OMICS A Journal of Integrative Biology
Volume 14, Number 4, 2010

DOI: 10.1089/omi.2009.0144

Inference of the Molecular Mechanism of Action from Genetic Interaction and Gene Expression Data

Mojca Mattiazzi,[1] Tomaž Curk,[2] Igor Križaj,[1] Blaž Zupan,[2,3] and Uroš Petrovič[1]

Abstract

Inference of new and useful hypotheses from heterogeneous sources of genome-scale experimental data requires new computational methods that can integrate different types of data. Gene expression and genetic interaction data are two most informative data types, each allowing the identification of genes at different levels of cellular regulatory network hierarchy. We present an integrative data analysis approach, which, rather than correlating the findings from the two data sets, uses each type of data independently to identify the components of molecular pathways and combines them into a single directed network. Our computational genomics approach is based on a set of inference rules traditionally used for reasoning on genetic experiments, which we have formalized and implemented in a software tool. The approach uses chemogenetic interaction and expression data to infer the type of relation between the chemical substance (perturber) and a transcription factor by using previous knowledge on the set of genes whose expression the transcription factor in question regulates. We have used the proposed approach to successfully infer the models for the action of the drug rapamycin and of a DNA damaging agent on their molecular targets and pathways in yeast cells. The developed method is available as a web-based tool at http://www.ailab.si/perturbagen.

Introduction

Finding the relations among diseases, genetic makeup of individuals, and the action of drugs is one of the main challenges for which functional genomics could provide answers and solutions. Gene expression and genetic/chemogenetic interaction data are two of the most informative data types in functional genomics (von Mering et al., 2002). Since the introduction of the whole genome DNA microarrays (DeRisi et al., 1997) it has become clear that every perturbation of cellular processes causes a set of genes to change their expression levels, thus enabling to determine a gene expression profile that is potentially specific for the perturbation. A nice example of this is the recently proposed concept of Connectivity Maps (Lamb et al., 2006). On the other hand, chemogenetic interactions identify genes that are functionally related to the cellular response to the perturbation, and it has been shown that chemogenomic profiles are also perturbation-specific (Hillenmeyer et al., 2008; Parsons et al., 2004).

Traditionally, combining these two types of data has been based on finding overlapping sets of genes in interaction with the perturber as determined from each type of experiment. An example of the reasoning behind this approach is that genes with upregulated expression under some experimental condition are important for the cellular adaptation to this condition, and that therefore their mutation would lead to impaired growth under the same condition. For instance, in a study to identify the molecular mechanism of action of an immunosuppressive and anticancer drug rapamycin in yeast *Saccharomyces cerevisiae*, Xie and coworkers (2005) have identified all yeast single deletion knock-out strains that exhibited either decreased or increased resistance when compared to the wild-type strain. They have then looked at the overlap between the identified set of genes and the set of genes whose expression had previously been shown to change significantly in response to rapamycin treatment (Hardwick et al., 1999). Similarly, as in other studies employing such correlation rationale (Ooi et al., 2006), only less than 10% of the genes in chemogenetic interaction with rapamycin have had significantly different expression level when cells were treated with rapamycin (Xie et al., 2005).

An explanation of this phenomenon has recently been proposed (Li and Zhan, 2008; Yeger-Lotem et al., 2009) that considers the hierarchical architecture of cellular regulatory networks and which claims that gene expression profiling and chemogenetic interactions reveal different but related sets of genes. Based on this rationale, we here present a method

[1]Department of Molecular and Biomedical Sciences, Jožef Stefan Institute, Ljubljana, Slovenia.
[2]Faculty of Computer and Information Sciences, University of Ljubljana, Ljubljana, Slovenia.
[3]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas.
The first two authors contributed equally to this work.

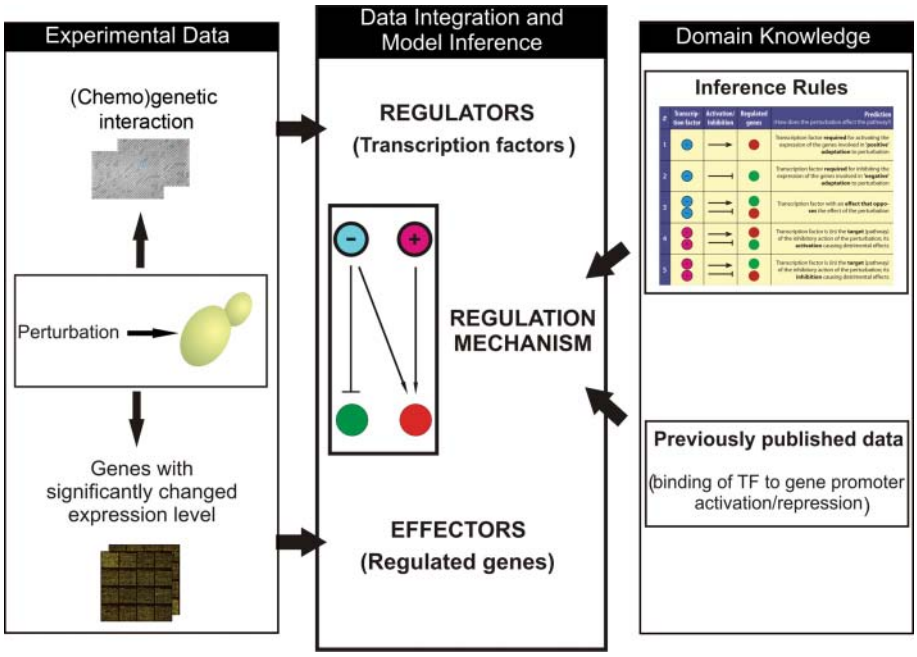


FIG. 1. Schematic representation of the method's input and output data. As input data, experimental data on the effects of the perturbation on the transcriptome and on its (chemo)genetic interactome, together with literature data on the nature of the effect of the transcription factors that are in genetic interaction, on genes with a significantly changed expression level, are used. The method generates a wiring diagram of a hypothetical model of the molecular mechanism of the action of the perturbation. See the text (the Materials and Methods section) for details.

(outlined in Fig. 1) that uses the information from experimental data on the chemogenetic interaction between the perturber and genes coding for transcription factors, gene expression data obtained by the same perturbation, and literature data on transcription factors' target genes. The output of this method is a two-layered perturbation-specific network between the regulator (coding for transcription factors) and the effector (expression regulated by the transcription factor) genes.

The proposed approach is derived from the notion that regulatory pathways in most cases have a pyramid-like structure (Yeger-Lotem et al., 2009; Yu and Gerstein, 2006), where one regulator controls the activity of several effectors (Fig. 2). Examples of this structure include pathways of transcription factors regulating the expression of their target genes, whereby an average number of genes putatively controlled by a yeast transcription factor is 37.6 (Lee et al., 2002). Obviously, in these pathways gene expression analysis will most often identify effector genes. By currently available genome-wide (chemo) genetic interaction assays, on the other hand, only mutations of single genes can be examined in the context of different perturbers, be it presence of a toxic chemical (e.g., chemogenomics) (Parsons et al., 2004) or a mutation in the (second) gene of interest (e.g., synthetic genetic array analysis) (Tong et al., 2004). Due to the buffering effects of genes functioning either in the same or parallel pathways (Hartman et al., 2001), it is much more likely that an observable or measurable phenotype will emerge as a consequence of a mutation in one single regulator than in any one of the several effectors. Therefore, current (chemo)genetic interaction assays are more likely to identify regulators, whereas gene expression assays primarily identify effectors (Fig. 2).

