of Hog1 would affect e_1, e_2, e_3 and e_4 because they are all downstream of it. The deletion of Cka2, on the other hand, would only affect e_3 and e_4 . Therefore, the nesting of the effects of the deletion of Cka2 within the effects of the deletion of Hog1 places the former downstream of the latter.

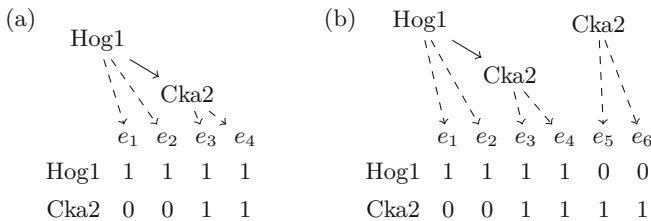


Fig. 1. (a) An example of effect nesting in an NEM, and (b) a partial intersection of effects as captured by a CSNEM. The table underneath each graph represents the differential expressions of the high-dimensional phenotypes observed in the screen, with rows corresponding to single-gene knockouts and each column corresponding to an *effect*, one dimension of a phenotype, where a ‘1’ indicates that a perturbation changed the response of the effect, and a ‘0’ indicates that it did not.

Such nesting of effects, however, does not always occur. The protein product of a gene may interact with those of other genes in a multitude of ways, and one might imagine a situation where two genes are interacting with each other upstream of a subset of the effects, but additionally have other roles independently of each other. This is the case in Fig. 1(b), where, upstream of effects e_1, e_2, e_3 and e_4 Cka2 and Hog1 interact as before, but Cka2 additionally affects

e_5 and e_6 independently of Hog1. In such a case, we see that the phenotype induced by the perturbations of each gene includes effects downstream of the common pathway, but each perturbation also shows unique effects, and rather than being nested, the effects show a partial intersection. The example in Fig. 1 is based on a pattern we identified in our application of CSNEM learning to experiments studying sodium chloride (NaCl) stress response in *Saccharomyces cerevisiae*.

In the CSNEM approach, we address this issue by explicitly considering the possibility that one gene may have multiple contexts of interaction. The model can be equivalently viewed either as a single graph model where multiple nodes may represent multiple roles of the same gene, or as a mixture of multiple NEMs, where each NEM describes a different subset of the effects. Notably, mixtures of NEMs have been used for analyzing single-cell expression data [22]. In that work, the mixture is used to account for variation of gene activation states across different cells. In contrast, in a CSNEM, the mixture represents different patterns of interaction among the same sets of genes across different subsets of the measured effects. The effect pattern in Fig. 1(b) can alternatively be accounted for by the introduction of a hidden node downstream of both Hog1 and Cka2, an approach explored by Sadeh et al. [21], where they introduce a statistical test to infer a partially resolved nested effect model. In fact, Sadeh et al. show that the presence of a hidden node downstream of a pair of genes is consistent with every possible configuration of effect responses. Their method aims to characterize all possible NEM models that are consistent with the data, and as a result it never rejects the possibility of a hidden node existing downstream of any pair of genes. In contrast, in our approach we aim to find a single parsimonious network model that optimally fits the data. We show how to cast the problem of learning a CSNEM as a modified version of NEM learning, evaluate the ability of this approach to recover a ground-truth network on simulated data, and present an application to the salt stress pathway in yeast.

2 Background: Nested Effects Models

Tresch and Markowitz [25] formulate nested effects models (NEMs) as a special case of effects models. In an effects model, there is a set of actions \mathcal{A} , and a set of effects \mathcal{E} , and we wish to model which effects change in response to each action. In earlier work on nested effects models [11], the actions and effects are respectively referred to as S-genes (S for signaling) and E-genes (E for effects). The actions correspond to perturbation experiments, while the effects correspond to the high-dimensional phenotype measured in the experiment. A general effects model can be represented by a binary matrix F where $F_{ae} = 1$ if action a leads to a response (or change) in effect e , and 0 otherwise.

Let $n_{\mathcal{A}}$ and $n_{\mathcal{E}}$ represent the number of actions and effects, respectively. An NEM is made up of a directed graph G the nodes of which are the actions \mathcal{A} , and an $n_{\mathcal{A}} \times n_{\mathcal{E}}$ binary matrix Θ of attachments, in which $\Theta_{ae} = 1$ if effect e is attached to action a , and 0 otherwise. A modeling constraint is that each effect is attached to at most one action.

The NEM is interpreted as follows: action a causes a response in effect e if and only if either e is attached directly to a , or there is a directed path in G from a to the action to which e is attached. Mathematically, this can be formulated in terms of matrix multiplication. Since what matters is which actions are reachable from other actions in G , we can work with Γ , the $n_A \times n_A$ accessibility matrix of G . Γ_{ab} is 1 if there is a directed path from a to b in G , and 0 otherwise. As a matter of convention and for mathematical convenience, the diagonal entries, Γ_{aa} are all 1s. Using Γ , we can express the effects matrix F of an NEM as $F = \Gamma\Theta$.

2.1 Likelihood Computation

The problem of inferring an NEM from a data set D can be viewed as that of maximizing a likelihood. In this section we review how the likelihood of an NEM is framed to illustrate how the likelihood of a CSNEM relates to it.

Supposing that we have some data consisting of measurements of the observable effects subject to each action included in the model, and assuming data independence, for a general effects model, the log-likelihood of the model is

$$\log L(F) = \log \mathbb{P}(D|F) = \sum_{(a,e) \in \mathcal{A} \times \mathcal{E}} \log \mathbb{P}(D_{ae}|F_{ae}). \quad (1)$$

Where $\mathbb{P}(D_{ae}|F_{ae})$ is the probability of the data we observed in regard to effect e subject to action a given that F_{ae} indicates whether we expect a response in e subject to a . When the observed phenotype is, for example, gene expression data, a typical indicator of a response in effect e is differential expression of effect e between the experimental condition a and a control, such as a wild-type phenotype.

