ChiNet uncovers rewired transcription subnetworks in tolerant yeast for advanced biofuels conversion

Yang Zhang[1], Z. Lewis Liu[2],[*] and Mingzhou Song[1],[*]

[1]Department of Computer Science, New Mexico State University, Las Cruces, NM 88003, USA and [2]National Center for Agricultural Utilization Research, Agricultural Research Service, U.S. Department of Agriculture, Peoria, IL 61604, USA

ABSTRACT

Analysis of rewired upstream subnetworks impacting downstream differential gene expression aids the delineation of evolving molecular mechanisms. Cumulative statistics based on conventional differential correlation are limited for subnetwork rewiring analysis since rewiring is not necessarily equivalent to change in correlation coefficients. Here we present a computational method ChiNet to quantify subnetwork rewiring by statistical heterogeneity that enables detection of potential genotype changes causing altered transcription regulation in evolving organisms. Given a differentially expressed downstream gene set, ChiNet backtracks a rewired upstream subnetwork from a super-network including gene interactions known to occur under various molecular contexts. We benchmarked ChiNet for its high accuracy in distinguishing rewired artificial subnetworks, *in silico* yeast transcription-metabolic subnetworks, and rewired transcription subnetworks for *Candida albicans* versus *Saccharomyces cerevisiae*, against two differential-correlation based subnetwork rewiring approaches. Then, using transcriptome data from tolerant *S. cerevisiae* strain NRRL Y-50049 and a wild-type intolerant strain, ChiNet identified 44 metabolic pathways affected by rewired transcription subnetworks anchored to major adaptively activated transcription factor genes *YAP1*, *RPN4*, *SFP1* and *ROX1*, in response to toxic chemical challenges involved in lignocellulose-to-biofuels conversion. These findings support the use of ChiNet in rewiring analysis of subnetworks where differential interaction patterns resulting from divergent nonlinear dynamics abound.

INTRODUCTION

Network rewiring refers to changes of network over time by either gain or loss of molecular interactions among distinct taxonomic entities (1). Rewiring of subnetworks, specific components in molecular networks, allows an organism to adapt to a defined environmental condition. Subnetwork rewiring alters molecule–molecule interactions as a second-order change that occurs in either strength or topology of molecular interactions. There must exist some input to which a rewired subnetwork responds differentially from the original subnetwork. We define the first-order change of a subnetwork as working zone change characterized by shift in the probability distributions of molecules in the subnetwork. A modified subnetwork response can be a consequence of either first- or second-order subnetwork changes. Modified metabolic network responses involving glycolysis and pentose phosphate pathways have been defined for a tolerant industrial yeast strain *Saccharomyces cerevisiae* NRRL Y-50049 under toxic chemical challenges (2). Mechanisms of *in situ* detoxification by the yeast strain and key regulatory elements involved in its tolerance were also identified (3,4). However, it remains unclear how upstream transcription networks may have been rewired to impact downstream metabolisms that confer toxic tolerance on Y-50049.

Large-scale rewiring of transcription programs in response to the loss of a *cis*-regulatory element was reported to elicit contrasting anaerobic/aerobic growth in yeasts (5). Variations in gene expression responses to stresses among four yeast species have been confirmed to be associated with the disparity of TATA boxes in promoter regions (6). Extracellular signaling was also shown to impose post-translational modifications on a transcription factor (TF) that reverses its function from activation to repression of gene expression (7). In fact, TF binding was found to be ultra-sensitive to disruptive sequence variations among human genomes (8). These findings raise the possibility that transcription network rewiring may have caused the outstanding detoxification capability of microbial strains for

---

[*]To whom correspondence should be addressed. Tel: +1 575 646 4299; Fax: +1 575 646 1002; Email: joemsong@cs.nmsu.edu
Correspondence may also be addressed to Z.L. Liu. Tel: +1 309 681 6294; Fax: +1 309 681 6427; Email: ZLewis.Liu@ars.usda.gov
Present address: Yang Zhang. Amyris Inc., Emeryville, CA, USA.

advanced biofuels production (2,4). In this study, we aim to discover rewired upstream molecular subnetworks that alter expression of target gene sets functioning in downstream biological processes at the genome scale.

By *subnetwork*, we mean a context-dependent component of a larger network, such as an active transcription subnetwork. A *gene set* refers to genes involved in a specifically defined metabolic or signaling pathway. Using cumulative rewiring statistics over a subnetwork, our rationale is to detect subnetworks that are most rewired and also those with individually weak but collectively strong rewiring signals. We aim to uncover both linearly and non-linearly rewired patterns in subnetworks that have been insufficiently addressed by existing approaches focusing mostly on linear interactions.

Pathway analysis aims at revealing activities of functional modules and is more robust to noise than gene-centric methods. Although tens of existing pathway analysis methods (9,10) are all sensitive to pathway response changes, most are unable to distinguish the source of change from pathway stimuli (first-order) or subnetwork rewiring (second-order). For example, evolved from early methods for over-representation of GO terms (11–13), gene set enrichment analyses (14–20) detect subnetwork activity by cumulative statistics of differential gene expression. Multivariate analysis of variance (MANOVA) was proposed to study differential expression of gene sets across conditions by implicitly accounting for gene–gene dependence but not rewiring (21). NetGSA (22) models gene–gene dependency explicitly for differential gene expression analysis of subnetworks, yet it assumed fixed gene-gene interaction coefficients that do not allow for rewiring analysis.

Moving beyond first-order gene-set analysis, current pathway analysis methods detect changes in second-order molecular interactions. Draghici *et al.* (23,24) developed impact analysis methods to compute perturbation factors to evaluate changes in differentially expressed genes in a pathway with respect to the topology of the pathway. SAMNet (25) solves a multi-commodity network flow problem formulated on the topology of protein and mRNA molecules involved in multiple conditions. Network rewiring is indicated by unequal flows passing through gene interactions associated with each condition. A recent method EDDY (26) statistically evaluates the difference in dependencies among genes in a given set between two conditions based on divergence of posterior probabilities modeled by Bayesian networks. Although these subnetwork analysis methods use both first- and second-order information from subnetwork topologies, they are still not designed to annotate the cause of altered subnetwork responses as being subnetwork input or rewiring.

Encouragingly, we start to see methods that are specific to subnetwork rewiring and enable one to separate the driver versus passenger pathways. An early method (27) identifies responsive subnetwork by maximizing a score based on covariance between genes in protein interaction subnetworks, and responsive subnetworks can be independently inferred for each condition and compared for rewiring. The gene set co-expression analysis (GSCA) method (28) sums absolute differential correlation of all gene pairs to obtain a subnetwork dispersion index, but is non-specific to subnetwork

topology and favors those differential interactions in the middle of a signaling cascade. Others use ranks of genes to compute pair-wise correlation (29). The method COSINE computes a score from both gene expression and gene interactions to find condition-specific subnetworks (30), where the gene interaction score is an expected F statistic derived from correlation coefficients. DINA (31) detects gene interactions using Spearman correlation coefficients and further utilizes an entropy measure to determine rewiring based on differences in the number of active interactions in each condition. Most recently, the gene set network correlation analysis (GSNCA) method (32) extended GSCA by assigning heavy weights to emphasize hub genes with strong correlations. However, all these subnetwork rewiring analysis methods follow the principle of subtraction of interaction statistics (33). Regardless of linear (e.g. Pearson correlation) or non-linear (e.g. Spearman correlation) statistics, such a principle is indiscriminate to many complex patterns that indeed differ, as our results will demonstrate.

Production of advanced biofuels including cellulosic ethanol poses major technical challenges for a sustainable bio-based economy (34,35). Development of the next-generation biocatalyst with robust and tolerance characteristics is a necessity for industrial applications. A robust prototype of tolerant industrial yeast *S. cerevisiae* NRRL Y-50049 was developed through evolutionary engineering to *in situ* detoxify furfural and 5-hydroxymethylfurfural (HMF), representative toxic inhibitory compounds liberated from lignocellulose biomass pretreatment (3,36–39). Strain Y-50049 is able to recover from a short lag phase and completes fermentation in the presence of 20 mM each of furfural and HMF (2). Its parental strain, in contrast, was unable to establish a viable culture after a 48-h lag phase and eventually lost function. Although key elements and modified detoxification pathway responses of the tolerant yeast have been described (2,4,40, altered global gene networks and pathways at the genome level that regulate inhibitor tolerance remain largely unknown. How the key regulatory gene *YAP1* is wired to impact specific downstream metabolic pathways for yeast tolerance is not clear. Yeast tolerance is a collective phenotype of gene functions and gene interactions in the cellular regulatory system. Genetic variations, causing upstream transcription subnetwork rewiring, consequently alter downstream metabolic pathway responses. Currently available bioinformatics methods are insufficient to dissect components of such adapted transcription networks. The lack of this knowledge hinders genetic engineering and development of the next-generation biocatalyst for advanced biofuels production.

Previously, we reported a comparative chi-square analysis ($CP\chi^2$) for single-interaction rewiring (41). In this work, we present a computational method ChiNet to detect rewired subnetworks of multiple gene interactions in transcription regulation of downstream metabolic pathways for the tolerant strain NRRL Y-50049. We address three technical challenges: (i) developing subnetwork rewiring statistics; (ii) approximating the null distribution of subnetwork heterogeneity; and (iii) incorporating known TFs and their target genes to enhance the estimation of changes in network connectivity. We fur-

ther demonstrate that ChiNet consistently outperforms differential-correlation based approaches in 60 realistic *in silico* yeast transcription-metabolic subnetworks. We also benchmarked ChiNet showing its substantial advantage over other rewiring analysis methods by simulated subnetworks with 459 configurations of characteristics. We further validated ChiNet for its high accuracy in distinguishing known rewired transcription subnetworks for *Candida albicans* versus *S. cerevisiae*, in comparison to differential-correlation based subnetwork rewiring approaches. Finally, we report transcription subnetwork rewiring anchored to adaptively activated key TF genes that potentially impact global adaptation of the tolerant yeast under toxic chemical stresses. This first insight into tolerant gene regulatory network rewiring aids dissection of genomic mechanisms of yeast tolerance and development of the next-generation biocatalyst for sustainable biofuels and a bio-based economy. ChiNet is most suitable for analysis of high-throughput omics data sets from not-well understood organisms and determine the deviation of their molecular subnetworks from related well-characterized organisms. It opens a new avenue to study the evolution of functional subnetworks in biological systems.

