

TEAK: Topology Enrichment Analysis framework for detecting activated biological subpathways

Thair Judeh¹, Cole Johnson², Anuj Kumar^{2,*} and Dongxiao Zhu^{1,*}

¹Department of Computer Science, Wayne State University, 5057 Woodward Avenue, Detroit, MI 48202, USA and ²Department of Molecular, Cellular, and Developmental Biology, University of Michigan, 830 North University Avenue, Ann Arbor, MI 48109, USA

Received August 6, 2012; Revised November 11, 2012; Accepted November 14, 2012

ABSTRACT

To mine gene expression data sets effectively, analysis frameworks need to incorporate methods that identify intergenic relationships within enriched biologically relevant subpathways. For this purpose, we developed the Topology Enrichment Analysis framework (TEAK). TEAK employs a novel in-house algorithm and a tailor-made Clique Percolation Method to extract linear and nonlinear KEGG subpathways, respectively. TEAK scores subpathways using the Bayesian Information Criterion for context specific data and the Kullback-Leibler divergence for case-control data. In this article, we utilized TEAK with experimental studies to analyze microarray data sets profiling stress responses in the model eukaryote *Saccharomyces cerevisiae*. Using a public microarray data set, we identified via TEAK linear sphingolipid metabolic subpathways activated during the yeast response to nitrogen stress, and phenotypic analyses of the corresponding deletion strains indicated previously unreported fitness defects for the *dpl1Δ* and *lag1Δ* mutants under conditions of nitrogen limitation. In addition, we studied the yeast filamentous response to nitrogen stress by profiling changes in transcript levels upon deletion of two key filamentous growth transcription factors, *FLO8* and *MSS11*. Via TEAK we identified a nonlinear glycerophospholipid metabolism subpathway involving the *SLC1* gene, which we found via mutational analysis to be required for yeast filamentous growth.

INTRODUCTION

With the exponential growth of high-dimensional gene expression data, biologists need versatile tools at their

disposal to efficiently extract important biological insights from their data. Clearly, many pathway resources are now available including KEGG (1,2), Reactome (3,4) and Biocarta (<http://www.biocarta.com>). The increasing availability of high-throughput gene expression data and high-fidelity pathways has led to an evolution in bioinformatics analysis from the analysis of single genes to gene sets and now to subpathways.

A classical approach for analyzing high-dimensional gene expression data is to use an over representation approach (ORA). Many methods exist (5) such as Pathway Processor (6), PathMAPA (7), PathwayMiner (8), ArrayXPath (9), GenMAPP (10) and Low Variance Pathway Predicator (11). In an ORA approach, one typically analyses the number of differentially expressed genes within a pathway gene set against the number of genes expected to be found by chance. While these previous approaches are useful, they may fail to take into account the inherent regulatory relationships found in biological pathways among the different genes. Biological pathways are effectively reduced to sets of gene sets using an ORA approach. In other words, a rich source of information, namely pathway topologies, remains untapped and unused.

SPIA (12) and Paradigm (13) are more recent tools that use whole pathway topologies. Whole pathways, however, may have different subpathways activated in response to a biological context. Thus, their subpathways may more accurately represent the underlying biological phenomena. One approach, SubpathwayMiner (SM) (14), extracts *k*-clique subpathways, i.e. the distance between any two nodes in a subpathway is not larger than *k*. Another approach (15) extracts linear subpathways using a depth-first search (DFS) algorithm. These approaches, however, may be limited since a hypergeometric test is used for SM and a Euclidean-based measure is used for the latter approach. Such approaches may fail to fully capture the underlying topological information present in biological pathways as permuting the structure of the subpathways using either approach will yield the same

*To whom correspondence should be addressed. Tel: +1 313 577 3104; Fax: +1 313 577 6868; Email: dzhu@wayne.edu
Correspondence may also be addressed to Anuj Kumar. Tel: +1 734 647 8060; Fax: +1 734 647 0884; Email: anujk@umich.edu

results. Currently, frameworks that exploit subpathway topologies have not been extensively studied.

The need to better identify pathway and subpathway topologies enriched within large-scale data sets is evident from studying the budding yeast *Saccharomyces cerevisiae*. As a prototypical eukaryote, cellular processes in yeast occur through complex sets of signaling modules. Genomic methods have been applied extensively in yeast to profile the signaling changes underlying these processes, but challenges remain toward effectively identifying pathways and subpathways within the gene lists resulting from these studies. In standard laboratory strains of *S. cerevisiae*, nutritional stress is mediated through several key signaling modules that act in concert with sets of genes responsive to a specific stimulus (16,17). In particular, nitrogen starvation is a common form of nutrient stress for the budding yeast, and upon nitrogen stress, yeast cells initiate both general and stimulus-specific responses. Nitrogen stress initiates a general response referred to as the environmental stress response characterized by stereotypic increases and decreases in mRNA levels for ~600 genes in response to a broad set of environmental/nutritional stresses (16). Nitrogen deprivation also induces stress-specific increases in the mRNA levels of ~300 additional genes, combined with post-transcriptional regulatory mechanisms to upregulate autophagy, alter endocytosis and downregulate protein synthesis (16,18).

Beyond the regulatory responses discussed earlier, filamentous strains of *S. cerevisiae* initiate a complex growth transition in response to nitrogen stress, forming pseudohyphal filaments of elongated and connected cells that extend outward and downward from a yeast colony (19,20). This filamentation is thought to be a foraging mechanism enabling nonmotile yeast to scavenge for nutrients. Interestingly, related processes of hyphal development are required for virulence in the opportunistic human fungal pathogen *Candida albicans* (21,22). In *S. cerevisiae*, the pseudohyphal growth response is regulated by at least four signaling pathways: the target-of-rapamycin kinase network, the Ras/cAMP-dependent protein kinase A (PKA) pathway, the Snf1p kinase pathway and the Kss1p mitogen-activated protein kinase (MAPK) pathway (20,23). The Ste12p/Tec1p transcription factor complex acts downstream of the Kss1p MAPK pathway, and the Flo8p transcription factor is phosphorylated and activated by the Tpk2p catalytic subunit of PKA (23,24). Both factors regulate transcription of the *MUC1/FLO11* gene encoding a Glycosylphosphatidylinositol (GPI)-anchored protein important for the enhanced cell-cell adhesion observed during filamentous growth (25,26). The filamentous growth response, however, is extensive and encompasses several other known transcriptional regulators, such as Mss11p, Phd1p and Dig1/2p, and hundreds of downstream genes and pathways (27,28).

While these works have identified key regulatory modules that function to transduce conditions of nitrogen stress into intracellular signals that affect cell growth/shape, the full scope of the signal transductions involved in the core regulatory modules has yet to be

determined. The problem is likely to be too complicated for experimental methods alone, and we believe that an integration of experimental and computational methods will be necessary to identify new subpathways within the filamentous growth network.

Thus, to detect active subpathways underlying biological processes, we developed the innovative Topology Enrichment Analysis framework (TEAK), which is freely available at <http://code.google.com/p/teak/> for Windows and MAC. TEAK uses an in-house developed graph traversal algorithm to extract all root to leaf linear subpathways of a given pathway while it uses a tailor-made Clique Percolation Method (CPM) (29,30) for nonlinear subpathways. For subpathways activated under a specific context or condition, e.g. a single data matrix corresponding to time series data or a set of samples corresponding to relevant mutants, TEAK deploys the Bayesian Information Criterion (BIC) (31) implemented in the Bayes Net Toolbox (32) to fully capture the topological information and regulatory relationships inherent in both linear and nonlinear subpathways. For differential subpathways between case and control conditions, TEAK instead uses the Kullback-Leibler (KL) divergence of two Bayesian networks, i.e. a case subpathway and a control subpathway, transformed into their multivariate Gaussian forms to score each subpathway. Thus, TEAK provides an innovative view of the data from a fresh angle allowing users to visualize a subpathway within its respective parent pathway as illustrated in Supplementary Figure S1.

Here, we utilized TEAK to analyze DNA microarray data profiling changes in transcript levels for the yeast genome during nitrogen stress responses. To validate that TEAK is effective in identifying biologically relevant pathways and subpathways, we analyzed transcriptional profiles of yeast responses to varying conditions of nitrogen stress (16). As a result, we identified a set of subpathways within the sphingolipid metabolic pathway that has not been phenotypically characterized previously for fitness defects during nitrogen stress, and through experimental studies, we report that the deletion of the *DPL1* and *LAG1* genes within these subpathways do confer a cell growth defect on low-nitrogen medium. Furthermore, we performed a microarray experiment and used TEAK to identify changes in mRNA levels upon deletion of two transcription factors, Flo8p and Mss11p, essential for the yeast filamentous growth response. Via TEAK we observed decreased transcript levels for a subset of genes within a subpathway involved in glycerophospholipid metabolism upon deletion of the *FLO8* and *MSS11* genes. By deletion analysis and phenotypic profiling, we report that the *SLC1* gene within this subpathway is required for yeast filamentous growth and that its deletion results in decreased expression of a *LacZ* reporter driven by promoter elements responsive to the filamentous growth MAPK pathway. Collectively, these laboratory studies validate TEAK's utility in analyzing large-scale data sets across multiple conditions. They also indicate the broad scope of the yeast filamentous growth response for which wild-type activity of the glycerol 3-phosphate pathway for *de novo* biosynthesis of phosphatidic acid is required.