Let $R \in \mathbb{R}^{n_{\mathcal{E}} \times n_{\mathcal{A}}}$ be a matrix of log-likelihood ratios such that $R_{ea} = \frac{\mathbb{P}(D_{ae}|F_{ae}=1)}{\mathbb{P}(D_{ae}|F_{ae}=0)}$, and let N represent the null model predicting no effect response to any action, Tresch and Markowitz [25] show that the log-likelihood of an effects model F is then

$$\log L(F) = \text{tr}(FR) + \underbrace{\log L(N)}_{\text{constant w.r.t. data}} \quad (2)$$

where $\text{tr}(\cdot)$ is the trace of a matrix. The above holds for any effects model in general. Since in an NEM, $F = \Gamma\Theta$, to maximize the likelihood of an NEM one would maximize $\text{tr}(\Gamma\Theta R)$.

Computationally, maximizing this expression is difficult because it is a search over a discrete but exponentially large space of all possible Γ and Θ matrices. Early work on NEMs reduces some of the complexity of this search by observing that since Θ can only have one 1 for each effect across all actions by construction, and since $\text{tr}(\Gamma\Theta R) = \text{tr}(R\Gamma\Theta)$, one can marginalize over all possible values of Θ , assuming that they are equally likely *a-priori*, yielding a marginal likelihood proportional to $\prod_{e \in \mathcal{E}} \sum_{a \in \mathcal{A}} \exp((R\Gamma)_{ea})$. This reduces the task to the search for a Γ that maximizes this marginal likelihood, an exhaustive search for which

is feasible for $n_{\mathcal{A}} \leq 5$ [11]. For larger graphs, however, the problem is still computationally restrictive, and multiple algorithms for learning nested effects model structure efficiently have been presented in the literature [6, 12], most of which have been implemented in the `nem` R package [5]. Other approaches to computing the likelihood have also been explored, such as the factor graph optimization approach by [26].

In this work, we show how learning a CSNEM can be cast as a more complex NEM learning problem. To solve the NEM learning problem, we use MC-EMiNEM, a method that does not attempt to optimize a marginal likelihood, as many of the above approaches do, but maximizes the log posterior

$$\log \mathbb{P}(I, \Theta | D) = \log L(I\Theta) + \sum_{(a,b) \in \mathcal{A} \times \mathcal{A}} \log \mathbb{P}(\Gamma_{a,b}) + \log \mathbb{P}(\Theta). \quad (3)$$

Where $\log \mathbb{P}(\Gamma_{i,i})$ is an edge-wise prior on the structure of the actions graph and $\mathbb{P}(\Theta)$ is a prior on the attachment matrix. MC-EMiNEM uses Monte Carlo (MC) sampling and Expectation Maximization (EM) within MC steps to search for the I and Θ that are optimal with respect to this posterior [16]. MC-EMiNEM is available as a part of the `nem` R package.

3 Methods: Context-Specific Nested Effects Models

As briefly mentioned in the introduction, the motivation for developing CSNEMs is that there are cases in which phenotype effects are not nested, as in the example in Fig. 1. In CSNEMs, we account for situations like the partial overlap in Fig. 1 by allowing an action in the graph to be represented by more than one node, and we call these different nodes that correspond to the same action different *contexts* of the action. Mathematically, this enables the model to represent relationships that are not representable by an NEM. Biologically, different contexts in a CSNEM may correspond to participation in different pathways, either due to physical separation such as localization of molecules, or temporal separation, such as participation in different stages of the cell cycle.

The CSNEM in Fig. 1(b) is presented as a single NEM-like graph with multiple contexts for the `Cka2` node. Note that the same diagram can also be viewed as a pair of NEMs: one containing `Hog1` and `Cka2`, which applies to effects e_1, e_2, e_3, e_4 , and another containing only `Cka2`, which applies to the effects e_5 and e_6 . This view of a CSNEM as a mixture of NEMs is most useful in understanding our approach to learning a CSNEM from data.

3.1 The Likelihood of a k -CSNEM

We define a k -CSNEM as a mixture of k NEM's, where the response of each effect e is governed by one of k NEMs, each of which can have a different graph G relating the actions \mathcal{A} . A k -CSNEM is therefore parameterized by k accessibility matrices $\Gamma^1, \dots, \Gamma^k$, each of which is $n_{\mathcal{A}} \times n_{\mathcal{A}}$ and by a vector θ , each coordinate

of which takes one of $kn_{\mathcal{A}} + 1$ values, specifying attachment to one of the $n_{\mathcal{A}}$ actions in one of the k NEMs, or the absence of attachment.

The parameter θ partitions the space of effects by assigning each effect to one of the k NEMs (or to none of them). As a matter of convention, we represent attachment of effect $e \in \mathcal{E}$ to an action $a \in \mathcal{A}$ in mixture member $i \in \{1, \dots, k\}$ by $\theta_e = (i - 1)n_{\mathcal{A}} + a$ (we slightly abuse notation, treating actions as natural numbers $1, \dots, n_{\mathcal{A}}$ here), and let $\theta_e = 0$ if the effect is not attached to any action in any NEM. We can then define the partition of \mathcal{E} into k sets $\mathcal{E}_1, \dots, \mathcal{E}_k$ as

$$\mathcal{E}_i = \{e \in \mathcal{E} \mid \exists a \in \mathcal{A} : \theta_e = (i - 1)|\mathcal{A}| + a\} \text{ for } i \in \{1, \dots, k\}. \quad (4)$$

Let us define a mapping of effect indices, which will be useful later: $\zeta : \{1, \dots, k\} \times \{1, \dots, |\mathcal{E}_i|\} \rightarrow \mathcal{E}$. Thus, $\zeta(i, j) = e$ when effect e is the j th member of partition \mathcal{E}_i . Given this partition, the likelihood of a CSNEM is defined as the product of the NEM likelihoods per partition:

$$L(\Gamma^{1, \dots, k}, \theta) = \prod_{i=1}^k L(\Gamma^i, \Theta^i) \quad (5)$$

where Θ^i is a matrix in $\{0, 1\}^{|\mathcal{A}| \times |\mathcal{E}_i|}$ and $\Theta_{aj}^i = 1$ iff $\theta_{\zeta(i, j)} = (i - 1) + a$, and 0 otherwise.