MATERIALS AND METHODS

The ChiNet method for analysis of subnetwork rewiring

We first give an overview of the ChiNet method, which decides whether a subnetwork is conserved, or rewired in either topology or interaction strength across two conditions. In the comparative chi-square framework $CP\chi^2$ first introduced in (41), rewired interactions were characterized by a heterogeneity chi-square. $CP\chi^2$ can detect single strongly rewired interactions, but was not designed to detect subnetwork rewiring. In ChiNet reported here, we introduce cumulative subnetwork statistics for subnetwork rewiring and gamma distributions to assess their null distributions. The strategy is illustrated in Figure 1. The input to ChiNet is an inclusive subnetwork topology and observed data for nodes in the subnetwork under two experimental conditions. Subnetwork topology is a required input from either prior knowledge or extracted from the data by other means such as backtracking we will present. The subnetwork topology can be extracted from a reference biological network module of either the species in question or its related species from KEGG Pathway (42) or other such databases. ChiNet adapts the topology to select only active interactions associated with current experimental conditions. The output of ChiNet consists of subnetwork homogeneity, heterogeneity and total activity across all experimental conditions. In our investigation of yeast tolerance to toxic chemical compounds, heterogeneity χ_D^2 measures subnetwork difference between the two yeast strains in response to the chemical stresses, and homogeneity χ_C^2 for the strength of subnetwork similarity. The total subnetwork activity χ_T^2 represents overall activity for all conditions. The three subnetwork statistics satisfy a decomposition rule cen-

tral to ChiNet:

$$\text{Subnet total activity } \chi_T^2 = \text{Subnet heterogeneity } \chi_D^2$$
$$+\text{Subnet homogeneity } \chi_C^2 \quad (1)$$

which implies that knowing any two statistics can determine the third. We evaluate the significance of the three statistics by gamma distributions to account for statistical dependencies among interactions in subnetworks. Supplementary Figure S1 illustrates that the gamma distribution approximates the null distribution of subnetwork heterogeneity χ_D^2 better than the chi-square approximation, when heterogeneity chi-squares of individual nodes in a subnetwork are not all independent. If a subnetwork only contains differentially expressed genes but is not detected as heterogeneous, its activity is most likely a consequence of differential signaling input rather than genetic variation within the subnetwork. Then ChiNet still considers such a subnetwork conserved despite its differentially expressed genes. When a subnetwork topology is dependent on the molecular context and not specified in advance, we extract one from a super network that contains all possible interactions by the BACKTRACK-REWIRED-SUBNETWORKS algorithm. Next we describe technical details of ChiNet.

By assessing subnetwork homogeneity and heterogeneity, ChiNet determines whether subnetworks of the same set of nodes are conserved, or rewired in either interaction strength or topology, across two or more conditions. We assume subnetwork G is given with its node set V and edge set E. Often a given subnetwork topology is a superset of all known interactions which can be either active or inactive in the current experiment. To use only active interactions, ChiNet has an option to identify a subnetwork from G that best fits to the current experimental data using a chi-square method (43). This step can be replaced by other network inference methods whose output network topology serves as input to ChiNet.

In an interaction, we call the cause variables the parents and the effect variable the child. Let π_1 and π_2 be the parent sets of child node i under two conditions. We form a pooled $r(i) \times s(i)$ contingency table: the $r(i)$ rows are values of parents in the union $\pi_1 \bigcup \pi_2$, and the $s(i)$ columns are the values of child node i. Let $[n_j[l, m]]$ ($j = 1, 2$) be $r(i) \times s(i)$ contingency tables using the same parent union for child i under each of the two conditions. We measure interaction homogeneity of child node i by Pearson's chi-square

$$\chi_c^2(i) = \sum_{l=1}^{r(i)} \sum_{m=1}^{s(i)} \frac{(n_1[l, m] + n_2[l, m] - \bar{n}_c[l, m])^2}{\bar{n}_c[l, m]}$$

$$\text{(Interaction homogeneity)} \quad (2)$$

with $v_c(i) = (r(i) - 1) \times (s(i) - 1)$ degrees of freedom (d.f.). $\bar{n}_c[l, m]$, the expected count in cell $[l, m]$ under the null hypothesis of parents and child being independent, is

$$\bar{n}_c[l, m] = \frac{\sum_{u=1}^{r(i)} (n_1[u, m] + n_2[u, m]) \sum_{t=1}^{s(i)} (n_1[l, t] + n_2[l, t])}{n_1 + n_2} \quad (3)$$

where n_1 and n_2 are the sample size for each condition, respectively. Next we assess interaction strength under each
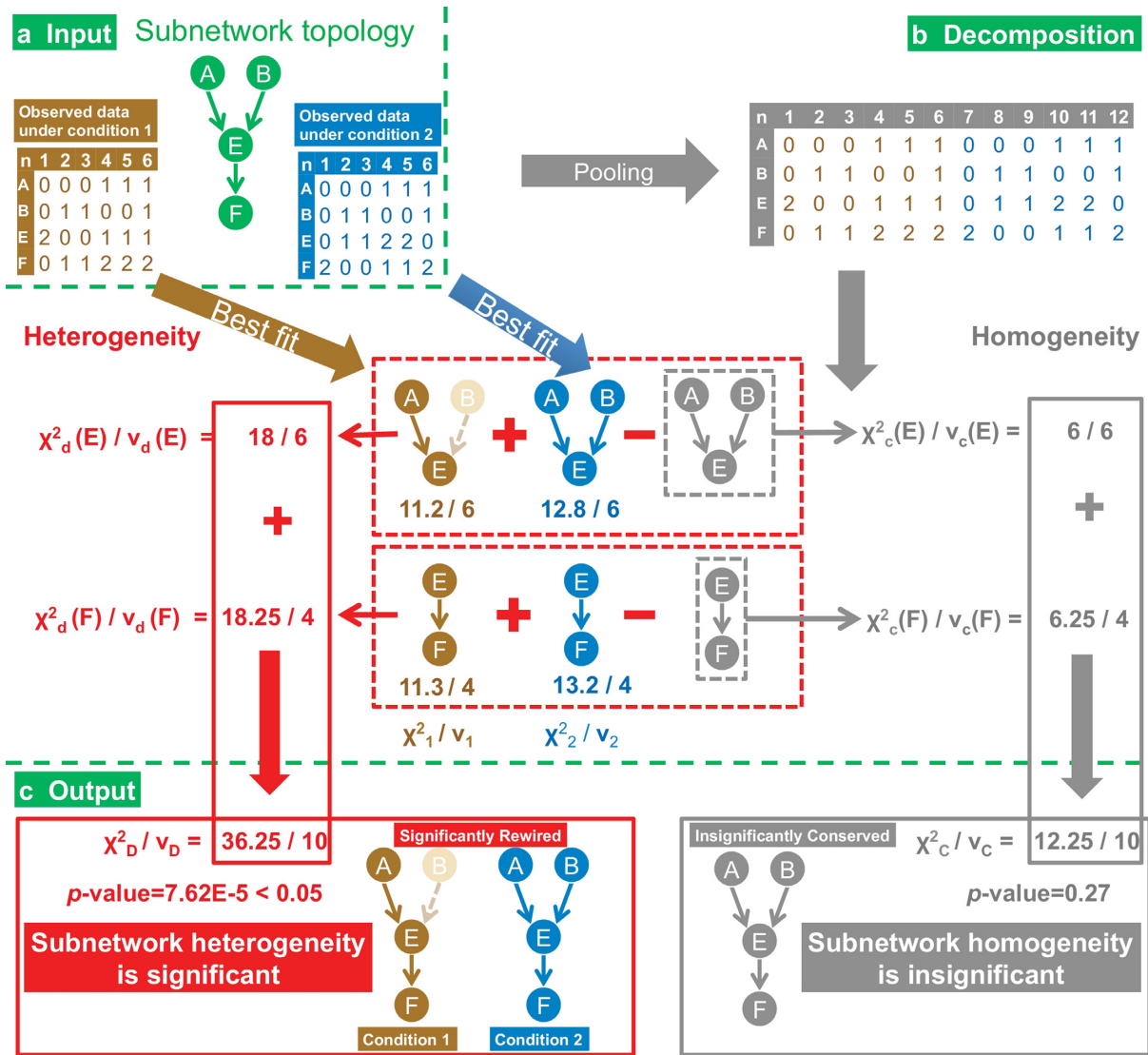
Figure 1. Comparative chi-square analysis of subnetwork rewiring (ChiNet). (**a**) Input. ChiNet requires two types of input: a subnetwork topology and observed data (dynamic or perturbed) for elements in the subnetwork under different conditions. Here the input is a subnetwork topology of four nodes and their observed values under two conditions. (**b**) Decomposition, the core of ChiNet. Using parent nodes in the subnetwork as candidates, the active parents of a child are selected for each condition, doable by various network inference methods. We used best fit to data under each condition as judged by the smallest *P*-value of chi-squares computed on contingency tables. Then a chi-square statistic measuring interaction homogeneity is computed for a contingency table pooling all data. By subtracting the homogeneity chi-square from a total interaction chi-square (sum of chi-squares under each condition)—the decomposition rule, we obtain a third chi-square measuring interaction heterogeneity. Summing up chi-squares of homogeneity and heterogeneity respectively over interactions, we obtain subnetwork heterogeneity and homogeneity. (**c**) Output. Based on the statistical significance of heterogeneity and homogeneity, the subnetwork in this example is determined to be rewired across conditions.

condition. Under the null hypothesis that parents and child are independent, the two tables come from the same null distribution of the pooled table. This gives rise to the expected count for each condition under the null hypothesis

$$\bar{n}_j[l, m] = n_j \frac{\sum\limits_{u=1}^{r(i)} n_1[u, m] + n_2[u, m]}{n_1 + n_2} \frac{\sum\limits_{t=1}^{s(i)} n_1[l, t] + n_2[l, t]}{n_1 + n_2} \quad (4)$$

A chi-square is computed using the observed and expected counts for each condition by

$$\chi_j^2(i) = \sum_{l=1}^{r(i)} \sum_{m=1}^{s(i)} \frac{(n_j[l, m] - \bar{n}_j[l, m])^2}{\bar{n}_j[l, m]}, \quad j = 1, 2 \quad (5)$$

with $v_1(i) = v_2(i) = (r(i) - 1) \times (s(i) - 1)$ d.f., respectively. Summing up interaction chi-squares over both conditions, we define the interaction total activity of node i by

$$\chi_t^2(i) = \chi_1^2(i) + \chi_2^2(i) \quad \text{(Interaction total activity)} \quad (6)$$

with $v_t(i) = v_1(i) + v_2(i)$ d.f.

Our previous CPχ^2 work on *single* interaction comparative chi-square analysis (41) established interaction heterogeneity $\chi_d^2(i)$ for each node in a network as follows:

$$\chi_d^2(i) = \chi_t^2(i) - \chi_c^2(i) \quad \text{(Interaction heterogeneity)} \quad (7)$$

which measures how far the interactions deviate from the pooled version. We proved that $\chi_d^2(i)$ asymptotically follows a chi-square distribution with $v_d(i) = v_t(i)-v_c(i)$ d.f. when no interactions exist across the conditions (41). Here, parent/cause and child/effect variables are provided on the input subnetwork topology to ChiNet analysis. When a parent–child relationship is one-to-one with no time delay, interaction heterogeneity chi-square statistic $\chi_d^2(i)$ is symmetrical with respect to the parent and the child. Thus, $\chi_d^2(i)$ reflects changes in the parent–child relationship but does not necessarily make a statement on causality. When a many-to-one or a temporally delayed parent-child relationship is considered, $\chi_d^2(i)$ is asymmetric and may reflect causal changes in such a relationship.