Our inference method is based on the set of rules that we formalized from traditional biological reasoning (Fig. 3) and that determine the type of regulator-effector relations. To demonstrate the utility of the proposed integrative data analysis method, we reanalyzed the perturber–gene interaction data and the gene expression data resulting from the action of rapamycin, an immunosuppressive and antifungal drug with a well-understood molecular mechanism of action. In yeast, as in mammals, rapamycin binds to Fpr1 protein ('FK506-binding protein' homologue) and the complex inhibits the activity of TOR (target of rapamycin) (Koltin et al., 1991), a phosphatidylinositol kinase-related protein kinase that controls cellular growth in response to nutrients (reviewed in Wullschleger et al., 2006). The major mechanism of TOR-mediated control is the inhibition of the activity of a number of transcriptional activators. Chemical epistasis and vector-based global expression analyses have revealed that transcription factors Gln3, Gat1, Rtg1/3, and Hap2/3/4/5 are inhibited by Tor1/2 proteins indirectly through Tap42 and/or Ure2 (Shamji et al., 2000), and a protein localization and biochemical study has shown that transcription factors Msn2/4 are inhibited by Tor1/2 proteins by retaining the transcription factors bound to cytoplasmic Bmh2 protein (Beck and Hall, 1999). Rapamycin thus has a pronounced effect on the expression of many genes (Shamji et al., 2000) and as such was an optimal perturber for our study. To test if the proposed approach is also useful in the cases of perturbations with less pronounced effects on transcription, we reanalyzed the data of methyl methanesulfonate (MMS), a DNA damaging agent with well-understood interactions with yeast cellular machinery (Workman et al., 2006).

Materials and Methods

Data

The extraction of relevant relations from publicly available data sets required for the proposed method, and related litera-

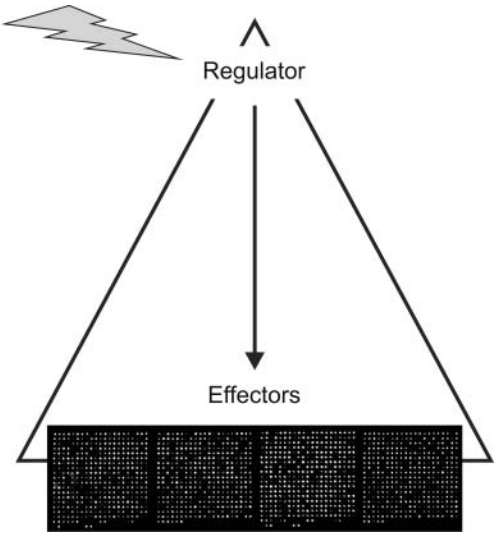


FIG. 2. Pyramid-like architecture of a typical regulator–effectors pathway (e.g., a transcription factor regulating the level of expression of several genes). (Chemo)genetic interactions with a cellular perturbation are more likely to occur with the regulator encoding gene, whereas the effectors genes are more likely to significantly change their expression levels under the conditions of a perturbation. See text for further explanation.

#	Transcription factor	Activation/ Inhibition	Regulated genes	Prediction (How does the perturbation affect the pathway?)
1	−	→	●	Transcription factor **required** for activating the expression of the genes involved in **'positive' adaptation** to perturbation
2	−	⊣	●	Transcription factor **required** for inhibiting the expression of the genes involved in **'negative' adaptation** to perturbation
3	− −	→ ⊣	● ●	Transcription factor with an **effect that opposes** the effect of the perturbation
4	+ +	→ ⊣	● ●	Transcription factor is (in) the **target** (pathway) of the inhibitory action of the perturbation; its **activation** causing detrimental effects
5	+ +	→ ⊣	● ●	Transcription factor is (in) the **target** (pathway) of the inhibitory action of the perturbation; its **inhibition** causing detrimental effects

FIG. 3. Rules for assigning predictions on the perturbation effect to each regulator–effector pair according to the input experimental data (relation of the perturber with the regulator and the effector gene, and relation between the regulator and the effector gene). Rules are based on formalization of traditional biological reasoning. Blue circles mark the regulators whose mutation causes sensitivity to the perturber (aggravating genetic interaction), whereas purple circles mark the regulators whose mutation causes resistance to the perturber (alleviating genetic interaction). Genes that are upregulated in the presence of the perturber are presented by red circles, and the ones that are downregulated by green circles. The relation between transcription factor and regulated gene can be either positive (activation, represented with an arrow) or negative (inhibition, represented with a blunt arrow). In predictive rules, "positive" adaptation refers to the adaptation of the organism by increasing the expression levels of the regulated genes, whereas "negative" adaptation refers to the adaptation of the organism by decreasing the expression levels of the regulated genes. The rules are applicable to null mutants only.

ture search, is not automated. The source of the data depends on the type of the perturber studied, and has to be identified and preprocessed according to the published format.

For the rapamycin pathway study, we determined rapamycin–gene interactions for transcription factor genes that were selected according to data from *Saccharomyces cerevisiae* rapamycin–gene whole-genome interaction data (Xie et al., 2005), and we measured whole-genome gene expression data at the most informative time point, according to a study using four different time points following treatment with rapamycin (Hardwick et al., 1999). For the inference of MMS pathway, the list of transcription factors was taken from the study of Workman and coworkers (2006), and the gene expression data was from Gasch et al. (2001).

Transcription factor–effector genes interactions were determined based on ChIP–chip experiments by Harbison et al. (2004), downloaded from SGD's FTP site: http://downloads.yeastgenome.org/chromosomal_feature/scerevisiae_regulatory.gff. For the 102 transcription factors, there were 1,918 putative effectors genes, which have at least one transcription factor binding site within 1,000 bp upstream of transcription start site. For gene interaction data we used data on 5,203 genes from BioGRID (http://www.thebiogrid.org/ downloads/datasets/SGD.tab.txt). We used gene annotation information from Gene Ontology (http://www.geneontology.org/) to infer the type of the influence of individual transcription factors. Transcription factors ($N = 26$) annotated with either of the terms "positive regulation of transcription" (GO:0045941; GO Biological Process), "transcription activator activity" (GO:0016563; GO Molecular Function), and "transcription coactivator activity" (GO:0003713; GO Molecular Function) were considered as having a positive influence on transcription. Transcription factors ($N = 40$) annotated with either of the terms "negative regulation of transcription" (GO:0016481; GO Biological Process), "transcription repressor activity" (GO:0016564; GO Molecular Function), "general transcriptional repressor activity" (GO:0016565; GO Molecular Function), "specific transcriptional repressor activity" (GO:0016566; GO Molecular Function), "transcription corepressor activity" (GO:0003714) were considered as having a negative influence on transcription.