MATERIALS AND METHODS

Yeast strains and growth conditions

Strains used for the analysis of sphingolipid metabolism during nitrogen stress were of the nonfilamentous S288c genetic background derived from BY4742 (*MAT α his Δ 1 leu2 Δ 0 lys2 Δ 0 ura3 Δ 0*) (33). For filamentous growth studies, the strains Y825 and HLY337 were derived from the Σ 1278b background (19,34). The haploid strain Y825 has the genotype *MAT α ura3-52 leu2 Δ 0*, and the haploid strain HLY337 has the genotype *MAT α ura3-52 trp1 Δ* . Standard growth media consisted of YPD prepared using 1% yeast extract, 2% bacto-peptone and 2% glucose. Synthetic media was prepared using 0.17% yeast nitrogen base (YNB) without amino acids and ammonia, 2% glucose and 5 mM ammonium sulfate (35). Hyperosmotic sensitivity was assayed using YPD supplemented with sterile 1 M Sorbitol. Nitrogen deprivation and filamentous growth phenotypes were assayed using Synthetic Low Ammonium Dextrose (SLAD) medium consisting of 0.17% YNB, 2% dextrose, 50 μ M ammonium sulfate and supplemented with the necessary amino acids (36). For plates autoclaved 2% agar was added to the media.

Yeast gene deletions and transformations

Deletion mutants were constructed in strains Y825 and HLY337 by using a one-step polymerase chain reaction (PCR)-based gene disruption strategy (37,38) with the G418 resistance cassette from plasmid pFA6a-KanMX6 (39). After confirming the haploid mutants via PCR, the strains were allowed to mate on YPD+G418 plates for ~20 h at 30°C. Mated cells were then streaked on SC-Trp-Leu plates to select for Y825 \times HLY337 diploids. Yeast transformations were performed according to standard lithium acetate-mediated protocols (40) with modifications (41–43).

Microarray experiments and analysis

After culturing the yeast strains as described earlier, RNA was prepared under standard protocols using the Poly(A) Purist kit (Ambion, Austin, TX, USA). RNA concentration and purity were determined spectrophotometrically and by gel electrophoresis. Microarray hybridization was performed with the Yeast Genome S98 Array using standard protocols (Affymetrix, Inc., Santa Clara, CA, USA). All microarray experiments were performed in quadruplicate for each strain as described by (27,44). For preprocessing the data for use by TEAK, we used GCRMA (45). To detect differentially expressed genes, we used significance analysis of microarrays (SAM) (46). For all deletion comparisons in this study, we used SAM's two-class unpaired analysis function with significance thresholds selected so that the corresponding false discovery rate was 0.

Growth assays

Yeast strains were inoculated in 5 ml YPD and incubated overnight at 30°C with constant shaking (250 rpm). Cultures were diluted the following morning in 5 ml YPD and allowed to grow until the culture reached an

OD₆₀₀ of ~0.6–0.8. A 1-ml aliquot of each strain was washed once with sterile water and then re-suspended in sterile water such that all cell titers were equal. Each strain was then diluted via five 10-fold serial dilutions, and 6 μ l of each dilution was spotted on YPD, YPD+1M Sorbitol, or SLAD media. The spotted plates were allowed to dry at room temperature for ~15 min before being placed at 30°C for 48 h. Yeast strains were inoculated in 5 ml YPD and incubated overnight at 30°C with constant shaking (250 rpm). Cultures were back diluted the following morning in 5 ml YPD and allowed to grow until the culture reached an OD₆₀₀ of ~0.6–0.8. A 1-ml aliquot of each strain was washed once with sterile water and re-suspended in 1 ml sterile water. Strains were triplicate loaded into either YPD or SLAD media within a 96-well plate such that the final concentration of each strain was an OD₆₀₀ of ~0.02 (~40 \times dilution) in a total volume of 100 μ l. Growth was measured optically every 5 min for a period of 30 h with constant shaking at 30°C in a Synergy HT Multi-Mode microplate reader (BioTek, Winooski, VT, USA). Time point measurements represent the average of the three replicates normalized by the average of the triplicate blank controls.

β -Galactosidase assays for *lacZ* activity

Yeast strains were inoculated in 5 ml SC-URA and incubated overnight at 30°C with constant shaking (250 rpm). Cultures were diluted the following morning in 5 ml SC-URA or SLAD media and allowed to grow until the culture reached an OD₆₀₀ of ~0.5–1. β -Galactosidase assays were performed using the Thermo Scientific Yeast β -Galactosidase Assay Kit (Thermo Fisher Scientific, Rockford, IL, USA) according to the manufacturer's protocol except that the β -galactosidase reaction was allowed to continue for a set time of 30 min for each sample before reading the absorbance. Measurements represent the average of three replicates.

Method overview

Figure 1 outlines TEAK. Using the KEGG API (1,2), TEAK first fetches all metabolic and nonmetabolic pathways for the organism under study. TEAK extracts a subset of the KEGG pathways consisting of gene products and/or complexes of gene products as nodes. For edges TEAK extracts all KEGG enzyme–enzyme relations, protein–protein interactions and gene expression interactions to create a set of unweighted adjacency matrices to represent the KEGG pathways (please refer to Supplementary Table S1 for more details). This process, including the extraction of linear and nonlinear subpathways, occurs only once per organism or as needed, and its results are pre-computed and included by default for *Homo sapiens*, *Mus musculus* and *S. cerevisiae*.

Subpathway extraction

Subpathways play a major role in biological processes since only part of a pathway may be activated at a specific time given an underlying condition. Often times a biological condition may be controlled by several specific subpathways, but the subpathways' contribution

may be obscured by their parent pathways. As such, subpathway extraction is an important component of TEAK. TEAK extracts two types of subpathways: linear and nonlinear. Linear subpathways are represented by root to leaf linear paths of a pathway, whereas nonlinear subpathways are represented by a union of adjacent and overlapping feed-forward loops.

Algorithm 1: Linear subpathways

```

1: Input: An unweighted, directed graph  $G$ 
2: Output: All root to leaf linear subpathways
3: Remove all self-loops from  $G$ 
4: Convert the graph  $G$  into the set of adjacency lists  $A$ 
5: Set the visit vector  $V$  of size  $|G|$  to false
6: Find the roots  $R$  and leaves  $L$  of graph  $G$ 
7: for  $j = 1, \dots, |R|$  do
8:   Add root  $r_j$  to the Stack  $S$ 
9:   Set  $V[r_j]$  to true
10:  while  $S$  is not empty do
11:    Let the node  $n$  be the top element of  $S$ 
12:    Remove every child  $c$  of  $n$  from the adjacency
13:    list  $A[n]$  that has  $V[c]$  as true
14:    if  $A[n]$  is not empty then
15:      Pop a child  $c$  of node  $n$  from  $A[n]$ 
16:      Set  $V[c]$  to true
17:      Add node  $c$  to  $S$ 
18:    else
19:      if  $n$  is a member of  $L$  then
20:        Append the contents of  $S$  as a new
21:        subpathway to the final output
22:      else
23:        Reconstruct  $A[n]$  using the graph  $G$ 
24:      end
25:      Pop a node from  $S$ 
26:      Set  $V[n]$  to false
27:    end
28:  end
  
```

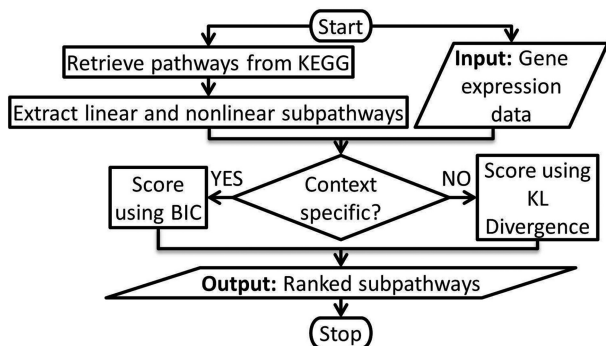


Figure 1. TEAK overview. TEAK requires gene expression data using one of the many label systems supported by KEGG including Entrez, NCBI-GI and ORF (for *S. cerevisiae*) as input. By default TEAK includes a set of subpathways for *H. sapiens*, *M. musculus* and *S. cerevisiae*. For other organisms, the extraction of subpathways from the KEGG pathways needs to be done once or as needed. For context specific data, TEAK uses the BIC to score the Bayesian network fitted for each subpathway. For case-control data, TEAK first fits two Bayesian networks, a case network and a control network, for each subpathway. After transforming each Bayesian network into its equivalent multivariate Gaussian form, TEAK then scores each pair of transformations per subpathway using the KL divergence. Via TEAK's GUI (Supplementary Figure S2), a user can then display a subpathway highlighted in its parent pathway as seen in Supplementary Figure S1.