Now, we extend the three chi-square statistics from the individual interaction level to the subnetwork level. Here, the null hypothesis is no interactions in the subnetwork are active under any condition. We first assume interactions in a subnetwork are statistically independent with respect to each node. We define, across all conditions, the subnetwork total activity chi-square and associated d.f. by

$$\chi_T^2 = \sum_{i \in V} \chi_t^2(i), \quad v_T = \sum_{i \in V} v_t(i) \quad \text{(Subnet total activity)} \quad (8)$$

which is chi-square distributed with v_T d.f. under the null hypothesis. Under the same null hypothesis, we define the subnetwork homogeneity χ_C^2 by

$$\chi_C^2 = \sum_{i \in V} \chi_c^2(i), \quad v_C = \sum_{i \in V} v_c(i) \quad \text{(Subnet homogeneity)} \quad (9)$$

χ_C^2 is chi-square distributed with v_C d.f. Now we define subnetwork heterogeneity χ_D^2 by

$$\chi_D^2 = \sum_{i \in V} \chi_d^2(i), \quad v_D = \sum_{i \in V} v_d(i) \quad \text{(Subnet heterogeneity)} \quad (10)$$

which is chi-square distributed with v_D d.f. All three subnetwork statistics are chi-square distributed, because the sum of independent chi-square random variables is also chi-squared with d.f. being the sum of d.f. for each node (44). From Equation (7), it follows that we can decompose subnetwork total activity χ_T^2 into the sum of subnetwork heterogeneity χ_D^2 and homogeneity χ_C^2. This gives us the subnetwork decomposition rule in Equation (1). We use p_T, p_C and p_D to represent the *P*-values of test statistics χ_T^2, χ_C^2 and χ_D^2, respectively, calculated from a given sample.

In establishing the chi-square null distributions for the above three subnetwork statistics, we have assumed interactions in a subnetwork be statistically independent. Even in an inactive subnetwork of connected nodes where we only observe noise, however, the noise dynamics can be dependent among the nodes. Thus, chi-square approximation of the null distributions can be inaccurate. Chuang and Shih (45) assumed individual chi-squares of 2 d.f. and estimated their correlation coefficients to approximate the dependent chi-square sum by a scaled chi-square distribution. When

sample sizes are limited, we found that individual interaction chi-squares of some nodes are not chi-square distributed. Additionally, estimation of the correlation matrix of individual chi-squares tends to be inaccurate. To overcome these issues, we present a gamma approximation for the sum of dependent chi-squares using a bootstrap strategy as shown in Supplementary Algorithm S1. CHINET-GAMMA-BOOTSTRAP.

To correct for the statistical effect of simultaneous testing on multiple subnetworks, we apply Bonferroni correction to control family-wise error rate or Benjamini–Hochberg (46) to control false discovery rate. Both are conservative in *P*-value adjustment but computationally efficient. Finally, we determine if a subnetwork is rewired or conserved based on p_D and p_C. Let α be specified maximum acceptable type I error. The subnetworks are rewired, if $p_D \leq \alpha$; or conserved, if $p_D > \alpha$ and $p_C \leq \alpha$. It is useful to point out that two rewired subnetworks may have strong heterogeneity and homogeneity at the same time; two conserved subnetworks can only have strong homogeneity.

Performance evaluation of ChiNet and comparison with other methods using simulated and experimental benchmark datasets

We first evaluated the performance of ChiNet in reference to GSCA and GSNCA, two differential-correlation based subnetwork rewiring methods, on simulated yeast transcription-metabolic networks. Then we benchmarked ChiNet, GSCA and GSNCA under 459 simulation settings associated with four network characteristics: noise level, sample size, complexity of dynamics (number of quantization levels) and subnetwork sparsity (number of parents, or in-degree, per child node). Both studies used a house noise model (Supplementary Note S1). Finally, we evaluated ChiNet, GSCA and GSNCA to identify rewired subnetworks among mitochondria ribosome protein (MRP), cytoplasmic ribosome protein (RP), rRNA genes and their TFs using microarray gene expression data collected from two yeast species fungus pathogen *C. albicans* and *S. cerevisiae*. Full detail about the three performance evaluation studies is described in Supplementary Notes S2, S3, and S4.

Biological experimental design and data collection

An industrial yeast strain *S. cerevisiae* NRRL Y-12632 and its inhibitor-tolerant derivative NRRL Y-50049 obtained through evolutionary engineering (Agricultural Research Service Culture Collection, Peoria, IL, USA) were used in this study. Experimental design, microarray gene expression, outlier processing, normalization, discretization and gene selection are described in full detail in Supplementary Note S5.

Backtracking upstream transcription subnetworks from downstream metabolic pathways

To identify upstream transcription subnetworks that may have induced downstream metabolic responses during adaptive growth against biomass conversion inhibitors in yeast, we backtrack shortest paths linking a differential transcription interaction to genes with differential

metabolic responses in Y-50049. From YEASTRACT, we identified 183 TFs and 42,524 TF-gene pairs of documented transcription regulatory interactions. This constitutes the known network topology for our study. Using this TF network topology, we first build a super graph that fits the data of the two strains the best using a chi-square test (43). Then we detect all differential gene interactions between the two strains. For every differentially expressed gene on a given downstream metabolic pathway, we find a shortest path to this gene from the closest upstream TF that is involved in a differential interaction by Dijkstra's algorithm. A subnetwork is obtained by joining all such shortest paths reaching a common metabolic pathway. Then we assess by ChiNet if the subnetwork is statistically significantly rewired across the two strains. The algorithm is presented as Supplementary Algorithm S2. BACKTRACK-REWIRED-SUBNETWORKS. The underlying assumption is that a differentially expressed enzyme is caused by the most adjacent rewired upstream transcription regulation.

Biological validation

For biological validation on tolerance impact of TF genes detected by this study, we examined six single gene deletion mutations from *Saccharomyces* Genome Deletion Sets for growth response to challenges of 10 mM each of furfural and HMF on a synthetic medium. A wild-type *S. cerevisiae* strain BY4742 (*MAT α his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*) grown with and without the inhibitor challenges served as a control. Each tested strain was grown on 4 ml synthetic medium in a 15 ml tube at 30°C with agitation of 250 rpm. Cell growth was monitored by absorbance at 600 nm. Cells grown without the inhibitor challenges served as controls. Experiments were repeated for all tests.

RESULTS

Advantage of ChiNet by *in silico* benchmarking

We first demonstrate the capability of ChiNet to identify nonlinear differential interaction patterns in *in silico* yeast subnetworks. With 60 known yeast metabolic pathways in KEGG Pathway (42) and their upstream transcription subnetworks from YEASTRACT (47), we artificially created 60 pairs of rewired and another 60 pairs of conserved dynamic subnetworks using the generalized logical network (GLN) model (43). From these models, we simulated dynamic data at different levels of noise. Here we compare ChiNet, GSCA and its two variants, and also GSNCA. Extending the original GSCA based on linear correlation, the GSCA-order1 variant examines temporal dependencies and the GSCA-Spearman variant integrates temporal dependencies, subnetwork topology and non-linear correlation. On data from each pair of subnetwork models, we applied ChiNet, the GSCA cohort and GSNCA to determine if the pair of underlying subnetworks is rewired or conserved. Figure 2 shows the advantage of ChiNet over the GSCA cohort and GSNCA in receiver operating characteristic (ROC) curves over a wide range of noise levels. The area under the ROC curve (AUROC) of value 1 indicates a perfect performance, 0.5 for a random guess and 0 for a systematic error. As the noise level inflates from 0.2 to 0.45, the
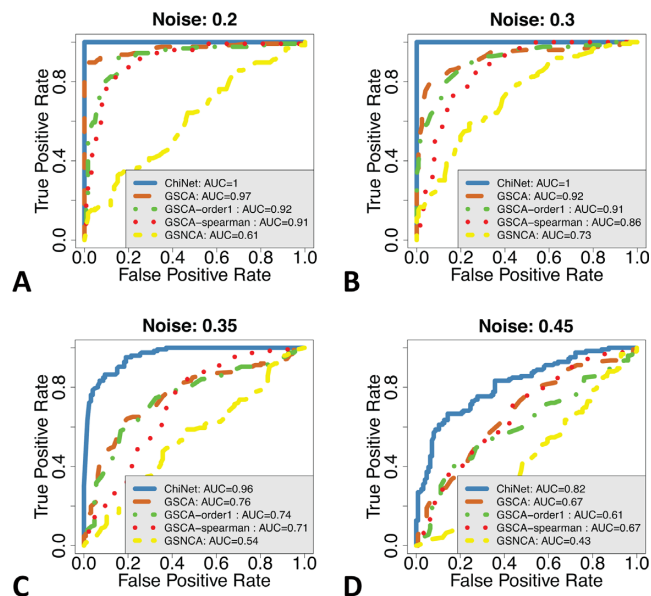

Figure 2. Advantage of ChiNet in detecting rewired upstream transcription subnetworks for downstream metabolic pathways in yeast. The results were obtained by applying ChiNet, GSCA and GSNCA methods on *in silico* data simulated from 120 pairs of rewired or conserved dynamic subnetworks from the known *Saccharomyces cerevisiae* transcription network and metabolic pathways. A true positive is a pair of rewired subnetworks detected as so; a false positive is a pair of conserved subnetworks detected wrongly as rewired. Thus, ChiNet demonstrates outstanding robustness to noise over the GSCA cohort and GSNCA, evidenced by the increasing advantage in area under the ROC curves over deteriorating noise. In the shown four settings, quantization level (3), sample size (15 × 5) and number of parents (1) are all fixed. The noise levels are (**A**) 0.2, (**B**) 0.3, (**C**) 0.35 and (**D**) 0.45, respectively.

gain in AUROC by ChiNet over the GSCA cohort also increases from 0.01–0.05 to 0.18–0.27; GSNCA did not function well, with AUROC notably less than ChiNet or GSCA at all noise levels in this study. Thus this result demonstrates potentially outstanding robustness of ChiNet to noise in realistic biological networks.

To evaluate sensitivity of ChiNet, GSCA and GSNCA to various subnetwork characteristics, we benchmarked their performance by a second simulation study. We generated a total of 91 800 pairs of subnetworks under 459 simulation settings characterized by noise level, complexity of interaction dynamics (as indicated by the number of quantization levels), sample size and sparsity of network topology (as indicated by the in-degrees of each node). Figure 3a illustrates ROC curves of the three types of methods at a specific simulation setting, demonstrating the notable strength of ChiNet. The distributions and box plots of empirical AUROC for each method are shown in Figure 3b and c. The mean AUROC over all 459 settings in decreasing order was observed at 0.77 for ChiNet, followed by 0.60, 0.63 and 0.64 for GSCA, GSCA-order1, GSCA-Spearman and 0.53 for GSNCA, respectively.