In relation to the CSNEM, let us combine the mixture of NEMs into one structure by defining the block diagonal matrix Γ made of blocks Γ^i , define $\Theta \in \{0, 1\}^{|\mathcal{A}| \times |\mathcal{E}|}$ by $\Theta_{ae} = 1$ iff $\theta_e = a$, and let be a block matrix made up of k appended $|\mathcal{A}| \times |\mathcal{A}|$ identity matrices:

$$\Gamma = \begin{bmatrix} \Gamma^1 & 0 & \dots & 0 \\ 0 & \Gamma^2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \Gamma^k \end{bmatrix}, \quad \Psi = \underbrace{[I_{|\mathcal{A}|} \ I_{|\mathcal{A}|} \ \dots \ I_{|\mathcal{A}|}]}_{k \text{ copies}}. \quad (6)$$

Let R^i be a matrix in $\mathbb{R}^{|\mathcal{E}_i| \times |\mathcal{A}|}$ where $R_{ja}^i = R_{\zeta(i, j), a}$ (i.e., R^i is a selection of effects from R based on the partition \mathcal{E}_i). Given these definitions the log-likelihood of the CSNEM can be written as¹

$$\log \prod_{i=1}^k L(\Gamma^i, \Theta^i) = \text{tr}(\Gamma \Theta (R \Psi)) + \log L(N). \quad (7)$$

Thus, the likelihood of a k -CSNEM is equal to the likelihood of an NEM with $k|\mathcal{A}|$ actions for the data matrix $R\Psi$, subject to the constraint that Γ is block diagonal as in (6). We can consequently use any NEM learner to learn a k -CSNEM mixture, as long as it supports constraining Γ to be block-diagonal.

¹ For a detailed derivation see <https://github.com/sverchkov/mc-em-cs-nem/blob/master/recomb-2018-supplement/recomb-2018-supplement.pdf>, commit 98b01f19357e3d58eae81764d42a6903624e3433 at the time of submission.

Analogously to (3), we can obtain a posterior probability for the CSNEM by introducing priors for Γ and Θ , and applying MC-EMiNEM to maximize that posterior. The block-diagonal constraint can be enforced using the edge-wise prior on the structure of Γ , by setting the priors on edges that would violate block-diagonality to zero.

3.2 Compact Visualization and Identifiability of a k -CSNEM

Having obtained k NEMs and the corresponding partitioning of the effect set, a single graph can be composed by merging all action nodes across the graphs that have the same ancestors (are reachable from the same set of actions). Figure 2 provides an example: Fig. 2(a) shows three graphs that describe the structures of three NEMs that compose a mixture, and Fig. 2(b) shows the result of merging them. Note that Hog1 is reachable from no nodes but itself in all three NEMs. Consequently, in the compact CSNEM, there is only one version of Hog1. In contrast, Cka2 is reachable from Hog1 in one of the NEMs, and is only reachable from itself in the others, which is why it has two contexts in the CSNEM. Similarly, Ckb14 is reachable from both Hog1 and Cka2 in one of the three NEMs, but not the others, and has two contexts as well. To keep track of the various contexts, we append the list of genes from which a context is reachable when displaying the graph, e.g. the context of Cka2 that is reachable from Hog1 is labeled ‘Cka2 [Hog1],’ while the context that is not reachable from other nodes is labeled simply ‘Cka2.’ This is particularly helpful when viewing graphs with many nodes and many contexts.

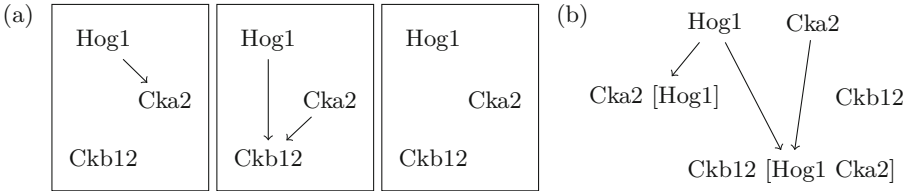


Fig. 2. Building a CSNEM from a mixture of NEMs. (a) Three NEMs that compose a mixture. (b) A single graph obtained by an edge-preserving merge of the three NEMs.

The merged graph in Fig. 2(b) preserves the edges that were present in the mixture of NEMs, but it is not necessarily a unique maximizer of the likelihood, rather, it is a member of an equivalence class of equally likely CSNEMs. What characterizes the equivalence class is the set of *inclusive ancestries* of the nodes in the CSNEM. The inclusive ancestry of a node is a set of actions; this set contains the action at the node and all actions from which it is reachable: e.g. the inclusive ancestry of the Cka2 node in the leftmost NEM in Fig. 2(a) is {Hog1, Cka2}, while the inclusive ancestry of the Cka2 node in the middle NEM is simply {Cka2}. The set of inclusive ancestries for the example in Fig. 2 is

therefore $\{\{\text{Hog1}\}, \{\text{Cka2}\}, \{\text{Ckb12}\}, \{\text{Hog1}, \text{Cka2}\}, \{\text{Hog1}, \text{Cka2}, \text{Ckb12}\}\}$. Any two CSNEMs with identical sets of inclusive ancestries necessarily have the same set of unique accessibility matrix columns $\Gamma_a^i : i \in \{1, \dots, k\}, a \in \mathcal{A}$, and consequently, have the same likelihood for likelihood-maximizing attachments Θ . The characterization of equivalence classes in terms of inclusive ancestry sets relates to previous results about NEM identifiability: for transitively closed Γ , cycles form fully connected components that can be merged into single nodes [12]. All nodes in such connected components have identical ancestry sets, yielding a one-to-one mapping from the NEM’s nodes to the ancestry sets, where the edges in the transitive closure of the NEM correspond to the set inclusion relations between ancestry sets. This can also be extended to the case of non-transitive Γ and the result on identifiability of non-transitive NEMs up to cycle reversals [25], the full discussion of which we omit here for brevity. Note that while set of ancestries characterizes the likelihood equivalence class, the posterior maximized by MC-EMiNEM would be, for example, higher for CSNEMs with fewer edges in Γ under a sparsifying edge prior.

4 Results

We have introduced the CSNEM model and showed how the CSNEM likelihood can be viewed as the likelihood of an NEM with $kn_{\mathcal{A}}$ actions learned from a modified differential expression log-likelihood ratio matrix $R\Phi$. Below, we use this transformation in conjunction with an existing NEM learning approach, MC-EMiNEM to learn CSNEMs and evaluate the ability of this approach to recover a CSNEM from data that is generated by a known multiple-context model in simulation. Finally, we present the results of learning a CSNEM from the results of knockout experiments on *S. cerevisiae* cells under NaCl stress, and discuss the biological significance of some patterns of context-specificity that are identified in the CSNEM.