Additional manual literature search was performed to determine the influence of three transcription factors of interest for the studies we report. Transcription factors Rtg3 and Ste12 were determined to activate expression, and transcription factor Ume6 to repress expression.

Strains and media, growth assay

The BY4741 (*MATa his3*Δ *leu2*Δ *met15*Δ *ura3*Δ) strain was used as the wild-type (WT) reference and all deletion mutants were haploid derivatives of BY4741. Standard YPD medium was used (1% yeast extract, 2% peptone, 2% glucose). Rapamycin (LC Laboratories, USA) was added to 5 nM final concentration, unless otherwise stated. Growth assay was performed as described in (St Onge et al., 2007), with some modifications. Individual deletion strains were grown overnight in YPD. Cell cultures were then diluted to an initial OD_{600} of ~ 0.05 in YPD containing 5 nM rapamycin and control YPD and grown in 300 μL volumes in 96-well plates in Tecan Sunrise microplate reader until stationary phase. The growth of each culture was monitored by measuring the OD_{595} every 2 min with constant shaking between measurements. The growth rate (K) was determined as the slope of the growth curve during exponential growth. The relative growth fitness (R) of a strain was calculated as: $R = (K_{\mathrm{rapa}\Delta}/K_{\mathrm{rapaWT}})/(K_{\Delta}/K_{\mathrm{WT}})$, where $K_{\mathrm{rapa}\Delta}$ and K_{Δ} are the growth rates of the deletion strain in rapamycin containing and control medium, respectively, and K_{rapaWT} and K_{WT} are the growth rates of the wild-type strain in rapamycin containing and control medium, respectively. The final R was calculated from three independent biological experiments performed in two parallels each.

DNA microarray experiment

Yeast cultures for the DNA microarray experiment were prepared as described by Hardwick and coworkers (1999), with some modifications. Briefly, wild-type strain or *ume1*Δ strain (strain with deleted *UME1* gene from the yeast deletion collection; see EUROSCARF at http://web.uni-frankfurt.de/fb15/mikro/euroscarf/) was inoculated in 50 mL of YPD medium and grown overnight. The culture was diluted 1:200 in YPD prewarmed to 30°C and allowed to grow to $OD_{600} \sim 1.0$. Rapamycin was added to a final concentration of 100 nM. After 90 min the cells were harvested by filtration, frozen in liquid nitrogen, and stored at −80°C until RNA isolation.

DNA microarrays were a kind gift from J.L. DeRisi. Microarray postprocessing, RNA sample preparation, cDNA synthesis, labeling, and DNA microarray hybridizations followed published procedures (DeRisi et al., 1997; http://derisilab.ucsf.edu/data/microarray/protocols.html), with some modifications. For cDNA synthesis in the presence of amino-allyl dUTP 5 μg mRNA were used. The test cDNA sample (i.e., rapamycin treatment) was labeled with Cy5 and the control sample (i.e., no treatment) with Cy3 fluorescent dye. For cleanup of cDNA synthesis and dye-coupling reactions QIAquick PCR purification kit (Qiagen, Chatsworth, CA, USA) was used. Microarray hybridizations were incubated at 65°C for 16 h. Microarrays were scanned with GenePix Personal 4100A microarray scanner (Axon Instruments, Union City, CA, USA) and analyzed using GenePix Pro 5.0 software (Axon Instruments). Cy5 and Cy3 intensities were background-corrected (median background intensities subtracted from median intensities) and normalized using a regression correlation (NOMAD; http://ucsf-nomad.sourceforge.net/). We used the sliding window method described in (Quackenbush, 2002) and in Yang et al. (2002) to calculate Z-scores based on the R-I (ratio-intensity) plot, where $R = \log_2(\mathrm{Cy5}/\mathrm{Cy3})$ is the normalized ratio of intensities and $I = 0.5 \cdot [\log_2(\mathrm{Cy5}) + \log_2(\mathrm{Cy3})]$. Two thresholds were used to select differentially expressed genes: genes with absolute Z-score above or equal to 1 were selected in the less stringent filtering procedure where the number of false negatives was minimized (true positives at 86% confidence level), and genes with absolute Z-score above or equal to 1.96 were selected in the more stringent filtering procedure (true positives at 95% confidence level). In the experiment where the transcriptomes of the wild-type and *ume1*Δ strains were compared, we used the statistical approach for replicate filtering described in Yang et al. (2002) to detect the outliers between the two experiments: if the two experiments are similar, then the Cy5/Cy3 intensity ratios of a given gene in the two experiments should be the same. If we define $R_{\mathrm{WT}}(g) = \log_2[\mathrm{Cy5}_{\mathrm{WT}}(g)/$

$Cy3_{WT}(g)]$ and $R_{ume1d}(g) = \log_2[Cy5_{ume1d}(g)/Cy3_{ume1d}(g)]$, then for each gene g, the ratio $R_{WT}(g) - R_{ume1d}(g)$ should be equal to zero. We calculated the mean and standard deviation (SD) of this difference for all genes in the two experiments. Genes, where difference $|R_{WT}(g) - R_{ume1d}(g)|$ was more than 2 SD from the mean, were considered outliers, that is, they significantly differed between the two experiments.

All DNA microarray data is available from the ArrayExpress database (accession number: E-MEXP-1782).

Inference of pathways

The proposed method infers the molecular mechanism of the action of perturbers from (chemo)genetic interaction and gene expression data and is outlined in Figure 1. As an input, the method uses two experimental data sets related to the perturber under study: (1) a list of all perturber–gene interactions, and (2) a gene expression matrix of all genes in the genome of cells treated with the perturber. It can handle generators of perturbations of any kind: a mutation, a bioactive small molecule, or an overexpressed protein, either endo- or exogenous. The method also requires the user to define a set of regulators for which the pathways will be inferred. In our experimental setup reported above, this set was composed of *S. cerevisiae* genes coding for the transcription factors from Harbison et al. (2004).

The method then proceeds in the following steps:

1. For a set of regulators, identify genes that are present in the list of perturber–gene interactions.
2. Use previously published data to determine the relationship (activation, repression) between the regulators found in Step 1 and their effector genes. Besides the results of ChIP–chip experiments for a large number of transcription factors (e.g., Harbison et al., 2004; Lee et al., 2002), also some modifications of this dataset (e.g., Beyer et al., 2006) are useful examples of such data sets, and the activating or repressing function is derived by literature search.
3. From the set of effectors identified in Step 2, keep only those whose expression is significantly altered under the perturbation. The significance of gene expression change was measured as indicated above.
4. From the set of regulators from Step 1 keep only those that potentially regulate at least one effector gene from Step 3. Using the list of regulators and effectors, construct a gene interaction network that includes the relations identified in Step 2.
5. For each regulator in the interaction network from Step 4 there are N_{up} upregulated effector genes and N_{down} downregulated effector genes. Let N be the total number of effector genes for this regulator (Step 1). Assign each regulator a score computed as $\max(N_{up}, N_{down})/N$ to assess the proportion of effector genes which changed the expression in the prevailing direction.
6. Based on: (a) the type of (chemo)genetic interaction of the perturber with the regulator [alleviating—the effect of the mutation of the regulator is less severe than expected, or aggravating—the effect of the mutation of the regulator is more extreme than expected; see also Mani et al. (2008)]; (b) the type of effect of the regulator on the effectors (activating/repressing); and (c) the prevailing direction of the change of gene expression in the group of effectors (up-/downregulated), associate each regulator with a description of its interaction with the perturber, using the corresponding rule from Figure 3.