Algorithm 1, based on the algorithm *Network2GeneSets* (47,48), extracts root to leaf linear paths or subpathways from the directed edges of the KEGG nonmetabolic pathways. A root r has zero incoming links and a positive number of outgoing links, whereas a leaf l has a positive number of incoming links and zero outgoing links (please refer to Supplementary Figure S3 for an example of Algorithm 1). Compared with breadth-first search (BFS) and DFS, whose 'vanilla' implementations may be found in (49), TEAK extracts all linear subpathways as seen in Figure 2.

From a biological perspective, root to leaf subpathways are important as they represent one of the possible routes taken from the beginning of a pathway to its end. Furthermore, in terms of signaling pathways, we hypothesize that root to leaf subpathways effectively model signal transduction events. In one signal transduction paradigm, a growth factor binds to a cell membrane receptor that then propagates a signal via intracellular signaling pathways to reach the nucleus and cause a change in gene expression (50). Signal transduction pathways regulate cell proliferation, survival, motility and differentiation (50) and play vital roles in cancer (51), mammalian associative conditioning (52) and cellular response to stress (53). For the KEGG signaling pathways, roots may be growth factors. For example, the epidermal growth factor is a root of the *H. sapiens* MAPK signaling pathway.

As root to leaf linear subpathways may not effectively model the underlying biological condition under study for all of the pathways extracted, a different type of subpathway is needed. In this case, we chose to adapt and slightly tweak the CPM (29,30). TEAK implements CPM with one notable change in which feed-forward loops, which are directed cliques of size three, are extracted instead of the maximal cliques of a pathway. We believe the focus on feed-forward loops is justified since the feed-forward loop is one of the most common motifs in biological networks (54). Nevertheless, our method shares many of the advantages found in the original CPM: (i) genes may participate in multiple subpathways whereas most other approaches extract mutually exclusive subpathways. Biologically, the former approach may be more relevant as a gene may regulate multiple biological processes. (ii) There exists no gene or link whose removal would disjoin a subpathway, i.e. no single cut-node or cut-link exists in a subpathway.

The subpathway algorithms and the Bayesian networks used by TEAK are only applicable to directed networks. For undirected networks, TEAK first extracts the longest shortest paths that are not contained within other shortest paths. For example, for the set of shortest paths $\{1 \leftrightarrow 2 \leftrightarrow 3, 4 \leftrightarrow 3 \leftrightarrow 2, 1 \leftrightarrow 2 \leftrightarrow 3 \leftrightarrow 4\}$, TEAK extracts the shortest path $1 \leftrightarrow 2 \leftrightarrow 3 \leftrightarrow 4$ that contains the other two shortest paths. For directionality, TEAK then selects one of two directed linear paths that most closely resemble root to leaf linear subpathways. For the shortest path $1 \leftrightarrow 2 \leftrightarrow 3 \leftrightarrow 4$, the two directed linear paths are $1 \rightarrow 2 \rightarrow 3 \rightarrow 4$ and $4 \rightarrow 3 \rightarrow 2 \rightarrow 1$. They may be obtained by fixing one terminal end of the shortest path as a root and the other terminal end of the shortest path as

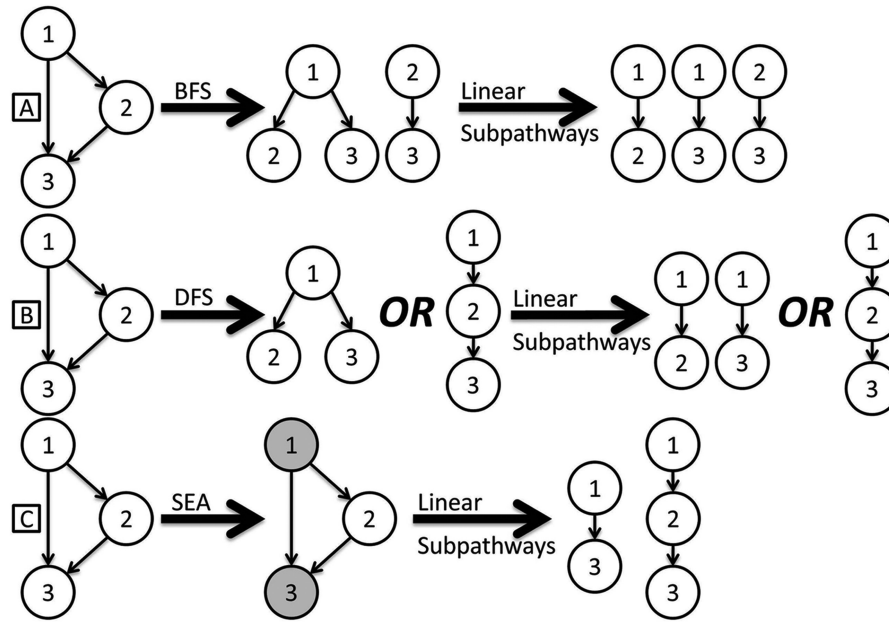


Figure 2. Extracting linear subpathways from a feed-forward loop. (A) By extracting shortest paths via BFS, the linear subpathways 1 → 2 and 2 → 3 are not root to leaf linear subpathways of the feed-forward loop. (B) DFS produces two different trees with two different sets of outputs, respectively. For the first set of outputs, 1 → 2 is not a root to leaf linear subpathway. For the second set of outputs, the linear subpathway 1 → 3 is missing. For more complex networks, the number of DFS trees may also be exponential such that combining and pruning their outputs may not be feasible. (C) TEAK properly extracts all root to leaf linear subpathways of the feed-forward loop.

a leaf as further illustrated in Figure 3. Since the pair of directed linear subpathways is I-equivalent, i.e. they can be represented by the same set of conditional independence assertions (55), either one can serve as a ‘root to leaf’ subpathway corresponding to the original shortest path. Then, for extracting directed nonlinear subpathways from an undirected pathway, TEAK extracts cliques of size 3. It then selects one of six I-equivalent feed-forward loops corresponding to the clique as seen in Figure 3.

Subpathway ranking

To rank the linear and nonlinear subpathways, TEAK first uses the Bayes Net Toolbox (32) to fit a context specific Gaussian Bayesian network for each subpathway. Briefly, a Gaussian Bayesian network is a Bayesian network in which all of its nodes are linear Gaussians. In other words, for a continuous node Y with m continuous parents X_1, \dots, X_m , the Conditional Probability Distribution of Y is

$$p(Y|x_1, \dots, x_m) = N(\beta_0 + \beta_1 x_1 + \dots + \beta_m x_m; \sigma^2) \tag{1}$$

where β_0, \dots, β_m are the regression coefficients and σ^2 is the variance (55). It should be noted that Bayesian networks are only applicable to directed acyclic graphs. However, as all the subpathways extracted by TEAK are directed acyclic graphs, this limitation is not applicable. We also further justify our choice of Bayesian networks since they have already been used to discover networks from gene expression data (56,57).

For scoring TEAK takes one of two approaches. For context specific data, TEAK fits a single Bayesian network for each subpathway and uses the BIC (31) implemented

in the Bayes Net Toolbox for scoring each Bayesian network. Briefly, the Bayes Net Toolbox implements BIC as

$$\text{Score}_{\text{BIC}} = \log P(D|\hat{\theta}) - 0.5 d \log N \tag{2}$$

where D corresponds to the gene expression data, $\hat{\theta}$ corresponds to the maximum likelihood estimate of the parameters used to represent the linear Gaussian node, d is the number of parameters and N is the number of samples in the gene expression data. As BIC is decomposable, i.e. each node’s score is calculated individually and then summed to return the final score, we normalize each subpathway by its number of nodes so that the scores are comparable.

Given that most researchers nowadays have case-control data, TEAK also supports case-control data, i.e. two data matrices, by fitting two Bayesian networks, one for each data matrix. It then transforms each context specific Bayesian network into its equivalent multivariate Gaussian form [please refer to the appendix of (58) for details and (59,60) for examples]. TEAK then calculates the KL divergence of the case multivariate Gaussian, q , from the control multivariate Gaussian, p , as

$$\begin{aligned} \text{KL}(q||p) = & \frac{1}{2} \log \frac{|\Sigma_p|}{|\Sigma_q|} + \frac{1}{2} \text{Tr}(\Sigma_p^{-1} \Sigma_q) \\ & + \frac{1}{2} (\mu_q - \mu_p)^T \Sigma_p^{-1} (\mu_q - \mu_p) - \frac{h}{2} \end{aligned} \tag{3}$$

where μ is the mean vector, Σ is the covariance matrix, $|\Sigma|$ is the determinant of the covariance matrix, Tr is the trace of a matrix and h is the number of nodes [please refer to the appendix of (61) for more details].