Thus, ChiNet here demonstrates a large margin of effectiveness over differential-correlation based methods.

The fundamental limitation of differential correlation employed by GSCA and GSNCA is illustrated by an example in Figure 4. As GSCA uses the dispersion index—the
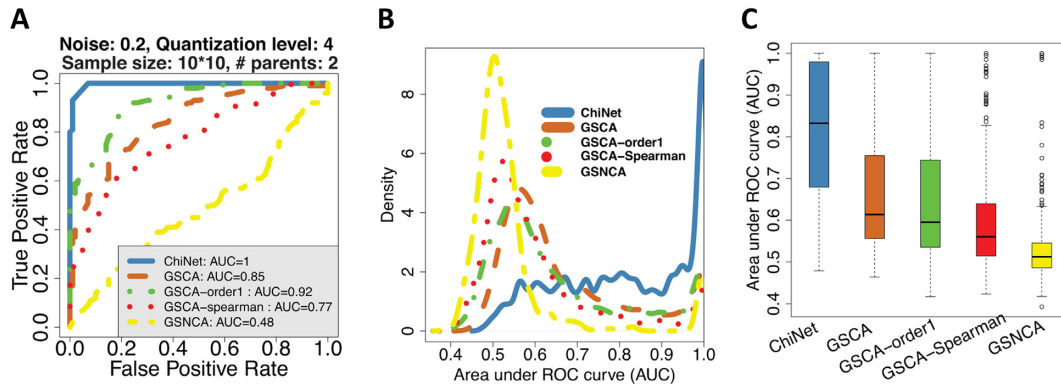
Figure 3. Advantage of ChiNet over the GSCA cohort and GSNCA on subnetwork rewiring with varying characteristics of noise level, complexity of interaction dynamics (number of quantization levels), sample size and sparsity (in-degrees). In each of 459 simulation settings, 200 pairs of subnetwork using generalized logical network (GLN) models were randomly generated: 100 pairs conserved and 100 pairs rewired. These form 200 pairs of trajectory collection files as input to all four methods. (**A**) An example of ROC curves under one simulation setting, where each GLN contains 10 nodes of 4 quantization levels and two parents and 10 trajectories of 10 time points are simulated from each GLN at a noise level of 0.2. ChiNet performs the best with the largest AUC of 0.996, substantially higher than the GSCA cohort and GSNCA. The original GSCA (AUC = 0.716) performed worse than both GSCA-Spearman (AUC = 0.862) and GSCA-order1 (AUC = 0.865). GSNCA achieved the lowest AUC of 0.48. (**B**) Distributions of area under the curves (AUCs), as empirical probability density functions of AUC for ChiNet, GSCA, GSCA-order1, GSCA-Spearman and GSNCA over 459 simulation settings. A density function distributed toward 1 in AUC is desirable for good performance. ChiNet (blue) is evidently the best performer. Although at high noise levels, all methods performed close to random guessing (AUC = 0.5), the density of ChiNet AUC is unarguably the lowest around AUC = 0.5 and the highest around AUC = 1, thus the best among the methods studied. (**C**) Box plots of the AUCs over the 459 settings. The ChiNet demonstrates an evident advantage in the AUCs as judged by the quartiles. About 60–75% of examples have an AUC >0.7 but 75% of the AUCs of the GSCA cohort or GSNCA are <0.7.
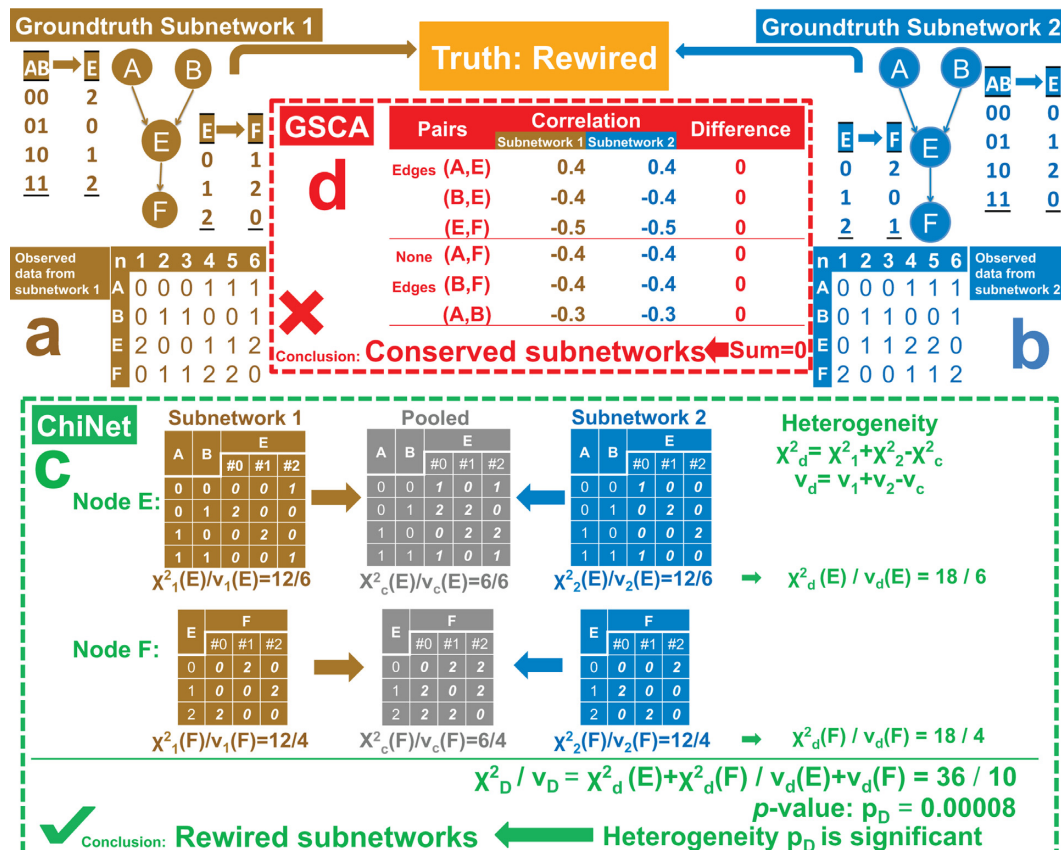
Figure 4. ChiNet overcomes a fundamental problem of differential-correlation based approaches for subnetwork rewiring. (**a** and **b**) Two 4-node ground-truth subnetworks of zero Markovian order are differential because they have different generalized truth tables for both nodes E and F. Observed data from both networks are also given. (**c**) ChiNet correctly declares the subnetworks significantly rewired ($P_D \leq 0.05$) by accumulating interaction heterogeneity over each node. (**d**) GSCA wrongfully concludes the subnetworks are conserved, based on the sum of the squares of differential correlation coefficients over all pairs of nodes being zero.

summation of the squares of differences in pairwise correlation coefficients between conditions—to evaluate network rewiring, it can be insensitive to complex pattern differences in rewired subnetworks. In this example, truly rewired subnetworks scored an undesired zero dispersion index (Figure 4). Although GSNCA uses L_1 norm and weighs differential correlation discriminatively for each node in the subnetwork, it still shares the same limitation with GSCA as correlation coefficients are subtracted in both methods. As a result, ChiNet outperformed the GSCA cohort and GSNCA at a markedly large margin.

Validating ChiNet using transcription subnetworks rewired between two yeasts

To understand how evolution may have rewired gene regulatory networks connecting TFs and their target genes between *C. albicans* and *S. cerevisiae*, we applied ChiNet to gene subsets that contain either diverged or conserved sequence motifs in their promoter regions on a gene expression microarray compendium including 1011 *S. cerevisiae* and 198 *C. albicans* samples (5). Loss of *cis*-regulatory elements for MRP genes due to genome evolution has been linked to rapid anaerobic growth in *S. cerevisiae* relative to other aerobic yeast species (5). Although gene clusters corresponding to differentially correlated expression patterns have been identified, expression patterns of these clusters do not directly suggest TF-gene rewiring. Both rewired and not-rewired genes expressed differentially between the two species (Supplementary Figures S11, S12 and S14). Meanwhile, their known TFs are equally enriched in the two species (Supplementary Figures S13 and S14). This implies that analyzing gene set enrichment by differential expression without looking at interaction patterns here would not logically lead to evidence for rewiring. Thus, we inspected rewired interaction patterns in subnetworks.

Our analysis (Supplementary Table S1) shows that the transcription subnetwork connecting MRP genes and their TFs are highly rewired with a normalized subnetwork heterogeneity chi-square of 53 (*P*-value = 0) between *C. albicans* and *S. cerevisiae* (Figure 5a). On the other hand, the transcription subnetwork connecting cytoplasmic ribosome protein (RB) and rRNA genes and their TFs are mostly not rewired (Supplementary Figure S2) with a normalized subnetwork heterogeneity chi-square of 10 (*P*-value = 4 × 10^{-11}). These findings by ChiNet are consistent with and complementary to transcription regulation rewiring suggested by the extent of sequence motif conservation (5). The rewired MRP gene regulation most likely contributes to the different capability for rapid anaerobic growth of *S. cerevisiae* versus aerobic growth of *C. albicans*.

Using this dataset, we again found ChiNet remarkably outperformed GSCA and GSNCA using the partially confirmed rewired and conserved genes between the two yeasts as a gold standard (5) (see Supplementary Note S4). ROC and precision-recall (PR) curves for all three methods (Supplementary Figure S3 to S7) were plotted under five values of subnetwork rewiring heterogeneity, which is defined as the ratio of rewired genes to the total number of genes excluding TFs in the subnetwork. Figure 5b and c shows AU-ROC and area under PR (AUPR) as a function of subnet-

work rewiring heterogeneity for the three methods. ChiNet exhibits a highly consistent advantage over GSCA and GSNCA at increasing subnetwork rewiring heterogeneity. Contrary to the two simulation studies, GSNCA performed better than GSCA here and demonstrates an advantage due to the implicit use of subnetwork topology. The dramatic under-performance of GSCA and GSNCA (Figure 5b and c) can be partially explained by false positives introduced by large differential correlations when the dynamic range of genes in one condition is fully covered by a larger dynamic range of another condition. In Figure 5d, e, and f, the expression pattern between transcription factor gene XBP1 and a target gene EBP2 (in the not-rewired gene group) in *C. albicans* is almost entirely enclosed within the pattern of *S. cerevisiae*. ChiNet did not score the two patterns high for rewiring because they do not contradict each other. However, the large differential XBP1-EBP2 correlation of $|-0.51 - 0.11| = 0.62$ would amount to falsely strong evidence for a rewired interaction across the two yeasts. There are a number of genes with overlapping dynamic interaction patterns in the not-rewired gene group and thus led to the poor overall performance of GSCA and GSNCA.