The final model for the molecular mechanism of the action for the perturber inferred by the method is thus composed of the regulator–effectors network (Step 5), the scores associated with each of the regulators, and the predictions for all regulator–effectors pathways affected by a perturbation. By including an edge with a type of regulation as determined from Step 6 from the perturber to all of the regulator genes, the model is then visualized as a network (see Fig. 4A for an example).

Results

We first determined chemogenetic interactions between rapamycin and transcription factors encoding genes in yeast *S. cerevisiae*. We have used the data from the study by Xie and coworkers (2005), where among 373 haploid single gene mutants with altered sensitivity to rapamycin in an agar-based screening, 29 mutated genes are annotated to the "transcription regulator activity" gene ontology term. Eight of these (*GAT1*, *GLN3*, *HAP2*, *HAP3*, *HAP5*, *RTG3*, *UME1,* and *UME6*) code for transcription factors for which data is available on their binding to regulatory sequences on the genome-wide scale (Harbison et al., 2004). We experimentally determined the relative growth rates compared to the wild-type strain in the presence of rapamycin for gene deletion strains of these eight genes, using growth curve measurements in liquid medium (Table 1). In addition, we determined the same parameter for strains with deleted *HAP4*, *MSN2,* and *MSN4* genes, which have not been identified in the screening (Xie et al., 2005), but have been previously reported to be functionally related to rapamycin (Beck and Hall, 1999; Shamji et al., 2000), and for which data exists on their binding to regulatory sequences (Harbison et al., 2004). Our results confirmed chemogenetic interaction between rapamycin and 10 out of the 11 genes included in the analysis—Msn2 being the only exception, because its deletion does not significantly alter the growth rate of rapamycin-treated cells. This set of 10 genes (Table 1) and their interactions was used in consequent analysis by the proposed method.

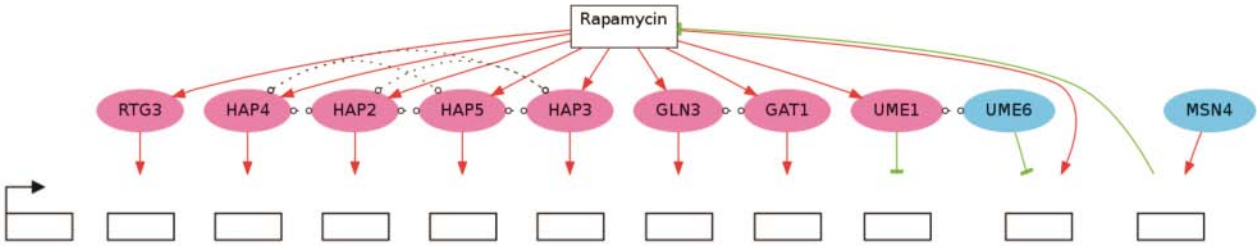
We used gene expression data from our own DNA microarray experiment following the same rapamycin treatment protocol as in Hardwick et al. (1999). To minimize the number of false negatives, we initially used a less stringent differentially expressed gene filtering procedure by selecting the genes with absolute Z-scores above or equal to 1.

The inferred network is shown in Figure 4A, and details on the functional relationships between rapamycin treatment and transcription factor–effectors pathways are presented in Figure 4B (html version available at http://www.ailab.si/perturbagen/examples/RapamycinZLs/results.html). The transcription factors are ranked according to the ratio between the number of putatively regulated genes with significantly changed expression level in the prevailing direction caused by rapamycin and the total number of genes putatively regulated by the transcription factor. Thus, the highest score ($4/7 = 0.571$) was assigned to the Rtg3 transcription factor, which putatively binds to the regulatory region of seven genes (Harbison et al., 2004), four of which were

upregulated. The predictions describing the relation between rapamycin and transcription factors were predominantly based on the formal rule *Transcription factor is (in) the target (pathway) of the inhibitory action of the perturbation; its activation causing detrimental effects*: this rule was assigned to Gln3, Gat1, Hap2, Hap3, Hap4, Hap5, Rtg3, and Ume1 transcription factors, whose activation, according to the prediction, in the presence of rapamycin harms the cells. The prediction for Ume6 was *Regulator with an effect that opposes the effect of rapamycin*, and for Msn4 the method's prediction is that this *Regulator is required for activating the expression of the genes involved in positive adaptation to perturber*.

The functional interaction between Ume1 and rapamycin was inferred based on a single gene with significantly altered

A



B

#	TF	Rapamycin - gene interaction	Expression of target genes	# genes data set	# genes all	Score	Activator/ Repressor	Prediction	Target genes and measure of their differential expression [Z-score]
1	RTG3	+	up	4	7	0.571	A	4	up (4 genes): ARG3 (1.6), COX9 (1.4), IDH1 (1.4), IDP1 (1.1)
	Transcription factor RTG3 is (in) the **target** (pathway) of the inhibitory action of Rapamycin; its **activation** causing detrimental effects.								
2	HAP4	+	up	17	34	0.5	A	4	up (17 genes): ATP16 (1.0), ATP17 (1.4), ATP20 (1.5), ATP7 (1.0), COR1 (1.4), COX12 (1.5), COX13 (1.0), COX5A (1.1), COX6 (1.2), COX7 (1.3), COX8 (1.2), COX9 (1.4), IDP1 (1.1), INH1 (1.5), POR1 (2.0), QCR7 (1.0), RIP1 (1.0)
	Transcription factor HAP4 is (in) the **target** (pathway) of the inhibitory action of Rapamycin; its **activation** causing detrimental effects.								
3	GLN3	+	up	19	40	0.475	A	4	up (19 genes): ARG4 (1.1), BAT2 (1.3), CPS1 (2.5), DAL1 (1.3), DAL2 (3.1), DAL3 (4.6), DAL5 (4.7), DAL7 (2.7), DIP5 (2.2), DUR3 (3.5), GDH2 (1.6), GLT1 (1.1), MCH4 (2.0), MIC14 (1.6), RAD28 (1.1), RRI2 (1.0), UGA3 (1.0), YHI9 (2.2), YMR118C (1.2)
	Transcription factor GLN3 is (in) the **target** (pathway) of the inhibitory action of Rapamycin; its **activation** causing detrimental effects.								
4	HAP2	+	up	8	23	0.348	A	4	up (8 genes): ATP17 (1.4), COX4 (1.1), COX6 (1.2), GLT1 (1.1), MEP1 (2.6), QCR7 (1.0), UGA1 (1.7), UGA3 (1.0)
	Transcription factor HAP2 is (in) the **target** (pathway) of the inhibitory action of Rapamycin; its **activation** causing detrimental effects.								
5	UME1	+	down	1	3	0.333	R	4	down (1 gene): RPS11B (-4.9)
	Transcription factor UME1 is (in) the **target** (pathway) of the inhibitory action of Rapamycin; its **activation** causing detrimental effects.								
6	HAP5	+	up	6	20	0.3	A	4	up (6 genes): COX4 (1.1), COX6 (1.2), QCR7 (1.0), SED1 (1.3), STE2 (1.2), STP2 (1.5)
	Transcription factor HAP5 is (in) the **target** (pathway) of the inhibitory action of Rapamycin; its **activation** causing detrimental effects.								
7	HAP3	+	up	4	14	0.286	A	4	up (4 genes): COX4 (1.1), COX6 (1.2), GLT1 (1.1), UGA3 (1.0)
	Transcription factor HAP3 is (in) the **target** (pathway) of the inhibitory action of Rapamycin; its **activation** causing detrimental effects.								
8	MSN4	-	up	6	29	0.207	A	1	up (6 genes): CUE4 (1.1), HSP26 (1.3), HSP78 (3.0), MDH1 (1.2), MMF1 (1.0), YGR050C (1.9)
	Transcription factor MSN4 **required** for activating the expression of the genes involved in **'positive' adaptation** to Rapamycin.								
9	GAT1	+	up	2	11	0.182	A	4	up (2 genes): DAL2 (3.1), YJR011C (1.1)
	Transcription factor GAT1 is (in) the **target** (pathway) of the inhibitory action of Rapamycin; its **activation** causing detrimental effects.								
10	UME6	-	up	17	105	0.162	R	3	up (17 genes): ARO9 (1.4), CAT2 (3.8), CRC1 (3.2), CSR2 (1.6), FYV5 (1.7), GDH2 (1.6), KNS1 (1.2), PST1 (1.0), RIP1 (1.0), SIP4 (2.4), SPL2 (1.6), UGA4 (4.8), YCR061W (1.3), YHR138C (1.8), YLR297W (1.1), YOL131W (1.0), YPL017C (1.5)
	Transcription factor UME6 with an **effect that opposes** the effect of Rapamycin.								