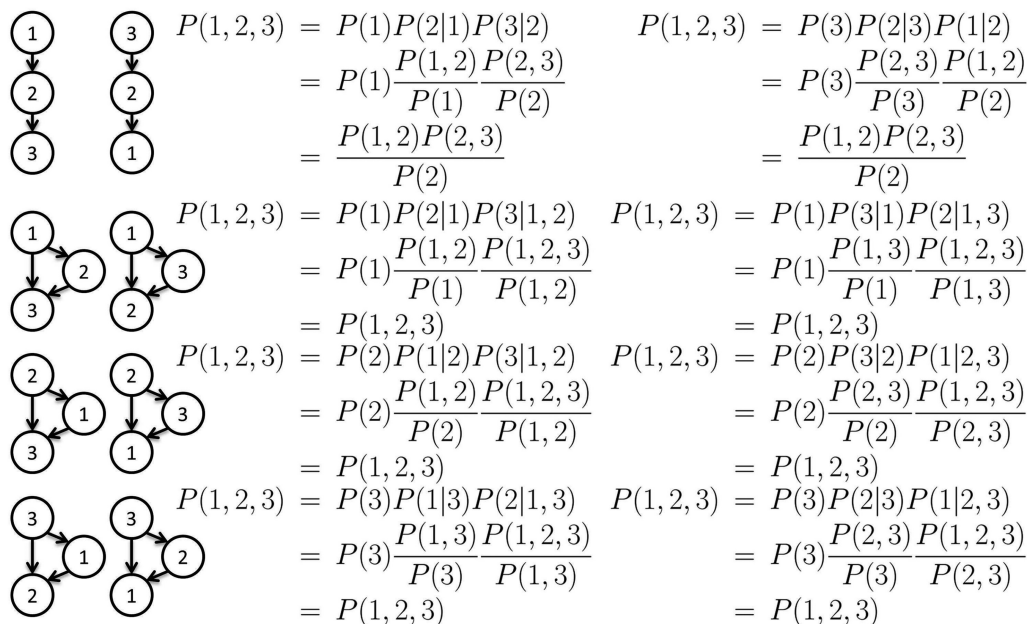


Figure 3. (Top) For the undirected linear graph $1 \leftrightarrow 2 \leftrightarrow 3$, there are two possible directed graphs starting at one of the two ends, namely $1 \rightarrow 2 \rightarrow 3$ and $3 \rightarrow 2 \rightarrow 1$. Both linear graphs are I-equivalent, and the result holds in general for linear graphs of arbitrary sizes. **(Bottom)** The six possible feed-forward loops for a clique of size 3. All six feed-forward loops are I-equivalent.

RESULTS

Wild-type sphingolipid metabolism is required for efficient yeast cell growth during nitrogen stress

The yeast cell response to nitrogen stress is extensive, encompassing changes at the mRNA level for hundreds of genes. Toward this point, Gasch *et al.* (16) profiled transcriptional changes in yeast upon exposure to conditions of nitrogen stress over 10 time points ranging from 30 min to 5 days. To consider specific pathways differentially regulated at the level of transcription during nitrogen stress, we analyzed these microarray data sets using TEAK for subpathways collectively showing significant changes in mRNA levels of constituent genes across the time points.

Here, we focused our studies on linear subpathways since genes within a linear subpathway are more likely to share a common phenotype and, consequently, are more amenable to genetic analysis. By TEAK a sphingolipid synthesis metabolic subpathway was highly ranked (within the top 25) with higher-ranking subpathways corresponding to processes that were already known to be important for cell growth during nitrogen stress (e.g. GPI-anchor biosynthesis and the yeast cell cycle). In eukaryotes, sphingolipids are an abundant membrane component, filling a structural role in membrane support and an increasingly well-studied role as bioactive compounds involved in signal transduction; sphingolipid functions in yeast are reviewed in (62). We were particularly intrigued by the identification of the sphingolipid metabolic pathway by TEAK since previous studies have reported that the modulation of sphingolipid synthesis impacts yeast lifespan and the regulation of the yeast high osmolarity stress-responsive pathway (63,64). Consequently, we chose to use TEAK to investigate subpathways of the

sphingolipid metabolism pathway that may contribute to yeast cell survival during nitrogen stress. By sampling subpathways strictly within the sphingolipid metabolic pathway, TEAK identified a gene set encompassing *LCB3*, *LCB5*, *YDC1*, *LAG1* and *DPL1*; these genes were consistently identified in the subpathways extracted by TEAK. We subsequently analyzed these genes further for a possible role in the yeast nitrogen stress response.

The genes selected for phenotypic analysis include *LCB5*, which along with *LCB4* encode long chain base kinases that catalyze formation of the phosphorylated long chain bases (LCBPs) dihydrosphingosine 1-phosphate and phytosphingosine 1-phosphate (65). *Lcb3p* is a phosphatase that dephosphorylates LCBPs (66). *YDC1* encodes an alkaline ceramidase that is specific for dihydroceramide (67). *Dpl1p* is an LCBP lyase that cleaves LCBPs at the C_{2-3} bond to yield a long chain aldehyde and ethanolamine phosphate (68). *Lag1p* is a component of ceramide synthase, synthesizing ceramide from C_{26} (acyl)-coenzyme A and dihydrosphingosine or phytosphingosine (69). Interestingly, LCBs, dihydrosphingosine and phytosphingosine have been identified as important second messengers in signaling pathways that regulate cellular responses to heat stress, and LCBPs, phytoceramide and additional sphingolipid metabolic intermediates contribute to the regulation of cell growth (70,71). It is important to note that a role in yeast cell growth under conditions of nitrogen stress had not been determined for these genes.

Consequently, we analyzed homozygous diploid yeast strains singly deleted for each of these genes in a nonfilamentous genetic background to assess their ability to grow under conditions of nitrogen stress. As indicated in Figure 4, deletion of *DPL1* and *LAG1* yields a strong reduction in cell growth under conditions of nitrogen

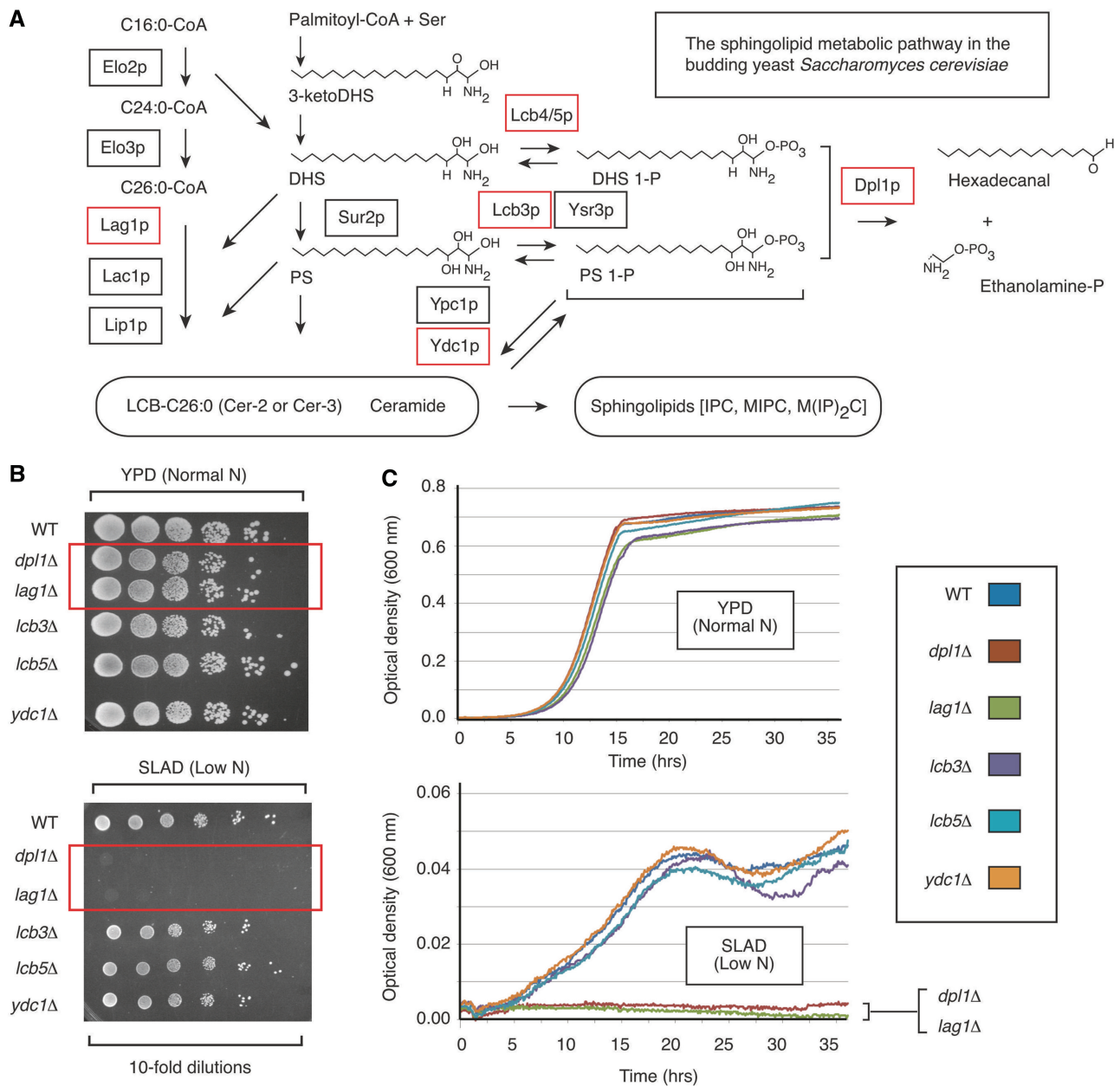


Figure 4. (A) A simplified representation of the sphingolipid metabolism pathway. The shaded genes were selected from TEAK for phenotypic analysis. (B) Colony assays of yeast cell growth in liquid culture under YPD (normal nitrogen) and SLAD (low nitrogen) across 10-fold dilutions. (C) Spectrophotometric analysis of yeast cell growth in liquid culture under YPD and SLAD. As seen in B and C, deletion of the genes DPL1 and LAG1 greatly reduces yeast cell growth under nitrogen stress.

stress, both by colony assay and by spectrophotometric analysis of yeast cell growth in liquid culture. This result indicates that wild-type activity of subpathways regulating sphingolipid synthesis and catabolism is required for efficient cell growth during conditions of nitrogen stress while validating TEAK's utility in extracting biologically relevant subpathways from large-scale data sets.