4.1 Evaluation on Simulated Data

We performed simulations to evaluate our ability to infer CSNEMs from data. We generated data from mixtures of NEMs of varying size: we varied the size of the NEMs in the mixture to contain $n_{\mathcal{A}} = 3, 5, 10$, or 20 actions, and we varied the number of NEMs in the generating model from $j = 1$ to $j = 5$, inclusive, with $j = 1$ being equivalent to a simple NEM model. The number of effects $n_{\mathcal{E}}$ was fixed at 1000. We generated 30 mixtures corresponding to each configuration of j and $n_{\mathcal{A}}$, resulting in a total of 600 generated models. To generate each mixture, first we generated j random directed graphs G_1, \dots, G_j of $n_{\mathcal{A}}$ nodes, by drawing each of the possible $n_{\mathcal{A}}^2$ edges of the graph with a probability of 0.2 for graphs of size $n_{\mathcal{A}} < 20$ and a probability of 0.04 for graphs of size $n_{\mathcal{A}} = 20$ (with the higher edge density of 0.2 for 20 nodes, all nodes become reachable from all other nodes, yielding degenerate effect patterns where each effect is either affected by all actions, or by none). Next, for each effect, with probability 0.3 we attach it

nowhere, otherwise, we uniformly randomly attach it to one of the $n_{\mathcal{A}} \times j$ nodes in all of these graphs. Given these graphs and effect attachments, we infer which effects are reachable from each node, and compute the $n_{\mathcal{A}} \times n_{\mathcal{E}}$ binary effect matrix F^T , where $F_{as}^T = 1$ if and only if effect s is reachable from action a in any one of the j graphs. Next, we generate a log-odds matrix that represents a noisy measurement of this effect matrix by drawing from $\log \frac{\text{Beta}(\beta, 1)}{\text{Beta}(1, \beta)}$ for each ‘true’ cell and from $\log \frac{\text{Beta}(1, \beta)}{\text{Beta}(\beta, 1)}$ for each ‘false’ cell, with $\beta = 10$. This process generates the log-odds matrix R that we use as input to our learning method. Additionally, to examine the effect of noise in the measurement of effects on model inference, we generated log-odds matrices using $\beta = 1, 2, 5$ from the first 10 generating mixtures with $n_{\mathcal{A}} = 20, j = 1, 3, 5$.

Since in real-world applications we usually do not know how many contexts are truly needed to describe a process under study, we sweep through values of k ranging from 1 to 8, and learn a k -CSNEM for each value of k from each generated log-odds matrix. CSNEMs were learned using the MC-EMiNEM implementation in the `nem` R package, with the learned network taken from the end of a 20000 sample chain, the empirical Bayes step performed every 5000 steps, an acceptance sparsity prior of 0.5, and $kn_{\mathcal{A}}$ edges changed in every MCMC step (see Niederberger et al. [16] for details on how these settings are used in MC-EMiNEM). The edge-wise prior for permissible edges was set to 0.2.

We evaluate each k -CSNEM learned from each log-odds matrix both in terms of the ability of the CSNEM to accurately model which effects are differentially expressed in response to each action and in terms of the relationships inferred among actions. In the former case, we use the F-measure to quantify how well the effect matrix F of the learned CSNEM matches that of the generating CSNEM, with the interpretation that if an effect responds to an action in both the learned and the generating model, it is a true positive, if it doesn’t respond in the learned model but does in the generating model it is a false negative, if it doesn’t respond in either model it is a true negative, and if it responds in the learned model but not the generating model it is a false positive. Figure 3(a) shows the F-measures for learning the effect matrix across our simulations for the almost-noiseless case of $\beta = 10$. Figure 3(c) shows the F-measures for learning the effect matrix of a 20-action network from log-odds matrices generated with varying settings of β .

To compare the learned graph structures to the generating graph structures, we must first determine which contexts in the learned model correspond to which contexts in the generating model. For each action a in each model, we obtain a list of contexts that are distinguishable in terms of which actions are ancestors of the action a . We then match each of these contexts in each model to their best match in the other model. Each ancestor that the two contexts in the best match have in common counts as a true positive, each ancestor that appears in the context from the learned model but not in the context from the generating model counts as a false positive, and each ancestor that appears in the context from the generating model but not in the context from the true model counts as a false negative. We use these counts to summarize agreement between the