### Globally rewired gene networks in the tolerant yeast

Applying ChiNet on transcriptome data of yeast in response to furfural and HMF, we found that the tolerant strain Y-50049 displayed significant alterations on gene regulatory networks at the global scale compared with its parental wild-type strain Y-12632. At least 44 pathways (Supplementary Table S2) were detected to significantly involve rewired upstream transcription subnetworks (Figure 6).

The oxidative phosphorylation pathway was detected to have the greatest differential expression between the two strains as suggested by its highly significant working zone change *P*-value (Supplementary Note S6). In addition to important central metabolic pathways, almost all amino acid metabolic pathways were affected, which represent comprehensive alterations of biosynthesis activities in the tolerant yeast. Other downstream pathways significantly affected by their upstream transcription subnetworks were involved in fatty acid metabolism and glycerolipid metabolism.

Transcription factor gene *YAP1* appeared to be the most dominant regulatory gene for Y-50049 in adaptation to the toxic compounds furfural and HMF. Its adaptive signature expression impacted at least 39 downstream pathways. Among them, the glycolysis and pentose phosphate pathways showed high statistical significance in upstream transcription subnetwork heterogeneity (Supplementary Table S2). The pentose phosphate pathway has a highly significant rewired upstream transcription network (*P*-value 1.97 × 10^{-14}) between the tolerant yeast strain Y-50049 and the wild-type (Figure 7). Eighteen TFs are involved and most rewired TF-enzyme interactions originated from *YAP1* and *IFH1*.

Another key regulatory gene *RPN4* was observed to be adaptively activated and affected more than 20 downstream pathways through enhanced activity of at least three downstream interactions of *RPN4→YOX1*, *REB1→RAP1* and
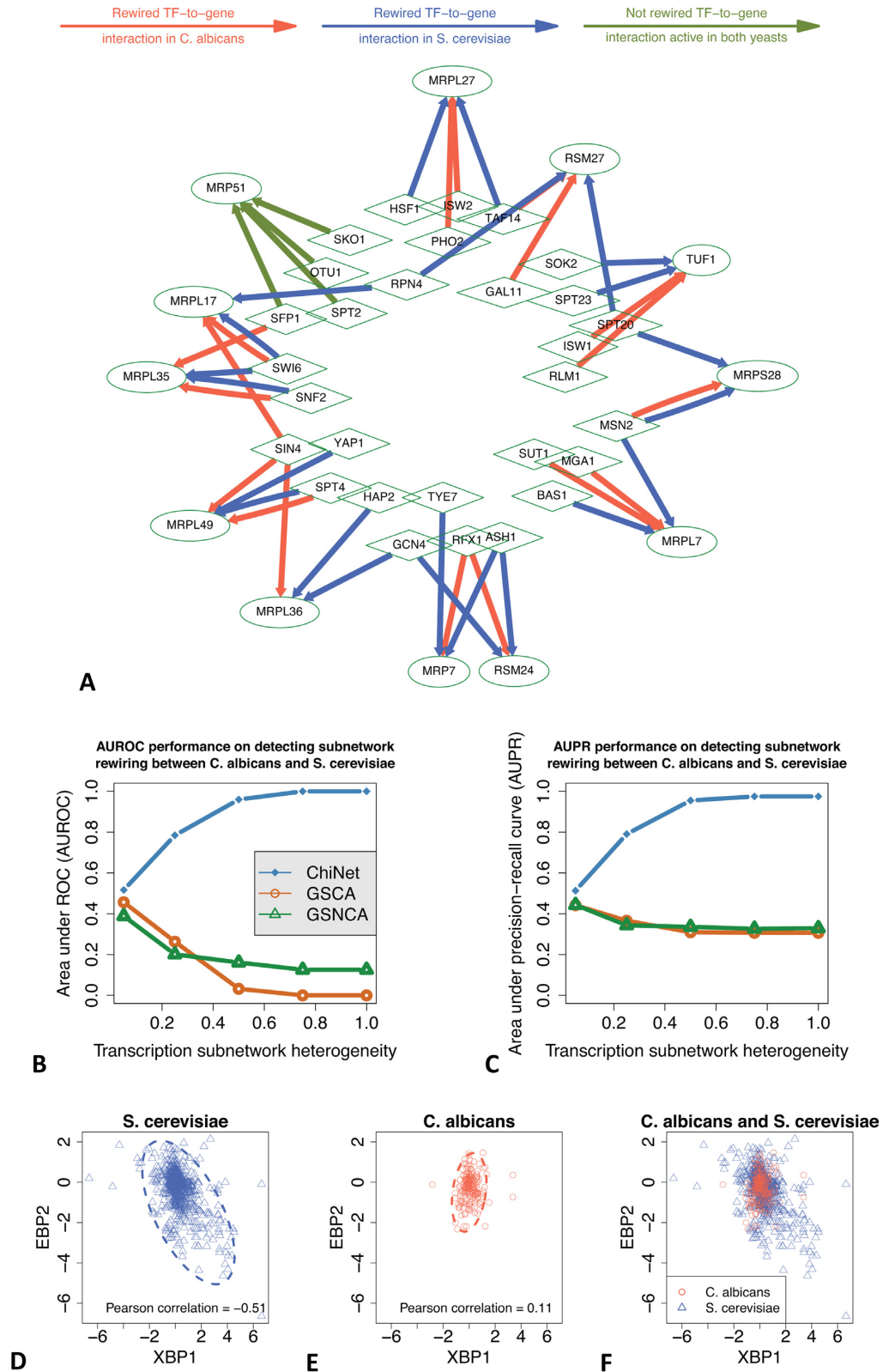
Figure 5. ChiNet uncovered rewired transcription subnetworks between *Candida albicans* and *Saccharomyces cerevisiae* and remarkably outperformed GSCA and GSNCA. (**A**) The most strongly rewired transcription subnetwork from 23 mitochondria ribosome protein (MRP) genes (oval) and their TF genes (diamond). The red edges represent ChiNet-detected transcription regulation rewiring in *C. albicans* and blue for those rewired in *S. cerevisiae*. The green edges are detected conserved interactions in both *C. albicans* and *S. cerevisiae*. (**B**) AUROC as a function of subnetwork heterogeneity for the three subnetwork rewiring methods—ChiNet, GSCA and GSNCA. (**C**) AUPR as a function of subnetwork heterogeneity for each method. (**D**–**F**) A not-rewired interaction XBT1→EBP2 can be a false positive by differential correlation. The ovals indicate the interaction patterns in the two species respectively. The interaction pattern of gene EBP2 and its TF XBP1 in *C. albicans* is almost fully covered by the interaction pattern in *S. cerevisiae* and did not show strong evidence for rewiring.
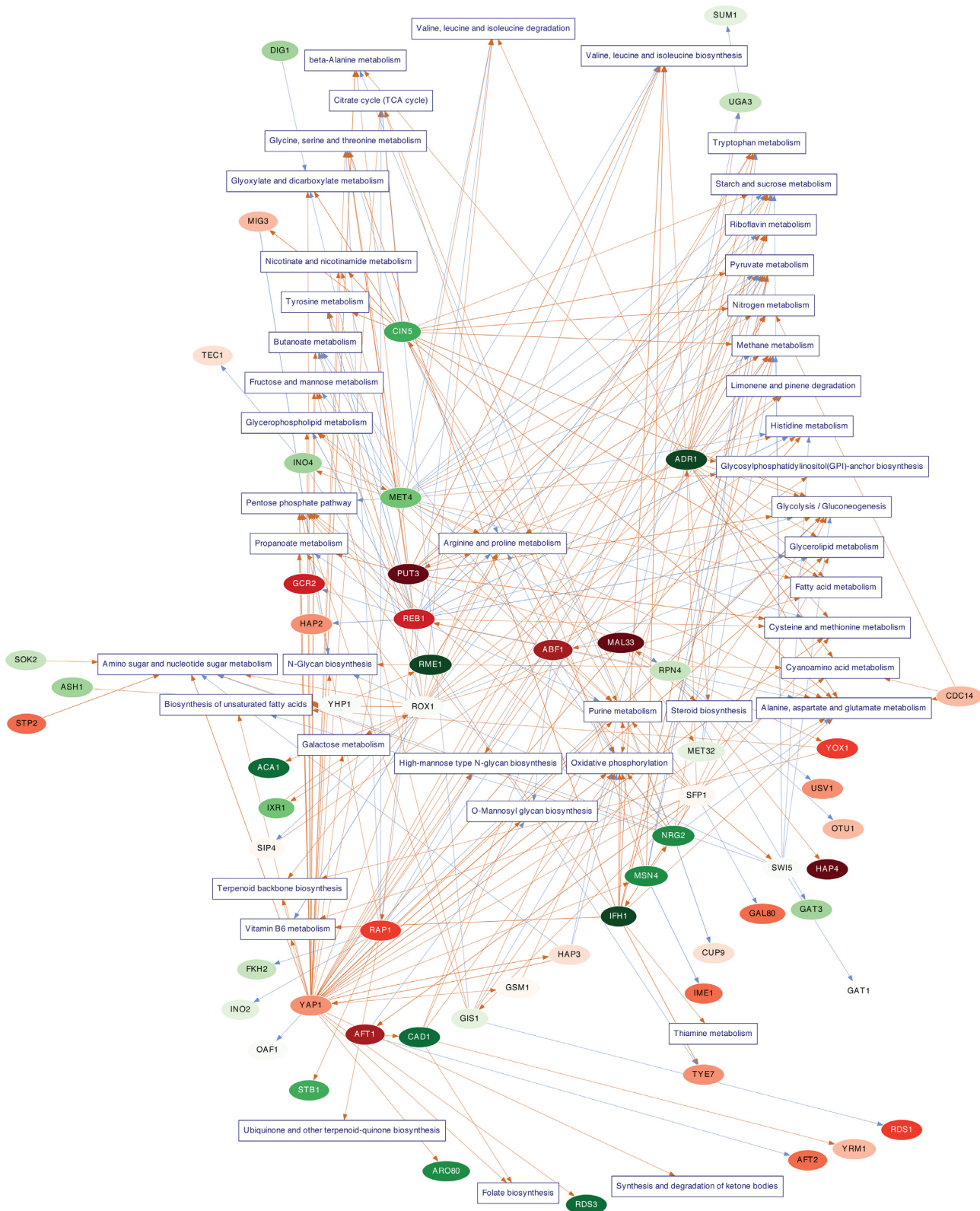
Figure 6. The rewired transcription network regulating metabolic pathways for tolerant industrial yeast *Saccharomyces cerevisiae* NRRL Y-50049. The network is formed by joining rewired upstream transcription subnetworks to downstream metabolic pathways in the tolerant strain Y-50049 versus the control during the lag phase of cell growth. Oval nodes represent over—(red) or under—(green) expressed transcription factors in Y-50049; the darkness is proportional to the level of over- or under-expression. Boxes represent metabolic pathways. An edge from a transcription factor to a metabolic pathway box denotes a significant interaction from the transcription factor to an enzyme-encoding gene on the pathway. Brown/Blue edges are statistically significantly rewired/conserved interactions across the two strains. All subnetworks before joining have adjusted heterogeneity *P*-value $P_D \leq 0.05$.
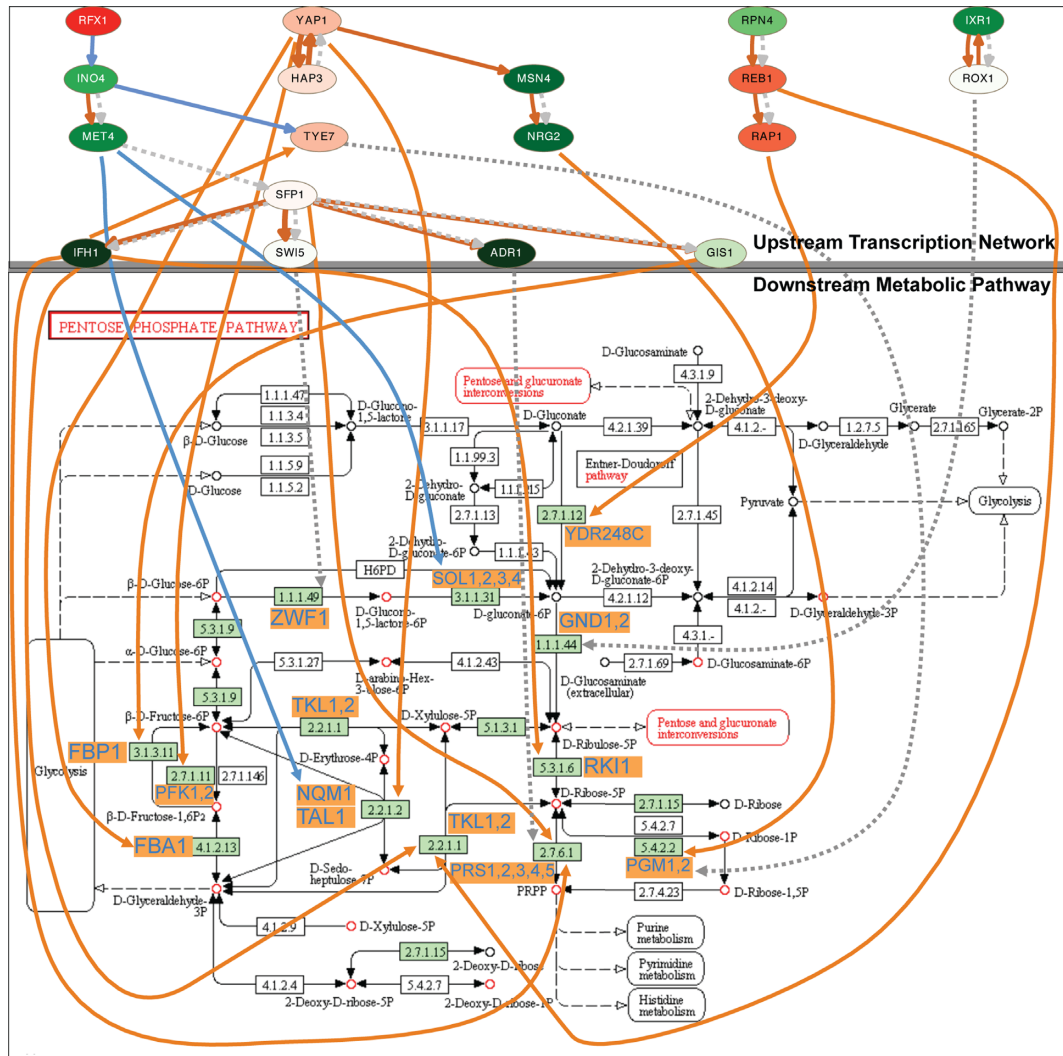
Figure 7. The rewired upstream transcription subnetwork in regulating the downstream pentose phosphate pathway in biomass conversion in Y-50049 yeast strain versus the wild-type. In the rewired upstream transcription subnetwork identified by ChiNet (P-value $= 1.97 \times 10^{-14}$), orange edges are significantly rewired transcription regulation; blue edges are significantly conserved transcription regulation; and short dashed edges are known transcription regulation but may be inactive. Oval nodes represent known TFs: those over-expressed in Y-50049 are in red, darker for strongly over-expressed; those under-expressed in Y-50049 are in green, darker for strongly under-expressed. In the downstream metabolic pathway, the square boxes labeled with enzyme commission numbers represent enzymes. Those in green are annotated. Those with an accompanying orange box with gene name were measured in our gene expression experiment. The open circles indicate metabolites. Edges associated with boxes and circles together represent enzymatic reactions. The pentose phosphate pathway diagram is based on KEGG Pathway (42).