Transcriptional programs regulated by the filamentous growth transcription factors Flo8p and Mss11p

In filamentous strains of *S. cerevisiae*, nitrogen stress also elicits a complex morphogenetic program resulting in the

transition to a filamentous form of growth. The transcription factors Flo8p and Mss11p are both required for this filamentous growth transition in that homozygous diploid *flo8Δ/Δ* and *mss11Δ/Δ* strains do not undergo filamentation under conditions of nitrogen stress (72,73). Flo8p is phosphorylated and activated by Tpk2p, a catalytic subunit of PKA (23), and is known to directly regulate transcription from the *MUC1/FLO11* promoter (Figure 5). Flo8p-binding sites have been assessed by chromatin immunoprecipitation/microarray studies (74), indicating hundreds of target sites across the yeast genome. Collectively, the DNA recognition pattern

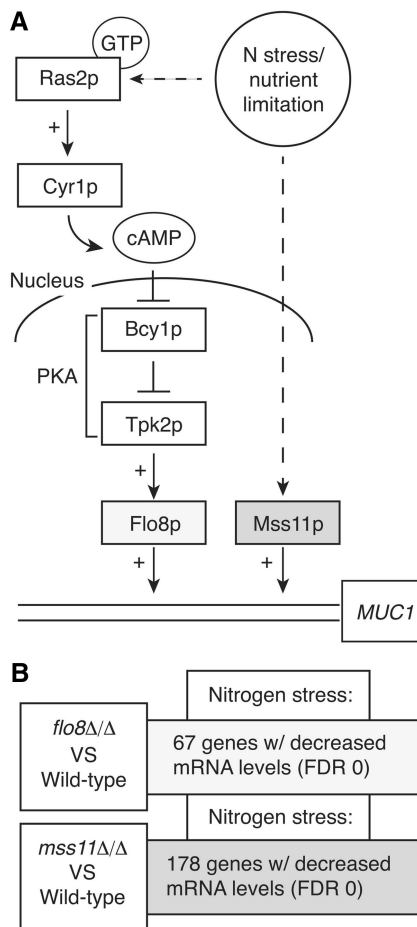


Figure 5. (A) Flop and Mss11p regulate the *MUC1/FLO11* promoter required for wild-type filamentous growth. (B) DNA microarray analysis to profile mRNA level changes in homozygous diploid strains of the filamentous $\Sigma 1278b$ genetic background deleted for *FLO8* and *MSS11*, respectively.

suggests regulation of a broad set of cellular processes although the downstream effects on mRNA levels of *FLO8* deletion have not been investigated extensively. By genetic analyses, Mss11p is thought to play a central role in the yeast filamentous growth response, putatively acting in concert with both the filamentous growth MAPK and PKA pathways (72) and also acting directly on the *MUC1* promoter. Mss11p target sites have not been assessed by chromatin immunoprecipitation/microarray analysis although mRNA levels have been profiled in a haploid *mss11Δ* strain under normal growth conditions (75). In addition to their role in regulating the filamentous growth response, Flo8p and Mss11p bind cooperatively to the *STAI* gene promoter under conditions of glucose depletion, regulating the transcription of this extracellular glucoamylase important for starch degradation (76). Thus, we expect that Flo8p and Mss11p mediate extensive and diverse transcriptional programs required for wild-type filamentous growth.

Here, we used DNA microarray analysis to profile changes in mRNA levels in homozygous diploid strains of the filamentous $\Sigma 1278b$ genetic background deleted for *FLO8* and *MSS11*, respectively, under conditions of

nitrogen stress. The results of this analysis are summarized in Figure 5. As indicated, deletion of *FLO8* resulted in decreased mRNA levels of 67 genes during nitrogen stress, and 178 genes exhibited decreased transcript abundance in the *mss11Δ/Δ* strain. A smaller set of genes showing increased transcript levels were also observed in the homozygous diploid deletion strains, and these genes can be accessed from the microarray data files provided (NCBI GEO Accession Number GSE40530). By a simple over-representation approach, we analyzed the resulting gene sets for enrichment of associated gene ontology (GO) biological process terms (Figure 5; Supplementary Figure S4). The smaller gene set exhibiting increased transcript levels showed little enrichment for informative GO terms, but the genes showing decreased mRNA levels were significantly enriched for several cellular processes involved in filamentation. From this analysis, *FLO8* deletion is likely to affect the biosynthesis/metabolism of sulfur-containing and nitrogenous compounds while also impairing wild-type transport of nitrogenous compounds. *MSS11* deletion resulted in decreased transcript levels for an overrepresented set of genes mediating cell-cell adhesion, β -glucan biosynthesis, cytoskeletal organization, and, more generally, signal transduction. These processes are known to be involved in the filamentous growth response, and the results validate the essential roles of both transcription factors during pseudohyphal development.

Interestingly, a relatively small set of genes was identified with differential mRNA levels in both the *flo8Δ/Δ* and *mss11Δ/Δ* strains (Figure 5). Within this small gene set, no GO term enrichment was observed, and the results may speak to the point that under conditions of nitrogen stress, Flo8p and Mss11p may not co-regulate an extensive set of gene targets.

Glycerophospholipid metabolism and the *SLC1* gene contribute to yeast pseudohyphal growth

To supplement the overrepresentation approach utilized earlier, we further analyzed the mRNA profiles of the *flo8Δ/Δ* and *mss11Δ/Δ* strains by TEAK. Here, we specifically focused on nonlinear subpathways as we expected that nonlinear relationships would be most prevalent in the highly interconnected cell processes underlying filamentous growth. The resulting TEAK analysis identified a gene set mediating glycerophospholipid metabolism as the top-ranking subpathway within the data sets (Figure 6). The corresponding GO biological process terms were not identified as being enriched in the data sets by the overrepresentation analysis. To identify a putative functional role for this subpathway within the filamentous growth response, we focused on the gene *SLC1*.

As indicated in Figure 6, *SLC1* encodes a 1-acyl-*sn*-glycerol-3-phosphate acetyltransferase catalyzing the acylation of *lyso*-phosphatidic acid to form the key glycerolipid biosynthesis intermediate phosphatidic acid (77). Deletion of *SLC1* does not abolish this enzymatic activity, however, as Ale1p also provides lysophospholipid acetyltransferase activity (78). Interestingly, Slc1p

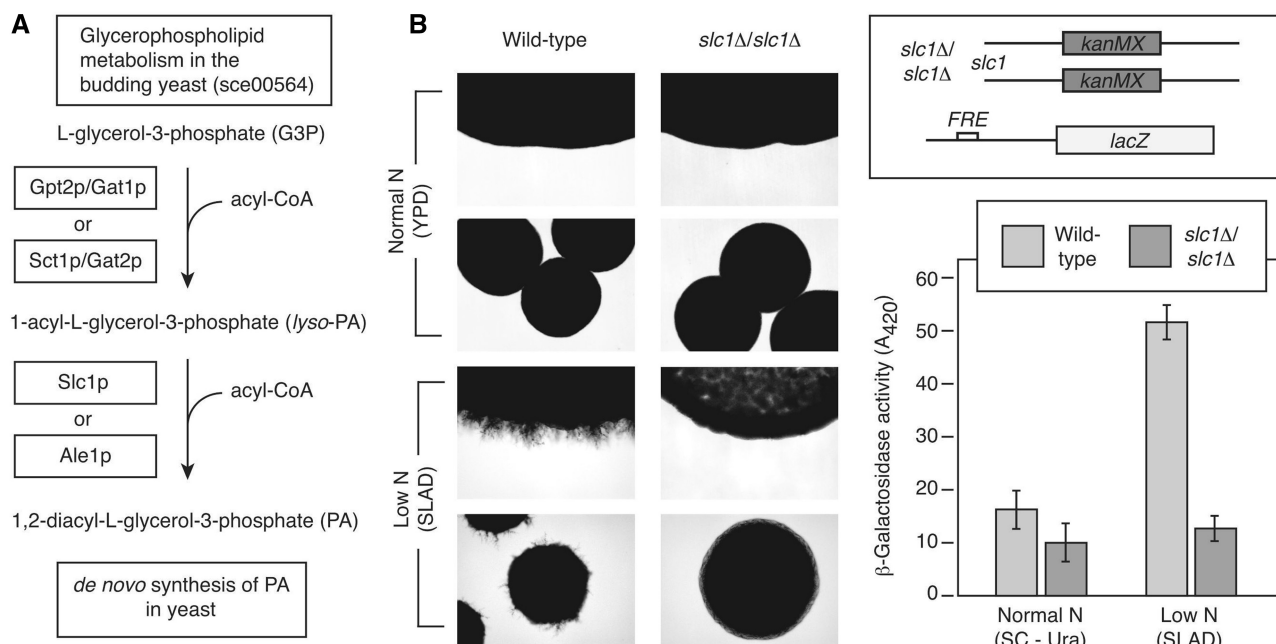


Figure 6. (A) A simplified representation of the glycerophospholipid metabolism pathway. (B) Deletion of the gene *SLC1* leads to loss of filamentous growth under SLAD growth conditions.

additionally exhibits magnesium-dependent acetyltransferase activity toward lyso forms of phosphatidylserine and phosphatidylinositol (77). While these functions of Slc1p are well established, its contributions toward filamentous growth had not been investigated previously, and its role, if any, in this stress response was unknown.