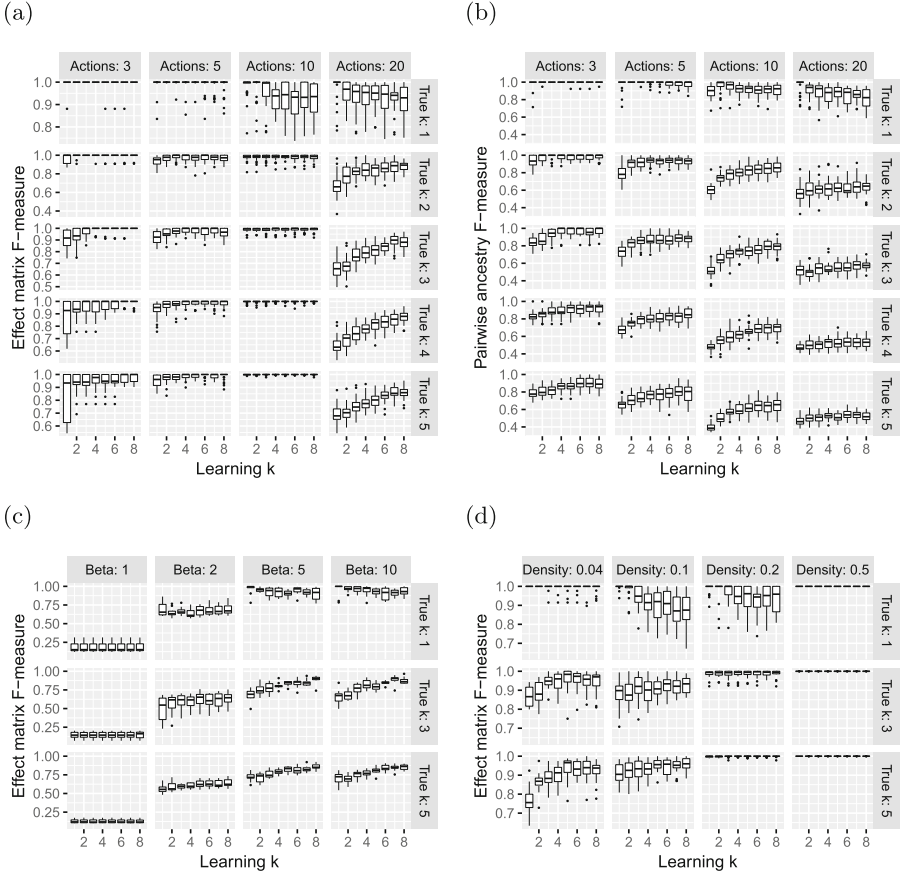


Fig. 3. Box plots of simulation F-measures. Each plot represents an aggregate of results from 30 random simulation replicates. Grid rows correspond to the number of contexts in the generating model, the x -axis in each of the grid cells indicates the number of contexts in the learned model, and the y -axis represents: (a) the F-measure of recovering the generating model’s effect matrix from the learned model across different sizes of action sets (grid columns) from log-odds matrices generated with $\beta = 10$, (b) the F-measure of recovering ancestry relationships, (c) the F-measure of learning the effect matrix of a 20-action network from log-odds matrices generated with varying settings of β (grid columns), and (d) the F-measure of learning the effect matrix from 10-action networks of varying density (grid columns) with log-odds generated using $\beta = 10$.

structures of two CSNEMs in terms of an F-measure which we call the pairwise ancestry F-measure. Figure 3(b) shows the pairwise ancestry F-measures across our simulations.

When the learned model is a plain NEM ($k = 1$), we see that as the generating model has more contexts, the recovery of both the effect and the ancestry pattern worsens (with the exception of the 10 actions case, examined below). This confirms that a CSNEM is necessary when multiple contexts are indeed in

play in the generating system. When the learned model has multiple contexts, even when the number of contexts in the learned model exceeds the number of contexts in the generating model, the approach does not seem to be susceptible to overfitting. This pattern holds as we increase noise (decrease β) in data generation.

At $n_{\mathcal{A}} = 10$ the NEM appears to recover the effects patterns well even when there are multiple contexts in the generating models, and we hypothesize that this is because of high connectivity in those ground truth networks: the average in-degree and out-degree of node is the product of one less than the number of actions times the edge density. We generated 20 mixtures for varying node densities (0.04, 0.1, 0.2, 0.5) with $j = 1, 3, 5$ contexts and $n_{\mathcal{A}} = 10$ nodes, and examined the effect-matrix F-measures across densities (Fig. 3(d)). Denser networks are perfectly recovered by single-context NEMs; this is likely because denser networks are more likely to lead to fully-connected transitive reductions, reducing the number of unique response patterns of effects, yielding data that is easier to capture in a simple NEM model. When the generating models are not too dense, CSNEMs are better than NEMs at recovering the effect patterns generated from multiple-contexts.

4.2 Application to NaCl Stress Response in *S. cerevisiae*

We apply our method to the exploration of NaCl stress response pathways in *S. cerevisiae*. We consider data obtained from a wild-type (WT) strain and 28 knockout strains. Transcript abundances were measured by microarray for each strain prior to NaCl treatment and 30 min after 0.7 M NaCl treatment. The data collection was described in detail in previous work [1, 8].

We are interested in how the gene knockouts change the cells' response to stress. Therefore, the actions \mathcal{A} in our model correspond to the knockouts. Since we use microarray data, the observations \mathcal{E} correspond to transcripts. The change in response is quantified as a change in log-fold-change. For each strain, we have the log-fold-change of transcript abundances in the sample 30 min after NaCl treatment as compared to the abundances in the sample prior to treatment. We then consider the difference between the log-fold-change in each knockout strain and that in the wild-type strain. To obtain the log-odds matrix R we use an empirical Bayes method to obtain log-posterior-odds of differential expression [10, 24] which is implemented in the `limma` R package [23]. Figure 4 shows the 3-CSNEM that was learned from the data.² The MC-EMiNEM settings used for learning both of these models are the same as those used for learning in the simulation experiments.

The inferred network captures many known and several new features of the yeast stress responsive signaling network. The Hog1 kinase is a master regulator of the osmotic stress response [15]. The CSNEM network correctly places Hog1 at the top of the hierarchy in paths with known co-regulators. For example, the

² An NEM learned from the data is at <https://github.com/sverchkov/mc-em-cs-nem/blob/master/recomb-2018-supplement/recomb-2018-supplement.pdf>.

network captures paths containing Hog1 and CK2 complex subunits Cka2 and Ckb1/2—Hog1 is known to interact physically with Cka2, and the two kinases regulate an overlapping set of genes [3]. The network also correctly predicts that the transcription factor Msn2 is regulated by Hog1, Pde2, and Snf1—all known regulators of Msn2 [9, 14, 18, 20]; yet a separate branch represents only Pde2 and Msn2, consistent with Pde2 playing a more significant role in regulating this transcription factor during salt stress [3]. Another example is seen in YGR122W, a poorly characterized protein required for processing the transcriptional repressor Rim101—the CSNEM correctly puts YGR122W and Rim101 in the same paths, with at least one regulatory branch shared with Hog1 control.

The CSNEM naturally produces groups of effects where each group comprises those effects (i.e. transcripts) that are reachable from contexts of actions in the graph. We examined the groups of effects in terms of Gene Ontology (GO) enrichments. Figure 5 shows a comparison of these enrichments to those obtained from grouping effects by the attachments from a learned NEM. The figure also shows a coarser split of the effects into groups based on CSNEM contexts: if an action was merged from two or more contexts in the single-network CSNEM representation, all the effects attached to it are considered reachable from both (or all three) contexts from which the action was merged. Each column in the figure corresponds to a GO term and each row corresponds to a combination of contexts or an action. A point in the figure indicates that the set of effects reachable from the context(s) or action was found to be significantly enriched for the GO term. Significance was defined according to a hypergeometric test with the Benjamini-Hochberg method used to control the false discovery rate at 0.05; only groups of five or more effects were considered for enrichment analysis.