MAL33→ABF1. However, altered regulatory interactions observed in Y-50049 genome adaptation were not limited to enhanced gene expression. As indicated by the rewired networks, regulatory genes with normally expressed and downregulated expression may also serve regulatory functions in adaptation to the furfural-HMF stress. The activated TF gene *SFP1* rooted more than 30 downstream pathways, including major biosynthesis and central metabolic pathways, through differential and conserved interactions including several regulatory genes with downregulated expression (Figure 6). For example, downstream of *SFP1*, TF gene *IFH1* was observed to be repressed but led to activation of *TYE7* and altered downstream interactions, including many amino acid metabolism pathways. TF gene *ROX1* also appeared to play an important role in the yeast adap-

tation involving at least 20 pathways. Under the challenge of furfural and HMF, *ROX1* was normally expressed mediating both altered and conserved interactions of genes and pathways. Downregulated TF gene *CIN5*, served as both a regulon and a regulator, was linked to up-regulated gene responses and conserved pathway interactions.

We performed single-gene-deletion mutations on selective TF genes including *YAP1, RPN4, MSN4, ROX1, SFP1* and *CIN5* to confirm gene functions in response to the toxic compounds. Strains with these mutations were all able to grow normally on a minimum medium (Supplementary Figure S8A). But when the medium was supplemented with furfural and HMF these strains were significantly repressed or unable to grow compared with a wild type control (Supplementary Figure S8B). For example,

mutation strains with *YAP1* knockout failed to grow on an inhibitor-containing medium at 72 h. TF gene *YAP1* activates response of anti-oxidant genes by recognizing a Yap1p response element (YRE), 5′-TKACTMA-3′, in the promoter region. *YAP1* was identified as a major responsible regulator for yeast tolerance to the inhibitors (4,48–49). Many genes showing induced expression possess the YRE sequence in their promoter region. Most *YAP1*-regulated genes were classified in a broad range of functional categories including redox metabolism, amino acid metabolism, stress response, DNA repair and others. Modified responses of glucose metabolic pathways for the tolerant yeast were defined in detail involving many genes with reductase activity and four major cofactor regeneration steps (2,50). Recent results of engineering efforts were consistent with these findings (51,52). While a wild-type strain was repressed to die under challenge of the toxic chemicals, the tolerant strain equipped with the reprogrammed glucose metabolic pathways detoxified the inhibitors *in situ* and produced ethanol. Glycolysis is one of the central metabolic pathways for cell survival and function. Our results of this study identified glycolysis as one of the most significant downstream pathways likely to be affected by rewired regulation and co-regulation of *YAP1*.

In addition to the indispensable *YAP1*, we found TF gene *SFP1* and *RPN4* as key regulatory genes affect downstream metabolic pathways such as pentose phosphate pathway (Figure 7) and many amino acid metabolism pathways (Figure 6; Supplementary Table S1). A functional pentose phosphate pathway is necessary for yeast tolerance involving both detoxification and damage repairs (2,53–59). In this study, we found the tolerant response of this pathway was mediated by altered *RPN4* expression and through more downstream regulatory interactions including *REB1* and *RAP1*. Chemical stress causes reactive oxygen species and damages RNA and protein conformation leading to protein unfolding and aggregation (54,60). Many candidate genes were found to have a proteasome-associated control element of Rpn4p in promoter regions and are potentially regulated by *RPN4* (4,61–62). Adapted *RPN4* expression by the chemical stress in the tolerant yeast apparently played a major regulatory role leading to a functional pentose phosphate pathway as suggested by this study. The highly sensitive response of these gene deletion mutation strains to the toxic chemicals further confirmed the essential roles of each gene involved in the rewired programs for the tolerant yeast. Our results suggest these TF genes are essential regulators for the yeast survival against the toxic compounds.

DISCUSSION

ChiNet developed in this study is innovative in pooling samples from all conditions to one contingency table, where conserved patterns reinforce each other while differential patterns cancel out. This allows ChiNet to detect fundamental interaction pattern rewiring in subnetworks that drive observed differential expression. Accurate detection of subnetwork rewiring enables a specific component in a network to be linked to changed biological function due to evolution.

ChiNet consistently outperformed previously reported methods based on differential correlation, including GSCA, its variants and GSNCA in both simulation and real experimental data studies. We observed that two GSCA variants with more ground-truth information did not improve much over the original GSCA, which was unexpected. It is known that zero differential correlation is neither a sufficient nor a necessary condition for the same slope for two linear patterns (63). The overall similarity in performance among GSNCA, GSCA and its variant GSCA-order1 suggests that differential correlation may constitute the bottleneck, despite the correct Markovian order being used in GSCA-order1. The GSCA-Spearman variant computes nonlinear Spearman correlation coefficients with correct subnetwork topology, but was not able to improve upon GSCA or GSCA-order1. A possible explanation is that a pair of nodes indirectly connected via a path can still be sensitive to differential correlation along the path. In addition, non-linear correlation coefficients may compress interaction patterns even further, resulting in true differential interaction patterns mapping to a similar value and becoming indistinguishable. For example, all monotonically increasing patterns representing very different interaction dynamics display equal Spearman correlation coefficients. Although the recently developed GSNCA method showed improved capability over GSCA through implicitly integrated network topology, our results suggest that summarizing an interaction pattern by correlation coefficient followed by comparing the statistic across conditions is fundamentally ineffective to capture diverse interaction patterns that may share similar correlation coefficient values.