To determine if *SLC1* is involved in the yeast pseudohyphal response to nitrogen stress, we generated a homozygous diploid deletion of the *SLC1* gene and analyzed the resulting mutant for the ability to form surface-spread filaments under conditions of nitrogen stress. Strikingly, on low-nitrogen medium, the *slc1Δ/Δ* mutant exhibited a loss of pseudohyphal filamentation (Figure 6). To assess whether this loss of filamentous growth occurs at least in part through decreased activity of the Kss1p MAPK pathway, we introduced a plasmid-based filamentation and invasion response element (FRE)-driven *lacZ* reporter in the *slc1Δ/Δ* mutant (79). The FRE promoter sequence is recognized by the Ste12p-Tec1p transcription factor complex that acts downstream of the Kss1p MAPK pathway; thus, expression of this reporter is a good indicator of filamentous growth MAPK pathway activity. Under conditions of nitrogen stress in the *slc1Δ/Δ* mutant, we observed a significant decrease in *lacZ*-encoded β-galactosidase activity relative to wild type, indicating decreased activity of the filamentous growth MAPK pathway upon deletion of *SLC1* during nitrogen deprivation. Thus, *SLC1* and glycerophospholipid metabolism do contribute to the yeast pseudohyphal response although the mechanism of this Flo8p- and Mss11p-mediated effect needs to be elucidated further. In addition, TEAK was indeed effective in extracting biologically relevant subpathways from the transcriptional profiling data sets.

Method comparison

We also ran SPIA (12) and SM (14) on the differentially expressed gene sets in Figure 5B. Table 1 presents TEAK's top 20 nonlinear subpathway results. In the top ranked nonlinear glycerophospholipid metabolism subpathway identified by TEAK alone, we experimentally validated that the *SLC1* gene is necessary for filamentous growth under SLAD growth conditions as seen in Figure 6B. In Supplementary Tables S2–S12, TEAK's top 20 linear subpathway results and SPIA and SM results are provided.

DISCUSSION AND CONCLUSION

TEAK, freely available at code.google.com/p/teak for Windows and MAC, is an innovative approach to detect activated subpathways. First, TEAK uses an in-house graph traversal algorithm to extract all root to leaf linear subpathways of a given pathway. Its major contributions include fully accounting for the topological information of subpathways and its ability to provide an interactive view of the data in the KEGG pathways. Furthermore, TEAK's GUI allows easy accessibility for a diverse set of users, and it implements an efficient algorithm to extract root-to-leaf linear subpathways where BFS and DFS algorithms may fail. Compared with previous approaches, TEAK also does not use differential gene expression analysis to determine modules of interest and is thus not sensitive to threshold values. Finally, by integrating the computational TEAK with experimental approaches, we have discovered and experimentally validated previously uncharacterized subpathways

Table 1. TEAK's top 20 nonlinear subpathway results

Name	SPIA	SM
Glycerophospholipid metabolism	No	No
Starch and sucrose metabolism	No	Yes
Glutathione metabolism	No	No
Terpenoid backbone biosynthesis	No	No
Cysteine and methionine metabolism	No	No
Purine metabolism	No	No
Cysteine and methionine metabolism	No	No
Arginine and proline metabolism	No	No
Cysteine and methionine metabolism	No	No
Sphingolipid metabolism	No	No
Terpenoid backbone biosynthesis	No	No
Alanine, aspartate and glutamate metabolism	No	No
Pentose phosphate pathway	No	No
Starch and sucrose metabolism	No	Yes
Purine metabolism	No	No
Purine metabolism	No	No
Pentose phosphate pathway	No	No
Sphingolipid metabolism	No	No
Fructose and mannose metabolism	No	No
MAPK signaling pathway—yeast	Yes	No

In the SPIA (12) and SM (14) columns, 'No' means that the result was absent from a method's top 20 results, 'Yes' otherwise. In Figure 6B, we experimentally validated the *SLC1* gene in the top ranked glycerophospholipid metabolism subpathway is necessary for filamentous growth under SLAD growth conditions.

involved in the yeast stress response to nitrogen starvation.

While our approach is effective for the extraction of biologically informative subpathways from large-scale data sets, the work presented here does possess three limitations. First, the microarray analysis of the *flo8Δ/Δ* and *ms11Δ/Δ* strains will identify both direct and indirect transcription factor targets. As the indirect targets are useful in defining the broader scope of processes affected by each transcription factor, this is not necessarily a major limitation as long as the results are interpreted with this in mind. Second, the current application of TEAK draws upon KEGG annotations as the pathway resource. Annotations of pathways, particularly regulatory pathways, cannot be considered to be all-inclusive in KEGG; however, the KEGG resource is the largest resource of yeast pathway data at present and the best option for this analysis to date. Third, the study presented here does not extend beyond transcriptional analyses of the events underlying yeast cellular responses to nitrogen stress. While it is obviously true that post-transcriptional regulatory mechanisms contribute significantly to all eukaryotic stress responses, transcriptional regulation is also a major control mechanism in modulating gene function during the yeast response to nitrogen stress. Furthermore, it should be noted that the TEAK approach can be applied easily to data sets other than transcriptional profiles. Thus, the applications of TEAK are not limited in this way.

The identification of sphingolipid metabolic pathways as a component of the yeast nitrogen stress response is interesting given the important role of sphingolipids as second messengers. Recently, decreased sphingolipid synthesis has been reported to increase yeast cell lifespan

through processes involving reduced Sch9p kinase activity and reduced chromosomal mutations (64). More to the point, previous studies have established that inhibition of the *de novo* sphingolipid synthesis pathway activates the MAPK pathway mediating the yeast response to high osmolarity (the HOG pathway) (63). HOG pathway activity inhibits the yeast filamentous growth response to nitrogen stress (80,81), which suggests that sphingolipid accumulation may be an activating signal toward pseudohyphal growth. This hypothesis can be tested in a filamentous strain of *S. cerevisiae*.

Our article focuses significantly on the pseudohyphal response to nitrogen stress in a filamentation-competent strain of *S. cerevisiae*. We specifically chose to study the filamentous growth transcription factors Flo8p and Mss11p because (i) they are required for wild-type filamentous growth; (ii) their respective transcriptional programs remain to be fully delineated; and (iii) they may function cooperatively during pseudohyphal development. By the microarray-based approach employed here, we detect changes in mRNA level that occur as both direct and indirect effects of *FLO8* and *MSS11* deletion. With respect to the transcriptional program controlled by Flo8p, the distinction between direct and indirect effects can be studied by comparing the results presented here with published chromatin immunoprecipitation/microarray analysis of Flo8p binding (74). As the latter method should be highly enriched for direct transcription factor-promoter interactions, additional transcript level changes reported here are likely to be enriched in indirect effects. Furthermore, the overlap between the data sets provides a high confidence set of Flo8p targets. In total, 15 genes were identified in common through both methods of study including the following genes required for wild-type filamentous growth: *MUC1*, *HMS1*, *NDE1*, *PDR11* and *CLN1*.

Interestingly, Flo8p and Mss11p are known to bind cooperatively to a short inverted repeat sequence in the *STAI* gene promoter under conditions of glucose limitation (76). Presumably, the transcription factors also cooperatively bind additional promoters and promoter elements. This suggests that we may observe a significant degree of overlap between the genes exhibiting differential transcript levels upon *FLO8* and *MSS11* deletion. In our analysis, however, we only observed 16 such genes in common between the two data sets. This overlapping set does include *MUC1/FLO11* whose promoter is known to be recognized by both factors during nitrogen stress-induced filamentous growth. In this case, though, Flo8p and Mss11p bind to Ste12p and Tec1p, which collectively recognize the FRE sequence in the *MUC1* promoter (76). Further chromatin immunoprecipitation/sequencing studies will be helpful in determining the binding dynamics of Flo8p and Mss11p.

The transcriptional profiles of *flo8Δ* and *ms11Δ* strains indicate a broad range of cellular processes encompassed within these transcriptional programs. Although additional transcription factors contribute to the full transcriptional regulation underlying filamentous differentiation, the gene set reported here is reflective of the broad scope of this growth transition. The development

of pseudohyphal filaments requires the output of complex processes regulating cell polarity, cell cycle progression, cell morphology, cell–cell adhesion, cytoskeletal organization and numerous metabolic systems. Consequently, genomics has been and continues to be very useful in identifying these higher-order processes for subsequent detailed follow-up analysis.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–12 and Supplementary Figures 1–4.