A key advantage of our approach is that regulators can be represented in multiple pathways, capturing regulators that may have distinct roles in different cellular compartments or cell cycle phases. In fact, several of the GO terms for which the CSNEM effect groups are enriched are associated with subcellular localization and include transcripts encoding proteins localized to the nucleus, nucleolus, plasma membrane, endoplasmic reticulum, mitochondria, peroxisome, and cytoskeleton. The coarser split of effects by contexts also shows that there are clear divisions of localization across contexts in the CSNEM.

An interesting example of the benefits of the CSNEM approach is seen in its ability to capture the disparate signaling roles of the phosphatase Cdc14, a key regulator of mitotic progression in dividing cells [27]. Inactive Cdc14 is tethered to the nucleolus during much of the cell cycle but released upon mitosis to other subcellular regions where it dephosphorylates cyclins and other targets [28]. Separate from its role in the cell cycle, Cdc14 was recently linked to the stress response in yeast [2, 3], although its precise role is not clear.

The CSNEM network places Cdc14 in multiple pathways that capture the distinct functions of the phosphatase. One path represents an isolated connection of Cdc14 to a group of genes regulated by the cell cycle network. Many of these genes are known to be regulated by Cdc14 during normal cell cycle progression. But consistent with a second role in the stress response, Cdc14 is

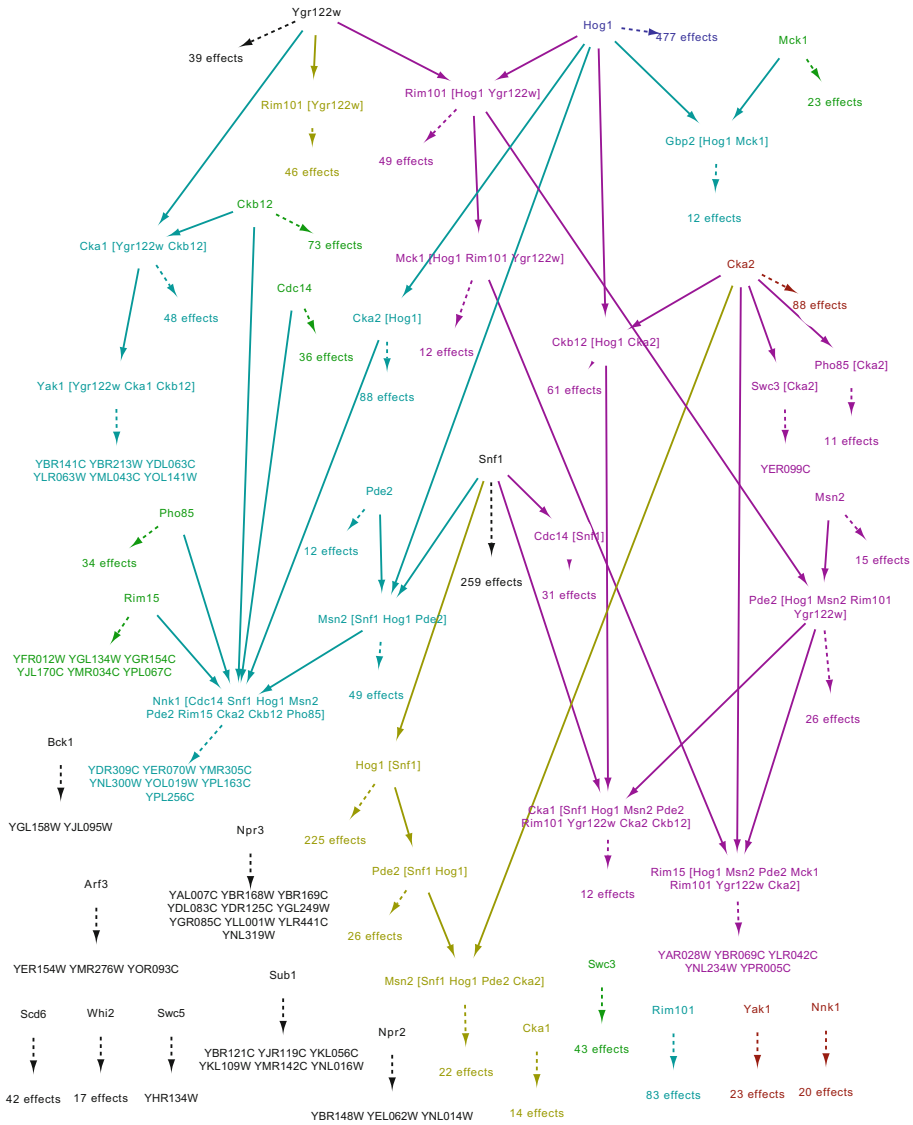


Fig. 4. The 3-CSNEM network learned from *S. cerevisiae* NaCl stress knockout microarray data. Action nodes and action-action edges are colored according to the NEM member in the mixture from which they came, in cyan, magenta, or yellow. Nodes that were merged because of identical ancestors in multiple mixture members are colored according to subtractive color mixing (cyan and magenta make blue, cyan and yellow make green, magenta and yellow make red, and all three make black). Effects are colored and grouped according to the actions to which they are attached. Where the number of effects in a group is fewer than 10, the effects are listed. Where it is 10 or more, the number of effects in the group is shown. Action-action edges are solid and action-effect edges are dashed. (Color figure online)