C

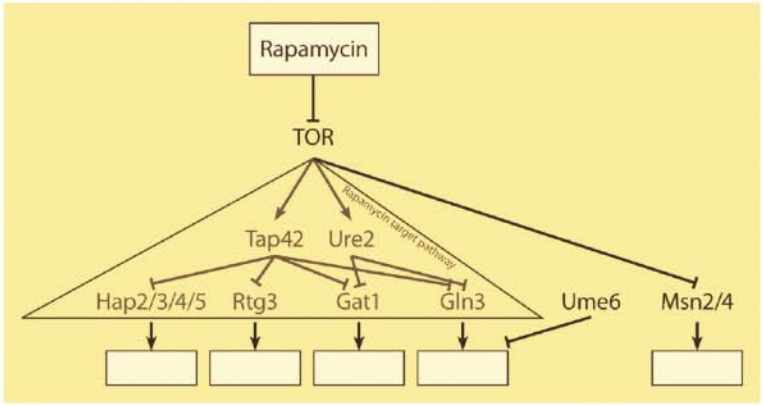


Table 1. Relative Growth Fitness of Selected Transcription Factor Mutant Strains, Compared to the Wild-Type Strain

Mutated gene	*Relative growth rate (±SD)*	*Type of interaction (aggravating, alleviating)*
GLN3	2.79 (±0.09)	Alleviating
HAP4	1.63 (±0.08)	Alleviating
UME6	0.40 (±0.11)	Aggravating
RTG3	1.42 (±0.10)	Alleviating
UME1	1.30 (±0.01)	Alleviating
HAP2	1.16 (±0.01)	Alleviating
MSN4	0.87 (±0.00)	Aggravating
HAP3	1.22 (±0.12)	Alleviating
HAP5	1.23 (±0.14)	Alleviating
GAT1	1.09 (±0.07)	Alleviating
MSN2	0.98 (±0.05)	None

Names of the genes constituting the set that represents genetic interaction input data for the method are marked in bold.

expression level, out from total of three genes putatively regulated by Ume1 according to Harbison et al. (2004). Because Ume1 has been identified as a component of the Rpd3 repression complex (Keogh et al., 2005; Kurdistani et al., 2002), three regulated genes are almost cetainly an underestimate. This fact thus gave us an opportunity to test the developed method on an example with incomplete data. To test whether the predicted interaction indeed exists, we performed whole-genome gene expression experiments comparing the wild-type and *ume1*Δ strains under rapamycin treatment and control conditions. The genes with different expression behaviour (i.e., the outliers, as described in the Materials and Methods section) were significantly enriched for the *maturation of SSU–rRNA from tricistronic rRNA transcript* biological process ($p = 10^{-5}$) and *snoRNA binding* molecular function ($p = 1.34 \times 10^{-5}$) Gene Ontology terms. The respective 18 genes were all downregulated in the presence of rapamycin in the wild-type strain, but did not significantly change their expression in the *ume1*Δ strain (Table S1 at http://www.ailab.si/supp/perturbagen/).

To verify the predicted relations between rapamycin and the transcription factors in chemogenetic interaction with it, we used gene expression data from an independent study (Hardwick et al., 1999) and employed the same fourfold change threshold for differential expression as proposed by the authors of the study. The predicted relations were exactly the same, with the only difference in the case of *UME1*, where the prediction [*Transcription factor UME1 is (in) the target (pathway) of the inhibitory action of rapamycin; its inhibition causing detrimental effects*] was again based on a single differentially expressed gene (Fig. S1 at http://www.ailab.si/supp/perturbagen/).

To test how different threshold levels for differential expression determination affect the results obtained with the method, we set a more stringent threshold and selected the genes with absolute *Z*-score above or equal to 1.96, to minimize the number of false positives. The results (Fig. S2 at http://www.ailab.si/supp/perturbagen/) somewhat differ from the ones obtained using less stringent filtering—the role of Rtg3 and Hap3 is not predicted, and the proposed rules are different for Hap5 and Msn4—but, importantly, in this case for six out of the eight transcription factors (Gat1, Hap2, Hap4, Hap5, Msn4, and Ume1), the rule is predicted on the basis of a single effector gene.