ACKNOWLEDGEMENTS

We thank the reviewers for their thoughtful and useful feedback. We also thank Lipi Acharya of Dow AgroSciences.

FUNDING

National Institutes of Health (NIH) [R21LM010137 to D.Z.]; NIH [1R01-A1098450-01A1 to A.K.]; March of Dimes [1-FY11-403 to A.K.]; Kumar Lab. Funding for open access charge: Wayne State University and University of Michigan research grants.

Conflict of interest statement. None declared.

REFERENCES

- Kanehisa, M. and Goto, S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.*, **28**, 27–30.
- Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. and Tanabe, M. (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.*, **40**, D109–D114.
- Matthews, L., Gopinath, G., Gillespie, M., Caudy, M., Croft, D., de Bono, B., Garapati, P., Hemish, J., Hermjakob, H., Jassal, B. *et al.* (2009) Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.*, **37**, D619–D622.
- Croft, D., O’Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., Caudy, M., Garapati, P., Gopinath, G., Jassal, B. *et al.* (2011) Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.*, **39**, D691–D697.
- Khatri, P. and Draghici, S. (2005) Ontological analysis of gene expression data: current tools, limitations, and open problems. *Bioinformatics*, **21**, 3587–3595.
- Grosu, P., Townsend, J.P., Hartl, D.L. and Cavalieri, D. (2002) Pathway processor: a tool for integrating whole-genome expression results into metabolic networks. *Genome Res.*, **12**, 1121–1126.
- Pan, D., Sun, N., Cheung, K.H., Guan, Z., Ma, L., Holford, M., Deng, X. and Zhao, H. (2003) PathMAP: a tool for displaying gene expression and performing statistical tests on metabolic pathways at multiple levels for *Arabidopsis*. *BMC Bioinformatics*, **4**, 56.
- Pandey, R., Guru, R.K. and Mount, D.W. (2004) Pathway Miner: extracting gene association networks from molecular pathways for predicting the biological significance of gene expression microarray data. *Bioinformatics*, **20**, 2156–2158.
- Chung, H.-J., Kim, M., Park, C.H., Kim, J. and Kim, J.H. (2004) ArrayXPath: mapping and visualizing microarray gene-expression data with integrated biological pathway resources using Scalable Vector Graphics. *Nucleic Acids Res.*, **32**(Suppl. 2), W460–W464.
- Dahlquist, K.D., Salomonis, N., Vranizan, K., Lawlor, S.C. and Conklin, B.R. (2002) GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. *Nat. Genet.*, **31**, 19–20.
- Zhu, D. (2009) Semi-supervised gene shaving method for predicting low variation biological pathways from genome-wide data. *BMC Bioinformatics*, **10**(Suppl. 1), S54.
- Tarca, A.L., Draghici, S., Khatri, P., Hassan, S.S., Mittal, P., Kim, J.-S., Kim, C.J., Kusanovic, J.P. and Romero, R. (2009) A novel signaling pathway impact analysis. *Bioinformatics*, **25**, 75–82.
- 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.
- Li, C., Li, X., Miao, Y., Wang, Q., Jiang, W., Xu, C., Li, J., Han, J., Zhang, F., Gong, B. *et al.* (2009) SubpathwayMiner: a software package for flexible identification of pathways. *Nucleic Acids Res.*, **37**, e131.
- Chen, X., Xu, J., Huang, B., Li, J., Wu, X., Ma, L., Jia, X., Bian, X., Tan, F., Liu, L. *et al.* (2011) A sub-pathway-based approach for identifying drug response principal network. *Bioinformatics*, **27**, 649–654.
- Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., Botstein, D., Brown, P.O. and Silver, P.A. (2000) Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell*, **11**, 4241–4257.
- Causton, H., Ren, B., Koh, S., Harbison, C., Kanin, E., Jennings, E., Lee, T., True, H., Lander, E. and Young, R. (2001) Remodeling of yeast genome expression in response to environmental changes. *Mol. Biol. Cell*, **12**, 323–337.
- Mousley, C., Yuan, P., Gaur, N., Trettin, K., Nile, A., Deminoff, S., Dewar, B., Wolpert, M., Macdonald, J., Herman, P. *et al.* (2012) A sterol-binding protein integrates endosomal lipid metabolism with TOR signaling and nitrogen sensing. *Cell*, **148**, 702–715.
- Gimeno, C.J., Ljungdahl, P.O., Styles, C.A. and Fink, G.R. (1992) Unipolar cell divisions in the yeast *S. cerevisiae* lead to filamentous growth: regulation by starvation and RAS. *Cell*, **68**, 1077–1090.
- Roberts, R. and Fink, G. (1994) Elements of a single MAP kinase cascade in *Saccharomyces cerevisiae* mediate two developmental programs in the same cell type: mating and invasive growth. *Genes Dev.*, **8**, 2974–2985.
- Braun, B. and Johnson, A. (1997) Control of filament formation in *Candida albicans* by the transcriptional repressor TUP1. *Science*, **277**, 105–109.
- Jayatilake, J., Samaranyake, Y., Cheung, L. and Samaranyake, L. (2006) Quantitative evaluation of tissue invasion by wild type, hyphal and SAP mutants of *Candida albicans*, and non-albicans *Candida* species in reconstituted human oral epithelium. *J. Oral Pathol. Med.*, **35**, 484–491.
- Pan, X. and Heitman, J. (1999) Cyclic AMP-dependent protein kinase regulates pseudohyphal differentiation in *Saccharomyces cerevisiae*. *Mol. Cell Biol.*, **19**, 4874–4887.
- Pan, X. and Heitman, J. (2002) Protein kinase A operates a molecular switch that governs yeast pseudohyphal differentiation. *Mol. Cell Biol.*, **22**, 3981–3993.
- Guo, B., Styles, C., Feng, Q. and Fink, G. (2000) A *Saccharomyces* gene family involved in invasive growth, cell-cell adhesion, and mating. *Proc. Natl Acad. Sci. USA*, **97**, 12158–12163.
- Karunanithi, S., Vadaie, N., Chavel, C., Birkaya, B., Joshi, J., Grell, L. and Cullen, P. (2010) Shedding of the mucin-like flocculin Flo11p reveals a new aspect of fungal adhesion regulation. *Curr. Biol.*, **20**, 1389–1395.
- Ma, J., Jin, R., Jia, X., Dobry, C., Wang, L., Reggiori, F., Zhu, J. and Kumar, A. (2007) An interrelationship between autophagy and filamentous growth in budding yeast. *Genetics*, **177**, 205–214.
- Jin, R., Dobry, C., McCown, P. and Kumar, A. (2008) Large-scale analysis of yeast filamentous growth by systematic gene disruption and overexpression. *Mol. Biol. Cell*, **19**, 284–296.
- Palla, G., Derenyi, I., Farkas, I. and Vicsek, T. (2005) Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, **435**, 814–818.
- Palla, G., Farkas, I.J., Pollner, P., Derenyi, I. and Vicsek, T. (2007) Directed network modules. *N. J. Phys.*, **9**, 186.
- Schwarz, G. (1978) Estimating the dimension of a model. *Ann. Stat.*, **6**, 461–464.