Integration of the rewired subnetworks into a global network (Figure 6) enabled us to zoom into a small number of highly involved TF genes as hubs. Although only *YAP1* has been elucidated to be activated in oxidative responses specifically due to furfural and HMF (64) as thiol-reactive electrophiles (65), many of the hub genes, yet to be studied for their biochemistry with furfural and HMF, are known to be involved in various stress responses in yeast. Another hub gene *RAP1* coordinates *IFH1* binding to ribosome protein to regulate protein synthesis in response to growth stimuli and environmental stresses (66). *RAP1* also directly regulates *YDR248C* in the pentose phosphate pathway (Figure 7). The protein product of *PUT3*, with rewired links to the proline metabolism pathway, activates *PUT1* and *PUT2* which encode enzymes of proline utilization (67). Proline is a protectant against stresses including freezing, desiccation, oxidation and ethanol in yeast (68). *MAL33* is activated at low glucose levels so that other sources of sugar such as maltose and galactose can be utilized (69). *CIN5*, with sequence homologous to *YAP1*, increased resistance to Cisplatin when over-expressed in *S. cerevisiae* (70). *RPN4* promotes DNA repair, antioxidant response and glucose metabolism under genotoxic stresses (71). Direct evidence (72) implicated *MSN2*/*MSN4* in induction of *NTH1* that controls trehalose hydrolysis under heat and osmotic stresses. *ROX1* is a transcriptional regulator of oxidative responses and up-regulated about 30 genes involved in cell stress under anaerobic conditions (73). Together, the functional coincidence of these TF genes in vari-

ous stress responses supports the global view highlighted by our ChiNet analysis that ties these genes and their downstream metabolic pathways. The absence of a single upstream hub gene seems to imply that multiple genomic loci may be responsible for the rewired transcription program in Y-50049. Thus, such rewiring most likely confers tolerance of toxic furfural and HMF on the yeast strain Y-50049.

Several considerations are to be made before applying ChiNet to detect subnetwork rewiring. First, the sample size needs to be sufficient such that the expected number of samples in each contingency table entry is at least five. Second, ChiNet discretizes molecular abundance to represent flexibly not-well-understood interaction patterns so as to prevent biases associated with an unvalidated parametric model. Quantization despite sacrifice in data resolution can be beneficial due to its noise removal effect, as demonstrated in Supplementary Note S4 where the minimal quantization level of two achieved the best performance in both AUROC and AUPR. If interaction patterns already have valid parametric models, information loss due to discretization can be avoided by alternative methods. For example, comparative dynamical system modeling (63) characterizes interaction heterogeneity from continuous time course data with nonlinear parametric models. Finally, ChiNet requires a subnetwork topology in the form of a directed graph. If a topology is unavailable but subnetwork rewiring analysis is desired, one can use a network inference method as a first step to reconstruct the topologies from observed data. Then as the second step in this workflow, one applies ChiNet to perform subnetwork rewiring analysis.

ChiNet is applicable to integrate both proteome and transcriptome data. Specifically, one can use the protein abundance of TFs as parent variables and RNA abundance of target genes as child variables in contingency tables to compute the interaction heterogeneity chi-square statistic defined in Equation (7). On transcriptomic data alone without protein activity information, the outcome of rewiring analysis may be incomplete when post-transcription regulation leads to non-monotonic or even random translational patterns. For example, a recent study on the proteogenome of human colon and rectal cancers revealed a positive sample-wise mRNA–protein correlation but also observed that mRNA abundance is not a reliable predictor of protein variation for individual genes (74). However, highly positively correlated mRNA and protein abundance were observed for most genes in fission yeast (75). Our benchmark study using transcriptome data alone from *C. albicans* versus *S. cerevisiae* by ChiNet highlighted subnetwork rewiring consistent with changed genotypes. The value of ChiNet lies in pointing to such candidate subnetworks for further investigation.

In conclusion, our benchmark studies have demonstrated substantially improved subnetwork rewiring analysis accuracy of ChiNet over alternative methodologies. Further, ChiNet revealed transcription subnetwork rewiring of the molecular mechanisms underlying yeast tolerance and robust strains development for advanced biofuels production. ChiNet is readily applicable to integrate transcriptomic, proteomic and metabolomic data to understand network rewiring fundamental to the evolution of biological systems.

AVAILABILITY

The ChiNet software is freely available to non-commercial users at www.cs.nmsu.edu/~joemsong/software/ChiNet.

ACCESSION NUMBERS

The tolerant yeast microarray data have been deposited in NCBI's Gene Expression Omnibus (76) and are accessible through GEO Series accession number GSE50492 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE50492).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Sun,M.G., Sikora,M., Costanzo,M., Boone,C. and Kim,P.M. (2012) Network evolution: rewiring and signatures of conservation in signaling. *PLoS Comput. Biol.*, **8**, e1002411.
2. Liu,Z.L., Ma,M. and Song,M. (2009) Evolutionarily engineered ethanologenic yeast detoxifies lignocellulosic biomass conversion inhibitors by reprogrammed pathways. *Mol. Genet. Genomics*, **282**, 233–244.
3. Liu,Z.L., Moon,J., Andersh,B., Slininger,P.J. and Weber,S. (2008) Multiple gene-mediated NAD(P)H-dependent aldehyde reduction is a mechanism of in situ detoxification of furfural and 5-hydroxymethylfurfural by *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.*, **81**, 743–753.
4. Ma,M. and Liu,Z.L. (2010) Comparative transcriptome profiling analyses during the lag phase uncover YAP1, PDR1, PDR3, RPN4, and HSF1 as key regulatory genes in genomic adaptation to the lignocellulose derived inhibitor HMF for *Saccharomyces cerevisiae*. *BMC Genomics*, **11**, 660.
5. Ihmels,J., Bergmann,S., Gerami-Nejad,M., Yanai,I., McClellan,M., Berman,J. and Barkai,N. (2005) Rewiring of the yeast transcriptional network through the evolution of motif usage. *Science*, **309**, 938–940.
6. Tirosh,I., Weinberger,A., Carmi,M. and Barkai,N. (2006) A genetic signature of interspecies variations in gene expression. *Nat. Genet.*, **38**, 830–834.
7. Filtz,T.M., Vogel,W.K. and Leid,M. (2014) Regulation of transcription factor activity by interconnected post-translational modifications. *Trends Pharmacol. Sci.*, **35**, 76–85.

8. Khurana,E., Fu,Y., Colonna,V., Mu,X.J., Kang,H.M., Lappalainen,T., Sboner,A., Lochovsky,L., Chen,J., Harmanci,A. *et al.* (2013) Integrative annotation of variants from 1092 humans: Application to cancer genomics. *Science*, **342**, 1235587.

9. Khatri,P., Sirota,M. and Butte,A.J. (2012) Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comput. Biol.*, **8**, e1002375.

10. Mitra,K., Carvunis,A.-R., Ramesh,S.K. and Ideker,T. (2013) Integrative approaches for finding modular structure in biological networks. *Nat. Genet.*, **14**, 719–732.

11. Khatri,P., Draghici,S., Ostermeier,G.C. and Krawetz,S.A. (2002) Profiling gene expression using onto-express. *Genomics*, **79**, 266–270.

12. Draghici,S., Khatri,P., Martins,R.P., Ostermeier,G.C. and Krawetz,S.A. (2003) Global functional profiling of gene expression. *Genomics*, **81**, 98–104.

13. Backes,C., Keller,A., Kuentzer,J., Kneissl,B., Comtesse,N., Elnakady,Y.A., Müller,R., Meese,E. and Lenhof,H.-P. (2007) GeneTrail–advanced gene set enrichment analysis. *Nucleic Acids Res.*, **35**, W186–W192.

14. Subramanian,A., Tamayo,P., Mootha,V.K., Mukherjee,S., Ebert,B.L., Gillette,M.A., Paulovich,A., Pomeroy,S.L., Golub,T.R., Lander,E.S. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 15545–15550.

15. Yi,M. and Stephens,R.M. (2008) SLEPR: a sample-level enrichment-based pathway ranking method—seeking biological themes through pathway-level consistency. *PLoS One*, **3**, e3288.

16. Irizarry,R.A., Wang,C., Zhou,Y. and Speed,T.P. (2009) Gene set enrichment analysis made simple. *Stat. Methods Med. Res.*, **18**, 565–575.

17. Simon,R., Lam,A., Li,M.-C., Ngan,M., Menenzes,S. and Zhao,Y. (2007) Analysis of gene expression data using BRB-array tools. *Cancer Inform.*, **3**, 11–17.

18. Yi,M., Mudunuri,U., Che,A. and Stephens,R.M. (2009) Seeking unique and common biological themes in multiple gene lists or datasets: pathway pattern extraction pipeline for pathway-level comparative analysis. *BMC Bioinformatics*, **10**, 200.

19. Sartor,M.A., Mahavisno,V., Keshamouni,V.G., Cavalcoli,J., Wright,Z., Karnovsky,A., Kuick,R., Jagadish,H.V., Mirel,B., Weymouth,T. *et al.* (2010) ConceptGen: a gene set enrichment and gene set relation mapping tool. *Bioinformatics*, **26**, 456–463.

20. Poisson,L.M., Sreekumar,A., Chinnaiyan,A.M. and Ghosh,D. (2012) Pathway-directed weighted testing procedures for the integrative analysis of gene expression and metabolomic data. *Genomics*, **99**, 265–274.

21. Hwang,T. and Park,T. (2009) Identification of differentially expressed subnetworks based on multivariate ANOVA. *BMC Bioinformatics*, **10**, 128.

22. Shojaie,A. and Michailidis,G. (2010) Network enrichment analysis in complex experiments. *Stat. Appl. Genet. Mol. Biol.*, **9**, Article ID 22.

23. Draghici,S., Khatri,P., Tarca,A.L., Amin,K., Done,A., Voichita,C., Georgescu,C. and Romero,R. (2007) A systems biology approach for pathway level analysis. *Genome Res.*, **17**, 1537–1545.

24. 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.

25. Gosline,S.J., Spencer,S.J., Ursu,O. and Fraenkel,E. (2012) SAMNet: a network-based approach to integrate multi-dimensional high throughput datasets. *Integr. Biol. (Camb)*, **4**, 1415–1427.

26. Jung,S. and Kim,S. (2014) EDDY: a novel statistical gene set test method to detect differential genetic dependencies. *Nucleic Acids Res.*, **42**, e60.

27. Guo,Z., Li,Y., Gong,X., Yao,C., Ma,W., Wang,D., Li,Y., Zhu,J., Zhang,M., Yang,D. *et al.* (2007) Edge-based scoring and searching method for identifying condition-responsive protein–protein interaction sub-network. *Bioinformatics*, **23**, 2121–2128.

28. Choi,Y. and Kendziorski,C. (2009) Statistical methods for gene set co-expression analysis. *Bioinformatics*, **25**, 2780–2786.

29. Alvo,M., Liu,Z., Williams,A. and Yauk,C. (2010) Testing for mean and correlation changes in microarray experiments: an application for pathway analysis. *BMC Bioinformatics*, **11**, 60.

30. Ma,H., Schadt,E.E., Kaplan,L.M. and Zhao,H. (2011) COSINE: COndition-SpecIfic sub-NEtwork identification using a global optimization method. *Bioinformatics*, **27**, 1290–1298.

31. Gambardella,G., Moretti,M.N., de Cegli,R., Cardone,L., Peron,A. and di Bernardo,D. (2013) Differential network analysis for the identification of condition-specific pathway activity and regulation. *Bioinformatics*, **29**, 1776–1785.

32. Rahmatallah,Y., Emmert-Streib,F. and Glazko,G. (2014) Gene Sets Net Correlations Analysis (GSNCA): a multivariate differential coexpression test for gene sets. *Bioinformatics*, **30**, 360–368.