To test the proposed method on a perturber with less pronounced transcription-related effects, we analyzed the data on genome-wide effects of a DNA-damaging substance MMS. We compared the results obtained by our method to those obtained in a comprehensive study that mapped the transcriptional network controlling the DNA damage response by measuring genome-wide binding locations for 30 DNA damage-related transcription factors after exposure of yeast cells to MMS, generating global hypotheses of how

FIG. 4. Comparison of the predicted model of the molecular mechanism of rapamycin activity in yeast cells generated using our method with the current model (Wullschleger et al., 2006). (**A**) The method predicted Rtg3, Gln3, Gat1, and Hap2/3/4/5 as transcription factors in the target pathway of rapamycin, which are activated by the perturber; Ume1 as a regulator in the target pathway of the inhibitory action of rapamycin and its activation causing detrimental effects; Ume6 as a transcription factor with an effect that opposes the effect of rapamycin, and Msn4 as a regulator that is required for activating the expression of the genes involved in a positive adaptation to rapamycin. Purple genes denote alleviating chemogenetic interaction, and blue ones aggravating chemogenetic interaction with rapamycin. Red arrows denote activation and green blunt arrows inhibition. Dotted lines mark physical interaction between transcription factors. (**B**) Snapshot of the results table of transcription factors evaluated by the method and ranked by the assigned score. Columns show (from left to right): name of the transcription factor encoding gene; type of interaction between perturber and the gene ("+"—alleviating interaction, "−"—aggravating interaction); orientation of the change in expression of transcription factor target genes under conditions of the perturber activity; the number of genes in the database to whose promoter sequences the transcription factor binds and that have significantly changed expression level under conditions of the perturber activity; the number of all genes in the database to whose promoter sequences the transcription factor binds; the arbitrary ranking score (calculated as the ratio of numbers from the former two columns); nature of the transcription factor (A—transcription activator; R—transcription repressor); the method's prediction of the type of interaction between the perturber and the transcription factor. (**C**) The current model proposes Rtg3, Gln3, Gat1, and Hap2/3/4/5 as being activated by rapamycin through canonical pathway (i.e., indirectly, through its inhibition of TOR, which normally inhibits the activity of these seven transcription factors through Tap42 and Ume2 proteins) (Wullschleger et al., 2006), Msn2/4 as being indirectly activated by rapamycin, which causes their translocation into nucleus (Shamji et al., 2000), and Ume6 as a transcription factor with an effect that opposes the effect of rapamycin by having an opposite effect on gene expression than Gln3 (Blinder et al., 1996). The proposed method builds from data on transcription factors binding to their target promoter sequences and therefore does not include other types of regulators, such as protein kinases (e.g., TOR proteins). (**A**) and (**B**) are snapshots of the Web-based tool results page accessible at http://www.ailab.si/perturbagen/examples/RapamycinZLs/results.html.

cellular signalling and transcription are integrated after DNA damage (Workman et al., 2006). Genes encoding the 30 transcription factors were used as the input data for chemogenetic interactions for our method, and gene expression data were taken from Gasch et al. (2001). Genes regulated by 10 of the 30 transcription factors, including all four implicated as MMS-connected by multiple criteria according to Workman et al. (2006), had significantly changed expression levels, enabling prediction of the interactions of these 10 transcription factors with MMS. Six of them (Dal81, Rtg3, Rpn4, Yap1, Gcn4, and Swi6) were predicted to be required for activating the expression of genes involved in (positive) adaptation to MMS, and four of them (Swi5, Fkh2, Swi4, and Sok2) were predicted to have an effect that opposes the effect of the perturbation (Fig. S3 at http://www.ailab.si/supp/perturbagen/).

Discussion

With traditional approaches, many years of experimentation are usually required to complete the model of action of a given perturber on cells. Provided the availability of the appropriate data, the method presented here infers a network of directed and annotated functional interactions between a perturber and transcription factors combining genome-wide (chemo)genetic interaction and gene expression data. In the test case of the reconstruction of the model of activity of rapamycin on its molecular targets in yeast cells, 8 out of 10 transcription factors, Gln3, Gat1, Hap2, Hap3, Hap4, Hap5, Rtg3, and Ume1 were assigned the formal rule *Transcription factor is (in) the target (pathway) of the inhibitory action of the perturbation; its activation causing detrimental effects*. With the exception of Ume1 (see below), these transcription factors are known to be under the negative control of TOR and thus in the rapamycin target pathway (Fig. 4C), and are as such indirectly activated by rapamycin treatment (Shamji et al., 2000). Rapamycin treatment induces expression of the genes they regulate, which is the main mechanism of its toxicity in yeast (Blinder et al., 1996; Shamji et al., 2000)—the method's prediction is thus in complete accordance with the accepted mechanism of action. Ume6, on the other hand, has an opposite effect on gene expression, as a function of nitrogen availability, compared to Gln3 (Dubois and Messenguy, 1997), and in accordance with this the method correctly predicted that Ume6 is a *Regulator with an effect that opposes the effect of rapamycin*. For Msn4, the method's prediction is that this *Regulator is required for activating the expression of the genes involved in positive adaptation to perturber*, which is in accordance with the function of this protein, which activates the expression of genes building a general stress response in yeast (Gasch et al., 2000). More specifically, rapamycin promotes nuclear localization of Msn4 through a different pathway as in the case of Gln3 and Gat1 transcription activators (Beck and Hall, 1999), and the orientation of chemogenetic interaction is opposite: whereas *gln3* and *gat1* mutants are more resistant to rapamycin treatment than the wild-type strain, *msn4* mutant is more sensitive, revealing that its activity protects the cells from detrimental effects of rapamycin. Because Msn2 and Msn4 have overlapping functions, it is reasonable to assume that a buffering effect exists between the two genes also in the case of the response to rapamycin, which could explain why the deletion of only one of these two genes, *MSN4*, resulted in a measurable phenotype. Alternatively, this result could indicate an *MSN4*-specific function in the cellular response to rapamycin. The currently accepted model of the pathways affected by rapamycin was thus successfully reproduced, demonstrating that the proposed network inference approach can provide biologically meaningful results.

The functional interaction between Ume1 and rapamycin was inferred based on a single gene with significantly altered expression level, providing an opportunity to test the developed method on an example with incomplete data. Gene expression experiment comparing the wild-type and *ume1*Δ

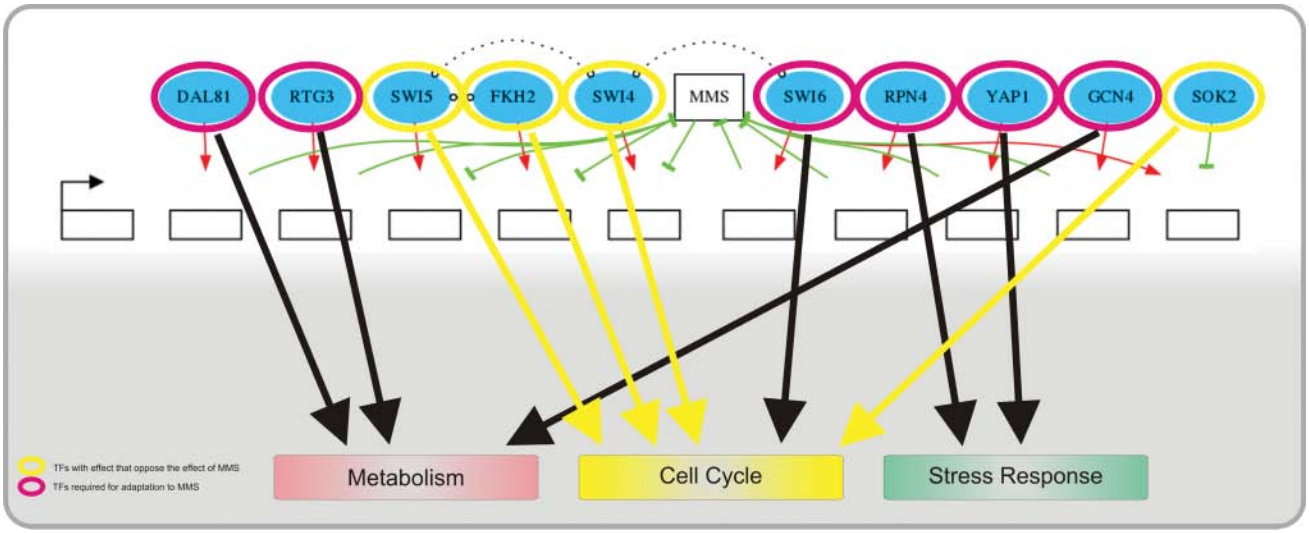


FIG. 5. Comparison of the prediction of MMS–transcription factor relations as predicted by our method with the model obtained in the study of Workman and coworkers (2006). The method's prediction is in absolute agreement with the model, and it takes the level of understanding a step further by predicting the causal relationship between the transcription factors and MMS. Transcription factors predicted to oppose the effect of MMS (circled in yellow) function in cell cycle, whereas the ones predicted to be required for adaptation to MMS (circled in purple) function in stress response and metabolism.