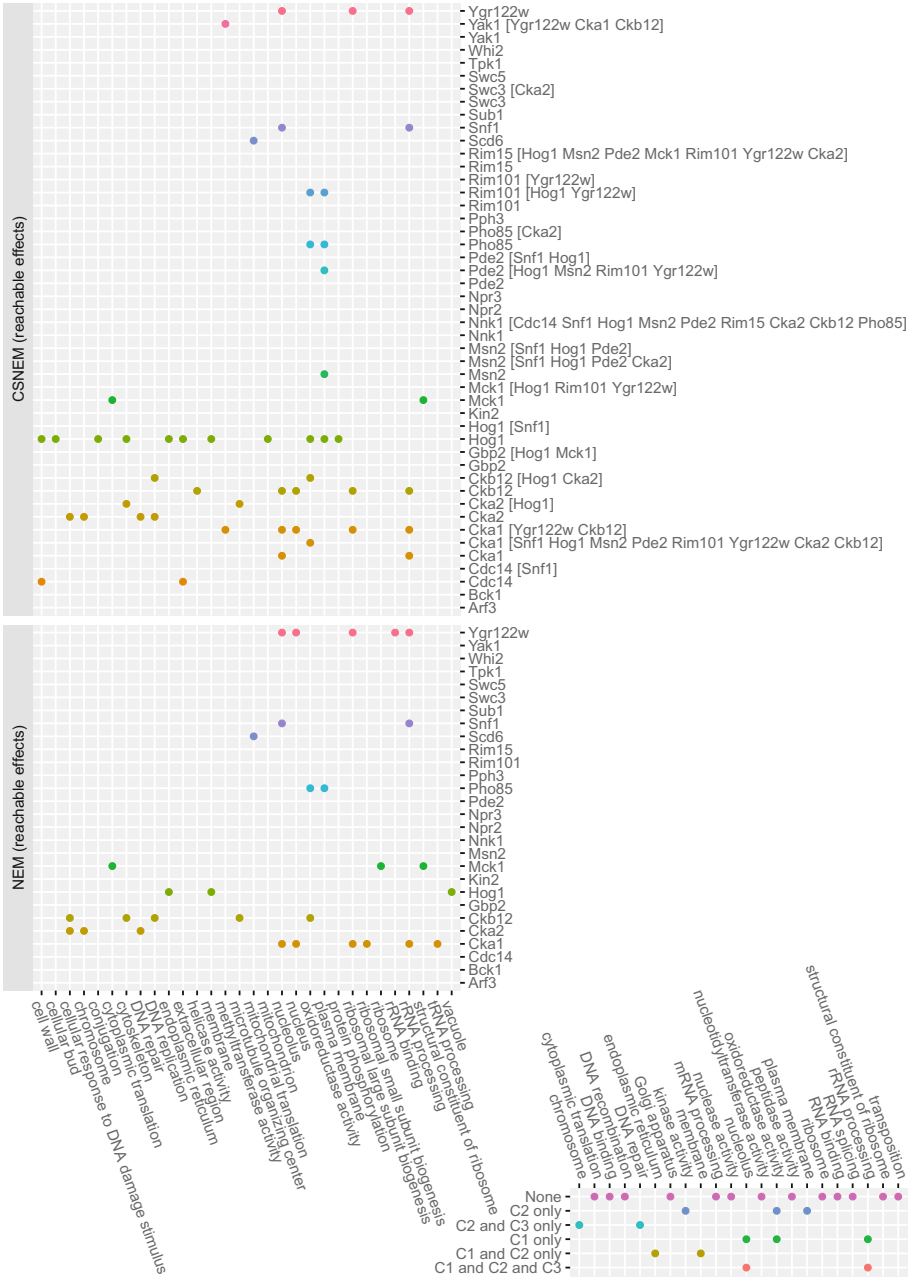


Fig. 5. Comparison of effect group GO term enrichments. Columns correspond to GO terms and rows correspond to actions in the NEM and CSNEM, and to possible combinations of contexts of the 3-CSNEM. A point indicates that a GO term was found to be significantly enriched. Points are colored by knockout in the NEM and CSNEM plots and by context in the context-membership plot. (Color figure online)

also nested in a path regulated by Snf1, a kinase that responds to both nutrient/energy restriction and osmotic stress resulting from salt treatment [29]. The Snf1-Cdc14 pathway is connected to 31 effectors that include genes induced by stress and related to glucose metabolism. Work from the Gasch Lab previously showed through genetic analysis that Snf1 and Cdc14 function, at least in part, in the same pathway during the response to salt stress [3]. Yet both Cdc14 and Snf1 have other functions in the cell, leading to the regulation of only partially overlapping gene sets. Thus, the CSNEM approach successfully captured this complex regulatory distinction for Cdc14 and Snf1.

5 Discussion

We have introduced CSNEMs, a generalization of NEMs which can explicitly model the different interactions that genes may have in different contexts. We have shown how a CSNEM can be viewed as a mixture of NEMs, and that the task of learning such a mixture can be cast as a single NEM-learning task with a modified data matrix and constrained action graph structure in which actions are replicated k times. Particularly, we took the approach of using a hard mixture where effects and actions are assigned to different contexts. A natural avenue for future investigation would be the exploration of soft-mixture approaches, which may prove more scalable for larger numbers of contexts and actions.

Applying our method to simulated data has shown that learning CSNEMs leads to good recovery of the effect patterns and ancestry relations that were present in the generating model. The results also show that a CSNEM is necessary when the generating model truly has multiple contexts, but slight over- or underestimation of the number of contexts does not seem to lead to overfitting. In practice, the correct number of contexts that a learned model should have is not known, and optimal selection of k is still an open problem that we plan to explore in future work. Existing approaches to model selection, such as a search for a plateau in likelihood or the use of model complexity measures such as AIC point to possible solutions to this problem.

Our analysis of a CSNEM network learned from *S. cerevisiae* NaCl-stress knockout microarray data revealed that the CSNEM does recover known regulatory patterns and moreover, captures known patterns of context-specificity in the genes under study. Analysis of GO term enrichments of the effects reachable from CSNEM nodes shows that many effect groups are associated with subcellular localization, a pattern even more evident in examining a coarser division of the effects, based on mixture contexts. We believe that localization may be one source of context-specificity that is relevant in many applications. The main motivation for developing CSNEMs was the observation that effect nesting may not be an appropriate assumption for some settings because of the context-specific nature of interactions that some genes can have, and perhaps more explicit modeling of contexts of interaction can lead to more faithful representations of the underlying biology.

Acknowledgments. We thank anonymous reviewers for many constructive comments. This research was supported by NIH/NLM grant T15 LM0007359, NIH/NIAID grant U54 AI117954, and NIH/NIGMS grant R01 GM083989.