32. Murphy, K.P. (2001) The Bayes Net Toolbox for MATLAB. *Comput. Sci. Stat.*, **33**, 2001.
33. Baeuerlein, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., Andre, B. *et al.* (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature*, **418**, 387–391.
34. Bharucha, N., Ma, J., Dobry, C., Lawson, S., Yang, Z. and Kumar, A. (2008) Analysis of the yeast kinome reveals a network of regulated protein localization during filamentous growth. *Mol. Biol. Cell*, **19**, 2708–2717.
35. Guthrie, C. and Fink, G. (1991) *Guide to Yeast Genetics and Molecular Biology*. Academic Press, San Diego, CA.
36. Lorenz, M. and Heitman, J. (1997) Yeast pseudohyphal growth is regulated by GPA2, a G protein alpha homolog. *EMBO J.*, **16**, 7008–7018.
37. Baudin, A., Ozier-Kalogeropoulos, O., Denouel, A., Lacroute, F. and Cullin, C. (1993) A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **21**, 3329–3330.
38. Wach, A., Brachat, A., Pohlmann, R. and Philippsen, P. (1994) New heterologous modules for classical or PCR-based gene disruptions in *Saccharomyces cerevisiae*. *Yeast*, **10**, 1793–1808.
39. Longtine, M., McKenzie, A. III, Demarini, D., Shah, N., Wach, A., Brachat, A., Philippsen, P. and Pringle, J. (1998) Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast*, **14**, 953–961.
40. Ito, H., Fukuda, Y., Murata, K. and Kimura, A. (1983) Transformation of intact yeast cells treated with alkali cations. *J. Bacteriol.*, **153**, 163–168.
41. Coelho, P., Kumar, A. and Snyder, M. (2000) Genome-wide mutant collections: toolboxes for functional genomics. *Curr. Opin. Microbiol.*, **3**, 309–315.
42. Kumar, A., DesEtages, S., Coelho, P., Roeder, G. and Snyder, M. (2000) High-throughput methods for the large-scale analysis of gene function by transposon tagging. *Methods Enzymol.*, **328**, 550–574.
43. Kumar, A., Vidan, S. and Snyder, M. (2002) Insertional mutagenesis: transposon-insertion libraries as mutagens in yeast. *Methods Enzymol.*, **350**, 219–229.
44. Ma, J., Dobry, C., Krysan, D. and Kumar, A. (2008) Unconventional genomic architecture in the budding yeast *Saccharomyces cerevisiae* masks the nested antisense gene NAG1. *Eukaryot. Cell*, **7**, 1289–1298.
45. Wu, Z., Irizarry, R.A., Gentleman, R., Martinez-Murillo, F. and Spencer, F. (2004) A model-based background adjustment for oligonucleotide expression arrays. *J. Am. Stat. Assoc.*, **99**, 909–917.
46. Tusher, V., Tibshirani, R. and Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA*, **98**, 5116–5121.
47. Acharya, L., Judeh, T., Duan, Z., Rabbat, M. and Zhu, D. (2012) GSGS: a computational approach to reconstruct signaling pathway structures from gene sets. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **9**, 438–450.
48. Acharya, L.R., Judeh, T., Wang, G. and Zhu, D. (2012) Optimal structural inference of signaling pathways from unordered and overlapping gene sets. *Bioinformatics*, **28**, 546–556.
49. Cormen, T.H., Leiserson, C.E., Rivest, R.L. and Stein, C. (2009) *Introduction to Algorithms*, 3rd edn. The MIT Press, Cambridge, MA.
50. Li, S. (2005) Mechanisms of cellular signal transduction. *Int. J. Biol. Sci.*, **1**, 152.
51. Bianco, R., Melisi, D., Ciardiello, F. and Tortora, G. (2006) Key cancer cell signal transduction pathways as therapeutic targets. *Eur. J. Cancer*, **42**, 290–294.
52. Selcher, J.C., Weeber, E.J., Varga, A.W., Sweatt, J.D. and Swank, M. (2002) Book review: protein kinase signal transduction cascades in mammalian associative conditioning. *Neuroscientist*, **8**, 122–131.
53. Shinozaki, K. and Yamaguchishinozaki, K. (1997) Gene expression and signal transduction in water-stress response. *J. Plant Physiol.*, **115**, 327–334.
54. Alon, U. (2007) Network motifs: theory and experimental approaches. *Nat. Rev. Genet.*, **8**, 450–461.
55. Koller, D. and Friedman, N. (2009) *Probabilistic Graphical Models: Principles and Techniques*. MIT Press.
56. Friedman, N., Nachman, I. and Pe'er, D. (1999) Learning Bayesian Network Structure from Massive Datasets. *The "Sparse Candidate" Algorithm*. In: *Proceeding Fifteenth Conference on Uncertainty in Artificial Intelligence*. UAI, pp. 206–215.
57. Friedman, N., Linial, M. and Nachman, I. (2000) Using Bayesian networks to analyze expression data. *J. Comput. Biol.*, **7**, 601–620.
58. Shachter, R. and Kenley, C. (1989) Gaussian Influence Diagrams. *Manag. Sci.*, **35**, 527–550.
59. Gomez-Villegas, M.A., Main, P. and Susi, R. (2007) Sensitivity analysis in Gaussian Bayesian networks using a divergence measure. *Commun. Stat. Theory Methods*, **36**, 523–539.
60. Gomez-Villegas, M.A., Main, P., Navarro, H. and Susi, R. (2010) Dealing with uncertainty in Gaussian Bayesian networks from a regression perspective. In: Petri Myllymaki, T.R. and Jaakkola, T. (eds). *Proceedings of the Fifth European Workshop on Probabilistic Graphical Models*, pp. 145–152.
61. Roberts, S.J. and Penny, W.D. (2002) Variational Bayes for generalized autoregressive models. *IEEE Trans. Signal Processing*, **50**, 2245–2257.
62. Dickson, R. (2010) Roles for sphingolipids in *Saccharomyces cerevisiae*. *Adv. Exp. Med. Biol.*, **688**, 217–231.
63. Tanigawa, M., Kihara, A., Terashima, M., Takahara, T. and Maeda, T. (2012) Sphingolipids regulate the yeast high osmolarity-responsive HOG pathway. *Mol. Cell. Biol.*, **32**, 2861–2870.
64. Huang, X., Liu, J. and Dickson, R. (2012) Down-regulating sphingolipid synthesis increases yeast lifespan. *PLoS Genet.*, **8**, e1002493.
65. Nagiec, M., Skrzypek, M., Nagiec, E., Lester, R. and Dickson, R. (1998) The LCB4 (YOR171c) and LCB5 (YLR260w) genes of *Saccharomyces* encode sphingoid long chain base kinases. *J. Biol. Chem.*, **273**, 19437–19442.
66. Qie, L., Nagiec, M., Baltisberger, J., Lester, R. and Dickson, R. (1997) Identification of a *Saccharomyces* gene, LCB3, necessary for incorporation of exogenous long chain bases into sphingolipids. *J. Biol. Chem.*, **272**, 16110–16117.
67. Mao, C., Xu, R., Bielawska, A., Szulc, Z. and Obeid, L. (2000) Cloning and characterization of a *Saccharomyces cerevisiae* alkaline ceramidase with specificity for dihydroceramide. *J. Biol. Chem.*, **275**, 31369–31378.
68. Saba, J., Nara, F., Bielawska, A., Garrett, S. and Hannun, Y. (1997) The BST1 gene of *Saccharomyces cerevisiae* is the sphingosine-1-phosphate lyase. *J. Biol. Chem.*, **272**, 26087–26090.
69. D'mello, N.P., Childress, A.M., Franklin, D.S., Kale, S.P., Pinswasdi, C. and Jazwinski, S.M. (1994) Cloning and characterization of LAG1, a longevity-assurance gene in yeast. *J. Biol. Chem.*, **269**, 15451–15459.
70. Kim, S., Fyrst, H. and Saba, J. (2000) Accumulation of phosphorylated sphingoid long chain bases results in cell growth inhibition in *Saccharomyces cerevisiae*. *Genetics*, **156**, 1519–1529.
71. Kolaczowski, M., Kolaczowska, A., Gaigg, B., Schneider, R. and Moye-Rowley, W. (2004) Differential regulation of ceramide synthase components LAC1 and LAG1 in *Saccharomyces cerevisiae*. *Eukaryot. Cell*, **3**, 880–892.
72. van Dyk, D., Pretorius, I.S. and Bauer, F.F. (2005) Mss11p is a central element of the regulatory network that controls FLO11 expression and invasive growth in *Saccharomyces cerevisiae*. *Genetics*, **169**, 91–106.
73. Xu, T., Shively, C., Jin, R., Eckwahl, M., Dobry, C., Song, Q. and Kumar, A. (2010) A profile of differentially abundant proteins at the yeast cell periphery during pseudohyphal growth. *J. Biol. Chem.*, **285**, 15476–15488.
74. Borneman, A., Leigh-Bell, J., Yu, H., Bertone, P., Gerstein, M. and Snyder, M. (2006) Target hub proteins serve as master regulators of development in yeast. *Genes Dev.*, **20**, 435–448.
75. Bester, M.C., Jacobson, D. and Bauer, F.F. (2012) Many *Saccharomyces cerevisiae* cell wall protein encoding genes are coregulated by Mss11, but cellular adhesion phenotypes appear only Flo protein dependent. *G3*, **2**, 131–141.
76. Kim, T., Kim, H., Yoon, J. and Kang, H. (2004) Recruitment of the Swi/Snf complex by Ste12-Tec1 promotes Flo8-Mss11-mediated activation of STA1 expression. *Mol. Cell. Biol.*, **24**, 9542–9556.

77. Benghezal,M., Roubaty,C., Veepuri,V., Knudsen,J. and Conzelmann,A. (2007) SLC1 and SLC4 encode partially redundant acyl-coenzyme A 1-acylglycerol-3-phosphate O-acyltransferases of budding yeast. *J. Biol. Chem.*, **282**, 30845–30855.
78. Jain,S., Stanford,N., Bhagwat,N., Seiler,B., Costanzo,M., Boone,C. and Oelkers,P. (2007) Identification of a novel lysophospholipid acyltransferase in *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **282**, 30562–30569.
79. Madhani,H. and Fink,G. (1997) Combinatorial control required for the specificity of yeast MAPK signaling. *Science*, **275**, 1314–1317.
80. O'Rourke,S. and Herskowitz,I. (1998) The Hog1 MAPK prevents cross talk between the HOG and pheromone response MAPK pathways in *Saccharomyces cerevisiae*. *Genes Dev.*, **12**, 2874–2886.
81. Pitoniak,A., Birkaya,B., Dionne,H., Vadaie,N. and Cullen,P. (2009) The signaling mucins Msb2 and Hkr1 differentially regulate the filamentation mitogen-activated protein kinase pathway and contribute to a multimodal response. *Mol. Biol. Cell*, **20**, 3101–3114.