strains under rapamycin treatment and control conditions were performed and 279 genes were identified, which had significantly different expression pattern in the *ume1*Δ strain compared to the wild-type strains. These genes are enriched for the *maturation of SSU–rRNA from tricistronic rRNA transcript* and *snoRNA binding* Gene Ontology terms, and all the respective genes were downregulated in the presence of rapamycin in the wild-type strain, but not in the *ume1*Δ strain. Because Ume1 is a negative regulator of gene expression (Keogh et al., 2005; Kurdistani et al., 2002; Mallory and Strich, 2003; Strich et al., 1989), this is exactly the group where differences between *ume1*Δ strain and the wild-type strain were expected, and expression of these genes is apparently repressed by Ume1 when cells are treated with rapamycin. These results are in line with the known function of Ume1, the transcription factor best known as a negative regulator of the expression of a subset of meiotic genes during vegetative growth, being a component of Rpd3 repression complex which is involved in the inactivation of yeast ribosomal DNA genes (Sandmeier et al., 2002). These experimental data confirmed the prediction of the method for Ume1 that *Transcription factor is (in) the target (pathway) of the inhibitory action of the perturbation; its activation causing detrimental effects*. When using data from an independent study (Hardwick et al., 1999) as gene expression input data, the predicted relations for nine of the transcription factors were exactly the same as with our own experimental data, with the only difference in the case of *UME1*, where the prediction [*Transcription factor UME1 is (in) the target (pathway) of the inhibitory action of rapamycin; its inhibition causing detrimental effects*] was again based on a single differentially expressed gene. However, as described above, input data on genes regulated by Ume1 is incomplete, and our own as well as published results clearly demonstrate that rapamycin activates, and not inhibits, Ume1 activity. This discrepancy nicely exemplifies how complete data sets—in this case one linking transcription factors to their target genes—will enhance the performance of methods such as the one developed in this study.

In line with this, when a stringent threshold was used to select for differentially expressed genes, resulting in the majority of predictions for rapamycin transcription factor relation being based on a single effector gene, the predictions were less accurate (Fig. S2). We therefore suggest using a less stringent threshold in the proposed method, because too stringent filtering results in a larger number of false negative results, leading to inaccurate predictions. On the other hand, the negative effect of a higher number of false positives is diminished due to the fact that in addition to gene expression data (chemo)genetic interaction data are used to generate predictions, and in this way the noise coming from relatively high false positive rate is filtered out, as demonstrated by the accuracy of prediction of the mechanism of action of rapamycin (Fig. 4).The method was also used to analyze the relation between MMS and 10 transcription factors implicated in the cellular response to DNA damage caused by MMS. Dal81, Rtg3, Rpn4, Yap1, Gcn4, and Swi6 were predicted to be required for activating the expression of genes involved in adaptation to MMS, and Swi5, Fkh2, Swi4, and Sok2 were predicted to have an effect that opposes the effect of MMS. As expected, because cell cycle is the process most directly affected by DNA damage, all four genes from the latter group of genes with opposite effect to MMS, and only one from the former (Swi6), are cell cycle regulators. The remaining five transcription factors, which were all predicted to be required for adaptation to MMS, are involved in either stress response or metabolism—processes that are indeed involved in the adaptation to cellular damage (Fig. 5) (Workman et al., 2006). Even though MMS does not directly affect transcription in a way that rapamycin does, Workman and coworkers (2006) concluded that differential expression of genes does not significantly overlap with the set of genes required for resistance to MMS, but that the transcription factors most essential for cellular recovery are also the most central to the MMS transcriptional response. Our results are in absolute agreement with this, and they take the level of understanding a step further by predicting the causal relationship between the transcription factors and the perturber—MMS.

The proposed integrated data analysis technique relies on the hierarchical structure of relationships and exploits the prevailing direction of change of gene expression to infer the rules that predict the effect of perturbation on the observed pathway. The approach requires the availability of the data that relates a perturber to the set of regulators and their relation to the effector genes. The availability of such data depends on the type of the perturber under study, and in the lack of databases that would systematically store this type of data sets the collection of the data set and their preprocessing could not be fully automated at this stage. Our study showed that the method is applicable, whereas the current implementation could be further extended with automated data preprocessing steps once the comprehensive repositories with standardized perturbation experiments become available. With the development of novel techniques (e.g., Badis et al., 2008), the method could in the future be applicable also to other organisms.

Conclusions

We developed an integrative computational method that infers a network of directed and annotated functional interactions between a perturber and transcription factors. Our approach is based on experimental gene expression and (chemo)genetic interaction input data and the set of predefined inference rules. As an example, the method was used to reproduce the model of molecular mechanism of action of rapamycin. We demonstrated that successful reconstruction of the model of activity of rapamycin on its molecular targets in yeast cells, which otherwise took many years of both genome-wide and focused experimentation, is attainable by combining only chemogenetic interaction and gene expression data and cannot be inferred from each of the two sources alone. The proposed method could thus be generally useful to infer the molecular mechanisms of the action of molecules with potential biomedical or biotechnological importance in a rapid and straightforward manner: two relatively simple experimental approaches in a well-annotated model organism generate sufficient data for the method to formulate a comprehensive and reliable hypothesis for the molecular mechanism of the molecule's biological activity, provided that the molecule has sufficient effect on gene expression and transcription factors. This development brings functional genomics approaches a step closer toward becoming a technology with high predictive potential.

Acknowledgments

We thank Sepp D. Kohlwein for fruitful discussions and Roger H. Pain for critically reading the manuscript. We also thank the anonymous reviewers for comments and suggestions. Part of the experimental work was done at the Center for Functional Genomics and Bio-Chips in Ljubljana, Slovenia. This work has been supported by grants J2-9699 and L2-1112 from the Slovenian Research Agency.

Author Disclosure Statement

The authors declare that no conflicting financial interests exist.

References

Badis, G., Chan, E.T., van Bakel, H., Pena-Castillo, L., Tillo, D., Tsui, K., et al. (2008). A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. Mol Cell 32, 878–887.

Beck, T., and Hall, M.N. (1999). The TOR signalling pathway controls nuclear localization of nutrient-regulated transcription factors. Nature 402, 689–692.