References

- Berry, D.B., Gasch, A.P.: Stress-activated genomic expression changes serve a preparative role for impending stress in yeast. *Mol. Biol. Cell* **19**(11), 4580–4587 (2008)
- Breitkreutz, A., Choi, H., Sharom, J.R., Boucher, L., Neduva, V., Larsen, B., Lin, Z.-Y., Breitkreutz, B.-J., Stark, C., Liu, G.: A global protein kinase and phosphatase interaction network in yeast. *Science* **328**(5981), 1043–1046 (2010)
- Chasman, D., Ho, Y.-H., Berry, D.B., Nemeč, C.M., MacGilvray, M.E., Hose, J., Merrill, A.E., Lee, M.V., Will, J.L., Coon, J.J.: Pathway connectivity and signaling coordination in the yeast stress-activated signaling network. *Mol. Syst. Biol.* **10**(11), 759 (2014)
- Friedman, N., Linial, M., Nachman, I., Pe'er, D.: Using Bayesian networks to analyze expression data. *J. Comput. Biol.* **7**(3–4), 601–620 (2000)
- Fröhlich, H., Beißbarth, T., Tresch, A., Kostka, D., Jacob, J., Spang, R., Markowetz, F.: Analyzing gene perturbation screens with nested effects models in R and bioconductor. *Bioinformatics* **24**(21), 2549–2550 (2008)
- Fröhlich, H., Fellmann, M., Sültmann, H., Poustka, A., Beißbarth, T.: Large scale statistical inference of signaling pathways from RNAi and microarray data. *BMC Bioinform.* **8**(1), 1 (2007)
- Irrthum, A., Wehenkel, L., Geurts, P.: Inferring regulatory networks from expression data using tree-based methods. *PLoS ONE* **5**(9), e12776 (2010)
- Lee, M.V., Topper, S.E., Hubler, S.L., Hose, J., Wenger, C.D., Coon, J.J., Gasch, A.P.: A dynamic model of proteome changes reveals new roles for transcript alteration in yeast. *Mol. Syst. Biol.* **7**(1), 514 (2011)
- Lee, P., Cho, B.-R., Joo, H.-S., Hahn, J.-S.: Yeast Yak1 kinase, a bridge between PKA and stress-responsive transcription factors, Hsf1 and Msn2/Msn4. *Mol. Microbiol.* **70**(4), 882–895 (2008)
- Lönnstedt, I., Speed, T.: Replicated microarray data. *Statistica Sinica* **12**(1), 31–46 (2002)
- Markowetz, F., Bloch, J., Spang, R.: Non-transcriptional pathway features reconstructed from secondary effects of RNA interference. *Bioinformatics* **21**(21), 4026–4032 (2005)
- Markowetz, F., Kostka, D., Troyanskaya, O.G., Spang, R.: Nested effects models for high-dimensional phenotyping screens. *Bioinformatics* **23**(13), i305–i312 (2007)
- Markowetz, F., Spang, R.: Inferring cellular networks a review. *BMC Bioinform.* **8**(6), S5 (2007)
- Mayordomo, I., Estruch, F., Sanz, P.: Convergence of the target of rapamycin and the Snf1 protein kinase pathways in the regulation of the subcellular localization of Msn2, a transcriptional activator of STRE (Stress Response Element)-regulated genes. *J. Biol. Chem.* **277**(38), 35650–35656 (2002)
- Nadal, E., Posas, F.: Osmostress-induced gene expression—a model to understand how stress-activated protein kinases (SAPKs) regulate transcription. *FEBS J.* **282**(17), 3275–3285 (2015)

16. Niederberger, T., Etzold, S., Lidschreiber, M., Maier, K.C., Martin, D.E., Frohlich, H., Cramer, P., Tresch, A.: MC EMiNEM maps the interaction landscape of the mediator. *PLoS Comput. Biol.* **8**(6), e1002568 (2012)
17. Ohya, Y., Sese, J., Yukawa, M., Sano, F., Nakatani, Y., Saito, T.L., Saka, A., Fukuda, T., Ishihara, S., Oka, S.: High-dimensional and large-scale phenotyping of yeast mutants. *Proc. Nat. Acad. Sci. U.S.A.* **102**(52), 19015–19020 (2005)
18. Petrenko, N., Chereji, R.V., McClean, M.N., Morozov, A.V., Broach, J.R.: Noise and interlocking signaling pathways promote distinct transcription factor dynamics in response to different stresses. *Mol. Biol. Cell* **24**(12), 2045–2057 (2013)
19. Piano, F., Schetter, A.J., Morton, D.G., Gunsalus, K.C., Reinke, V., Kim, S.K., Kemphues, K.J.: Gene clustering based on RNAi phenotypes of ovary-enriched genes in *C. elegans*. *Curr. Biol.* **12**(22), 1959–1964 (2002)
20. Rep, M., Krantz, M., Thevelein, J.M., Hohmann, S.: The transcriptional response of *Saccharomyces cerevisiae* to osmotic shock Hot1p and Msn2p/Msn4p are required for the induction of subsets of high osmolarity glycerol pathway-dependent genes. *J. Biol. Chem.* **275**(12), 8290–8300 (2000)
21. Sadeh, M.J., Moa, G., Spang, R.: Considering unknown unknowns: reconstruction of nonconfoundable causal relations in biological networks. *J. Comput. Biol.* **20**(11), 920–932 (2013)
22. Siebourg-Polster, J., Mudrak, D., Emmenlauer, M., Rämö, P., Dehio, C., Greber, U., Fröhlich, H., Beerenwinkel, N.: NEMix: single-cell nested effects models for probabilistic pathway stimulation. *PLoS Comput. Biol.* **11**(4), e1004078 (2015)
23. Smyth, G.K.: Limma: linear models for microarray data. In: Gentleman, R., Carey, V.J., Huber, W., Irizarry, R.A., Dudoit, S. (eds.) *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Springer, New York (2005). <https://doi.org/10.1007/0-387-29362-0-23>
24. Smyth, G.K.: Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**(1), 3 (2004)
25. Tresch, A., Markowetz, F.: Structure learning in nested effects models. *Stat. Appl. Genet. Mol. Biol.* **7**(1), 9 (2008)
26. Vaske, C.J., House, C., Luu, T., Frank, B., Yeang, C.-H., Lee, N.H., Stuart, J.M.: A factor graph nested effects model to identify networks from genetic perturbations. *PLoS Comput. Biol.* **5**(1), e1000274 (2009)
27. Weiss, E.L.: Mitotic exit and separation of mother and daughter cells. *Genetics* **192**(4), 1165–1202 (2012)
28. Wurzenberger, C., Gerlich, D.W.: Phosphatases: providing safe passage through mitotic exit. *Nat. Rev. Mol. Cell Biol.* **12**(8), 469–482 (2011)
29. Ye, T., Elbing, K., Hohmann, S.: The pathway by which the yeast protein kinase Snf1p controls acquisition of sodium tolerance is different from that mediating glucose regulation. *Microbiology* **154**(9), 2814–2826 (2008)