33. Ideker,T. and Kroganb,N.J. (2012) Differential network biology. *Mol. Syst. Biol.*, **8**, 565.

34. Wall,J.D., Harwood,C. and Demain,D. (2008) *Bioenergy*. American Society for Microbiology Press, Washington DC.

35. Vertes,A.A., Qureshi,N. and Yukawa,H. (2010) *Biomass to Biofuels: Strategies for Global Industries*, Wiley, Chichester.

36. Larsson,S., Palmqvist,E., Hahn-Hägerdal,B., Tengborg,C., Stenberg,K., Zacchi,G. and Nilvebrant,N.-O. (1999) The generation of fermentation inhibitors during dilute acid hydrolysis of softwood. *Enzyme Microb. Technol.*, **24**, 151–159.

37. Klinke,H.B., Thomsen,A. and Ahring,B.K. (2004) Inhibition of ethanol-producing yeast and bacteria by degradation products produced during pre-treatment of biomass. *Appl. Microbiol. Biotechnol.*, **66**, 10–26.

38. Liu,Z.L., Slininger,P.J. and Gorsich,S.W. (2005) Enhanced biotransformation of furfural and hydroxymethylfurfural by newly developed ethanologenic yeast strains. *Appl. Biochem. Biotechnol.*, **121-124**, 451–460.

39. Liu,Z.L. and Blaschek,H.P. (2010) Biomass conversion inhibitors and in situ detoxification. In: *Biomass to Biofuels: Strategies for Global Industries*. Blackwell Publishing Ltd., Chichester, pp. 233–259.

40. Liu,Z.L. (2011) Molecular mechanisms of yeast tolerance and in situ detoxification of lignocellulose hydrolysates. *Appl. Microbiol. Biotechnol.*, **90**, 809–825.

41. Song,M., Zhang,Y., Katzaroff,A.J., Edgar,B.A. and Buttitta,L. (2014) Hunting complex differential gene interaction patterns across molecular contexts. *Nucleic Acids Res.*, **42**, e57.

42. 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.

43. Song,M., Lewis,C.K., Lance,E.R., Chesler,E.J., Yordanova,R.K., Langston,M.A., Lodowski,K.H. and Bergeson,S.E. (2009) Reconstructing generalized logical networks of transcriptional regulation in mouse brain from temporal gene expression data. *EURASIP J. Bioinform. Syst. Biol.*, **2009**, Article ID 545176.

44. Casella,G. and Berger,R.L. (1990) *Statistical Inference*, Duxbury Press, Belmont, CA.

45. Chuang,L. and Shih,Y. (2012) Approximated distributions of the weighted sum of correlated chi-squared random variables. *J. Stat. Plan. Inference*, **142**, 457–472.

46. Benjamini,Y. and Hochberg,Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B*, **57**, 289–300.

47. Teixeira,M.C., Monteiro,P.T., Guerreiro,J.F., Gonçalves,J.P., Mira,N.P., dos Santos,S.C., Cabrito,T.R., Palma,M., Costa,C., Francisco,A.P. *et al.* (2014) The YEASTRACT database: an upgraded information system for the analysis of gene and genomic transcription regulation in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **42**, D161–D166.

48. Lin,F.-M., Qiao,B. and Yuan,Y.-J. (2009) Comparative proteomic analysis of tolerance and adaptation of ethanologenic *Saccharomyces cerevisiae* to furfural, a lignocellulosic inhibitory compound. *Appl. Environ. Microbiol.*, **75**, 3765–3776.

49. Gulshan,K., Lee,S.S. and Moye-Rowley,W.S. (2011) Differential oxidant tolerance determined by the key transcription factor Yap1 is controlled by levels of the Yap1-binding protein, Ybp1. *J. Biol. Chem.*, **286**, 34071–34081.

50. Jordan,D.B., Braker,J.D., Bowman,M.J., Vermillion,K.E., Moon,J. and Liu,Z.L. (2011) Kinetic mechanism of an aldehyde reductase of *Saccharomyces cerevisiae* that relieves toxicity of furfural and 5-hydroxymethylfurfural. *Biochim. Biophys. Acta*, **1814**, 1686–1694.

51. Moon,J. and Liu,Z.L. (2012) Engineered NADH-dependent GRE2 from *Saccharomyces cerevisiae* by directed enzyme evolution enhances HMF reduction using additional cofactor NADPH. *Enzyme Microb. Technol.*, **50**, 115–120.

52. Jayakody,L.N., Horie,K., Hayashi,N. and Kitagaki,H. (2013) Engineering redox cofactor utilization for detoxification of

glycolaldehyde, a key inhibitor of bioethanol production, in yeast *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.*, **97**, 6589–6600.

53. Gorsich,S.W., Dien,B.S., Nichols,N.N., Slininger,P.J., Liu,Z.L. and Skory,C.D. (2006) Tolerance to furfural-induced stress is associated with pentose phosphate pathway genes ZWF1, GND1, RPE1, and TKL1 in *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.*, **71**, 339–349.

54. Allen,S.A., Clark,W., McCaffery,J.M., Cai,Z., Lanctot,A., Slininger,P.J., Liu,Z.L. and Gorsich,S.W. (2010) Furfural induces reactive oxygen species accumulation and cellular damage in *Saccharomyces cerevisiae*. *Biotechnol. Biofuels*, **3**, 1–10.

55. Hasunuma,T., Sanda,T., Yamada,R., Yoshimura,K., Ishii,J. and Kondo,A. (2011) Metabolic pathway engineering based on metabolomics confers acetic and formic acid tolerance to a recombinant xylose-fermenting strain of *Saccharomyces cerevisiae*. *Microb. Cell Fact.*, **10**, 2–13.

56. Ding,M.-Z., Wang,X., Liu,W., Cheng,J.-S., Yang,Y. and Yuan,Y.-J. (2012) Proteomic research reveals the stress response and detoxification of yeast to combined inhibitors. *PLoS One*, **7**, e43474.

57. Andrew,E.J., Merchan,S., Lawless,C., Banks,A.P., Wilkinson,D.J. and Lydall,D. (2013) Pentose phosphate pathway function affects tolerance to the G-Quadruplex binder TMPyP4. *PLoS One*, **8**, e66242.

58. González-Ramos,D., van den Broek,M., van Maris,A.J., Pronk,J.T. and Daran,J.M. (2013) Genome-scale analyses of butanol tolerance in *Saccharomyces cerevisiae* reveal an essential role of protein degradation. *Biotechnol. Biofuels*, **6**, 1754–6834.

59. Hao,X.-C., Yang,X.-S., Wan,P. and Tian,S. (2013) Comparative proteomic analysis of a new adaptive *Pichia Stipitis* strain to furfural, a lignocellulosic inhibitory compound. *Biotechnol. Biofuels*, **6**, 34.

60. Goldberg,A.L. (2003) Protein degradation and protection against misfolded or damaged proteins. *Nature*, **426**, 895–899.

61. Wang,X., Xu,H., Ha,S.-W., Ju,D. and Xie,Y. (2010) Proteasomal degradation of Rpn4 in *Saccharomyces cerevisiae* is critical for cell viability under stressed conditions. *Genetics*, **184**, 335–342.

62. Kahar,P., Taku,K. and Tanaka,S. (2011) Enhancement of xylose uptake in 2-deoxyglucose tolerant mutant of *Saccharomyces cerevisiae*. *J. Biosci. Bioeng.*, **111**, 557–563.

63. Ouyang,Z., Song,M., Güth,R., Ha,T.J., Larouche,M. and Goldowitz,D. (2011) Conserved and differential gene interactions in dynamical biological systems. *Bioinformatics*, **27**, 2851–2858.

64. Song,M., Ouyang,Z. and Liu,Z. (2009) Discrete dynamical system modelling for gene regulatory networks of 5-hydroxymethylfurfural tolerance for ethanologenic yeast. *IET Syst. Biol.*, **3**, 203–218.

65. Kim,D. and Hahn,J.-S. (2013) Roles of the Yap1 transcription factor and antioxidants in *Saccharomyces cerevisiae*'s tolerance to furfural and 5-hydroxymethylfurfural, which function as thiol-reactive electrophiles generating oxidative stress. *Appl. Environ. Microbiol.*, **79**, 5069–5077.

66. Wade,J.T., Hall,D.B. and Struhl,K. (2004) The transcription factor Ifh1 is a key regulator of yeast ribosomal protein genes. *Nature*, **432**, 1054–1058.

67. Siddiqui,A.H. and Brandriss,M.C. (1989) The *Saccharomyces cerevisiae* PUT3 activator protein associates with proline-specific upstream activation sequences. *Mol. Cell. Biol.*, **9**, 4706–4712.

68. Takagi,H. (2008) Proline as a stress protectant in yeast: physiological functions, metabolic regulations, and biotechnological applications. *Appl. Microbiol. Biotechnol.*, **81**, 211–223.

69. dos Santos,S.C., Tenreiro,S., Palma,M., Becker,J. and Sá-Correia,I. (2009) Transcriptomic profiling of the Saccharomyces cerevisiae response to quinine reveals a glucose limitation response attributable to drug-induced inhibition of glucose uptake. *Antimicrob. Agents Chemother.*, **53**, 5213–5223.

70. Furuchi,T., Ishikawa,H., Miura,N., Ishizuka,M., Kajiya,K., Kuge,S. and Naganuma,A. (2001) Two nuclear proteins, Cin5 and Ydr259c, confer resistance to cisplatin in *Saccharomyces cerevisiae*. *Mol. Pharmacol.*, **59**, 470–474.

71. Spasskaya,D., Karpov,D., Mironov,A. and Karpov,V. (2014) Transcription factor Rpn4 promotes a complex antistress response in *Saccharomyces cerevisiae* cells exposed to methyl methanesulfonate. *Mol. Biol.*, **48**, 141–149.

72. Zähringer,H., Thevelein,J.M. and Nwaka,S. (2000) Induction of neutral trehalase Nth1 by heat and osmotic stress is controlled by STRE elements and Msn2/Msn4 transcription factors: variations of PKA effect during stress and growth. *Mol. Microbiol.*, **35**, 397–406.

73. Kwast,K.E., Lai,L.-C., Menda,N., James,D.T., Aref,S. and Burke,P.V. (2002) Genomic analyses of anaerobically induced genes in Saccharomyces cerevisiae: functional roles of Rox1 and other factors in mediating the anoxic response. *J. Bacteriol.*, **184**, 250–265.

74. Zhang,B., Wang,J., Wang,X., Zhu,J., Liu,Q., Shi,Z., Chambers,M.C., Zimmerman,L.J., Shaddox,K.F., Kim,S. *et al.* (2014) Proteogenomic characterization of human colon and rectal cancer. *Nature*, **513**, 382–387.

75. Marguerat,S., Schmidt,A., Codlin,S., Chen,W., Aebersold,R. and Bähler,J. (2012) Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells. *Cell*, **151**, 671–683.

76. Edgar,R., Domrachev,M. and Lash,A.E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.*, **30**, 207–210.