Beyer, A., Workman, C., Hollunder, J., Radke, D., Moller, U., Wilhelm, T., et al. (2006). Integrated assessment and prediction of transcription factor binding. PLoS Comput Biol 2, e70.

Blinder, D., Coschigano, P.W., and Magasanik, B. (1996). Interaction of the GATA factor Gln3p with the nitrogen regulator Ure2p in *Saccharomyces cerevisiae*. J Bacteriol 178, 4734–4736.

DeRisi, J.L., Iyer, V.R., and Brown, P.O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 278, 680–686.

Dubois, E., and Messenguy, F. (1997). Integration of the multiple controls regulating the expression of the arginase gene CAR1 of *Saccharomyces cerevisiae* in response to differentnitrogen signals: role of Gln3p, ArgRp–Mcm1p, and Ume6p. Mol Gen Genet 253, 568–580.

Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., et al. (2000). Genomic expression programs in the response of yeast cells to environmental changes. Mol Biol Cell 11, 4241–4257.

Gasch, A.P., Huang, M., Metzner, S., Botstein, D., Elledge, S.J. and Brown, P.O. (2001). Genomic expression responses to DNA-damaging agents and the regulatory role of the yeast ATR homolog Mec1p. Mol Biol Cell 12, 2987–3003.

Harbison, C.T., Gordon, D.B., Lee, T.I., Rinaldi, N.J., Macisaac, K.D., Danford, T.W., et al. (2004). Transcriptional regulatory code of a eukaryotic genome. Nature 431, 99–104.

Hardwick, J.S., Kuruvilla, F.G., Tong, J.K., Shamji, A.F. and Schreiber, S.L. (1999). Rapamycin-modulated transcription defines the subset of nutrient-sensitive signaling pathways directly controlled by the Tor proteins. Proc Natl Acad Sci USA 96, 14866–14870.

Hartman, J.L.t., Garvik, B., and Hartwell, L. (2001). Principles for the buffering of genetic variation. Science 291, 1001–1004.

Hillenmeyer, M.E., Fung, E., Wildenhain, J., Pierce, S.E., Hoon, S., Lee, W., et al. (2008). The chemical genomic portrait of yeast: uncovering a phenotype for all genes. Science 320, 362–365.

Keogh, M.C., Kurdistani, S.K., Morris, S.A., Ahn, S.H., Podolny, V., Collins, S.R., et al. (2005). Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. Cell 123, 593–605.

Koltin, Y., Faucette, L., Bergsma, D.J., Levy, M.A., Cafferkey, R., Koser, P.L., et al. (1991). Rapamycin sensitivity in *Saccharomyces cerevisiae* is mediated by a peptidyl-prolyl *cis–trans* isomerase related to human FK506-binding protein. Mol Cell Biol 11, 1718–1723.

Kurdistani, S.K., Robyr, D., Tavazoie, S., and Grunstein, M. (2002). Genome-wide binding map of the histone deacetylase Rpd3 in yeast. Nat Genet 31, 248–254.

Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., Wrobel, M.J., et al. (2006). The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science 313, 1929–1935.

Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., et al. (2002). Transcriptional regulatory networks in *Saccharomyces cerevisiae*. Science 298, 799–804.

Li, H., and Zhan, M. (2008). Unraveling transcriptional regulatory programs by integrative analysis of microarray and transcription factor binding data. Bioinformatics 24, 1874–1880.

Mallory, M.J., and Strich, R. (2003). Ume1p represses meiotic gene transcription in *Saccharomyces cerevisiae* through interaction with the histone deacetylase Rpd3p. J Biol Chem 278, 44727–44734.

Mani, R., St Onge, R.P., Hartman, J.L.t., Giaever, G., and Roth, F.P. (2008). Defining genetic interaction. Proc Natl Acad Sci USA 105, 3461–3466.

Ooi, S.L., Pan, X., Peyser, B.D., Ye, P., Meluh, P.B., Yuan, D.S., et al. (2006). Global synthetic-lethality analysis and yeast functional profiling. Trends Genet 22, 56–63.

Parsons, A.B., Brost, R.L., Ding, H., Li, Z., Zhang, C., Sheikh, B., et al. (2004). Integration of chemical–genetic and genetic interaction data links bioactive compounds to cellular target pathways. Nat Biotechnol 22, 62–69.

Quackenbush, J. (2002). Microarray data normalization and transformation. Nat Genet 32(Suppl), 496–501.

Sandmeier, J.J., French, S., Osheim, Y., Cheung, W.L., Gallo, C.M., Beyer, A.L., et al. (2002). RPD3 is required for the inactivation of yeast ribosomal DNA genes in stationary phase. EMBO J 21, 4959–4968.

Shamji, A.F., Kuruvilla, F.G., and Schreiber, S.L. (2000). Partitioning the transcriptional program induced by rapamycin among the effectors of the Tor proteins. Curr Biol 10, 1574–1581.

St Onge, R.P., Mani, R., Oh, J., Proctor, M., Fung, E., Davis, R.W., et al. (2007). Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. Nat Genet 39, 199–206.

Strich, R., Slater, M.R., and Esposito, R.E. (1989). Identification of negative regulatory genes that govern the expression of early meiotic genes in yeast. Proc Natl Acad Sci USA 86, 10018–10022.

Tong, A.H., Lesage, G., Bader, G.D., Ding, H., Xu, H., Xin, X., et al. (2004). Global mapping of the yeast genetic interaction network. Science 303, 808–813.

von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S.G., Fields, S., et al. (2002). Comparative assessment of large-scale data sets of protein–protein interactions. Nature 417, 399–403.

Workman, C.T., Mak, H.C., McCuine, S., Tagne, J.B., Agarwal, M., Ozier, O., et al. (2006). A systems approach to mapping DNA damage response pathways. Science 312, 1054–1059.

Wullschleger, S., Loewith, R., and Hall, M.N. (2006). TOR signaling in growth and metabolism. Cell 124, 471–484.

Xie, M.W., Jin, F., Hwang, H., Hwang, S., Anand, V., Duncan, M.C., et al. (2005). Insights into TOR function and rapamycin

response: chemical genomic profiling by using a high-density cell array method. Proc Natl Acad Sci USA 102, 7215–7220.

Yang, I.V., Chen, E., Hasseman, J.P., Liang, W., Frank, B.C., Wang, S., et al. (2002). Within the fold: assessing differential expression measures and reproducibility in microarray assays. Genome Biol 3, research0062.

Yeger-Lotem, E., Riva, L., Su, L.J., Gitler, A.D., Cashikar, A.G., King, O.D., et al. (2009). Bridging high-throughput genetic and transcriptional data reveals cellular responses to alpha-synuclein toxicity. Nat Genet 41, 316–323.

Yu, H., and Gerstein, M. (2006). Genomic analysis of the hierarchical structure of regulatory networks. Proc Natl Acad Sci USA 103, 14724–14731.

Address correspondence to:
Dr. Uroš Petrovič
Department of Molecular and Biomedical Sciences
Jožef Stefan Institute
Jamova 39, SI-1000 Ljubljana, Slovenia

E-mail: uros.petrovic@ijs.si