

Damage Recovery Pathways in *Saccharomyces cerevisiae* Revealed by Genomic Phenotyping and Interactome Mapping

Thomas J. Begley,¹ Ari S. Rosenbach,¹ Trey Ideker,^{2,3} and Leona D. Samson^{1,4}

¹Biological Engineering Division and Center for Environmental Health Sciences and ²Whitehead Institute for Biomedical Research, Massachusetts Institute of Technology, Cambridge, MA

Abstract

We have generated a genomic phenotyping database identifying hundreds of *Saccharomyces cerevisiae* genes important for viable cellular recovery after mutagen exposure. Systematic phenotyping of 1615 gene deletion strains produced distinctive signatures for each of four mutagens. Integration of the phenotyping database with mutagen-induced transcriptional profiling data demonstrated that being transcriptionally responsive to a mutagen does not predict whether or not a gene contributes to recovery from exposure to that mutagen. Computational integration of the database with 4025 interacting proteins, comprising the yeast interactome, identified several multiprotein networks important for damage recovery. Some networks were associated with DNA metabolism and cell cycle control functions, but most were associated with unexpected functions such as cytoskeleton remodeling, chromatin remodeling, protein, RNA, and lipid metabolism. Hence, a plethora of responses other than the DNA damage response is important for recovery. These network mapping results demonstrate how systematic phenotypic assays may be linked directly to underlying molecular mechanisms.

Introduction

The accumulation of somatic cells harboring permanent genetic change (*i.e.*, mutations) contributes to the onset of cancer, aging, and other degenerative diseases (1). Because most mutagens are also cytotoxic, such accumulation must be also influenced by the ability of cells to remain viable after mutagen exposure. Much attention has been given to the role of the DNA damage response for cellular recovery, a response that coordinates DNA repair, cell cycle checkpoints and, in the case of metazoan cells, programmed cell death (2). Although

these pathways are clearly important for recovery, recent global transcriptional profiling of mutagen-exposed *Saccharomyces cerevisiae* plus computational analysis of the data led to the finding that protein degradation pathways may also contribute to recovery from mutagen exposure (3–6). Indeed, thinking of commonly studied mutagens as simply DNA-damaging agents is perhaps misleading, because most mutagens also damage proteins, RNA, carbohydrates, lipids, and other cellular molecules.

Here we have systematically tested a set of 1615 haploid *S. cerevisiae* deletion strains to identify those deletion strains that display altered viable recovery after exposure to four classical DNA damaging agents, chosen because they are known to induce different kinds of DNA lesions (1). This approach has been termed genomic phenotyping because it systematically interrogates the genome of an organism to identify individual gene products that affect a particular phenotype, in this case the sensitive or resistant phenotype of *S. cerevisiae* strains on exposure to mutagens that produce alkylation damage, oxidative damage, or radiation damage. We have analyzed the contribution of one third of the yeast genome (1,615 *S. cerevisiae* gene products) to recovery after mutagen exposure in the framework of 12,232 protein-protein and protein-DNA interactions comprising the known yeast interactome [derived from both low- and high-throughput protein-protein and protein-DNA interaction studies (7–13)]. Our results reveal that many unexpected and hitherto uncharacterized pathways influence the recovery of eukaryotic cells after damaging agent exposure.

Results and Discussion

Genomic Phenotyping

We screened a library of 1615 *S. cerevisiae* gene deletion strains in a multidose, multireplicate phenotypic screen (Fig. 1A) to determine which proteins influence recovery after attack by the methylating agent methyl methanesulfonate (MMS), the bulky alkylating agent 4-nitroquinoline-*N*-oxide (4NQO), the oxidizing agent *tert*-butyl hydroperoxide (*t*-BuOOH), and 254 nm UV radiation. These four agents were chosen because they produce very different kinds of damage. The first third of the *S. cerevisiae* gene deletion library was used to establish this labor-intensive rigorous approach, to develop database tools, and to determine whether partial coverage of the genome (as

Received 10/29/02; accepted 11/04/02.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

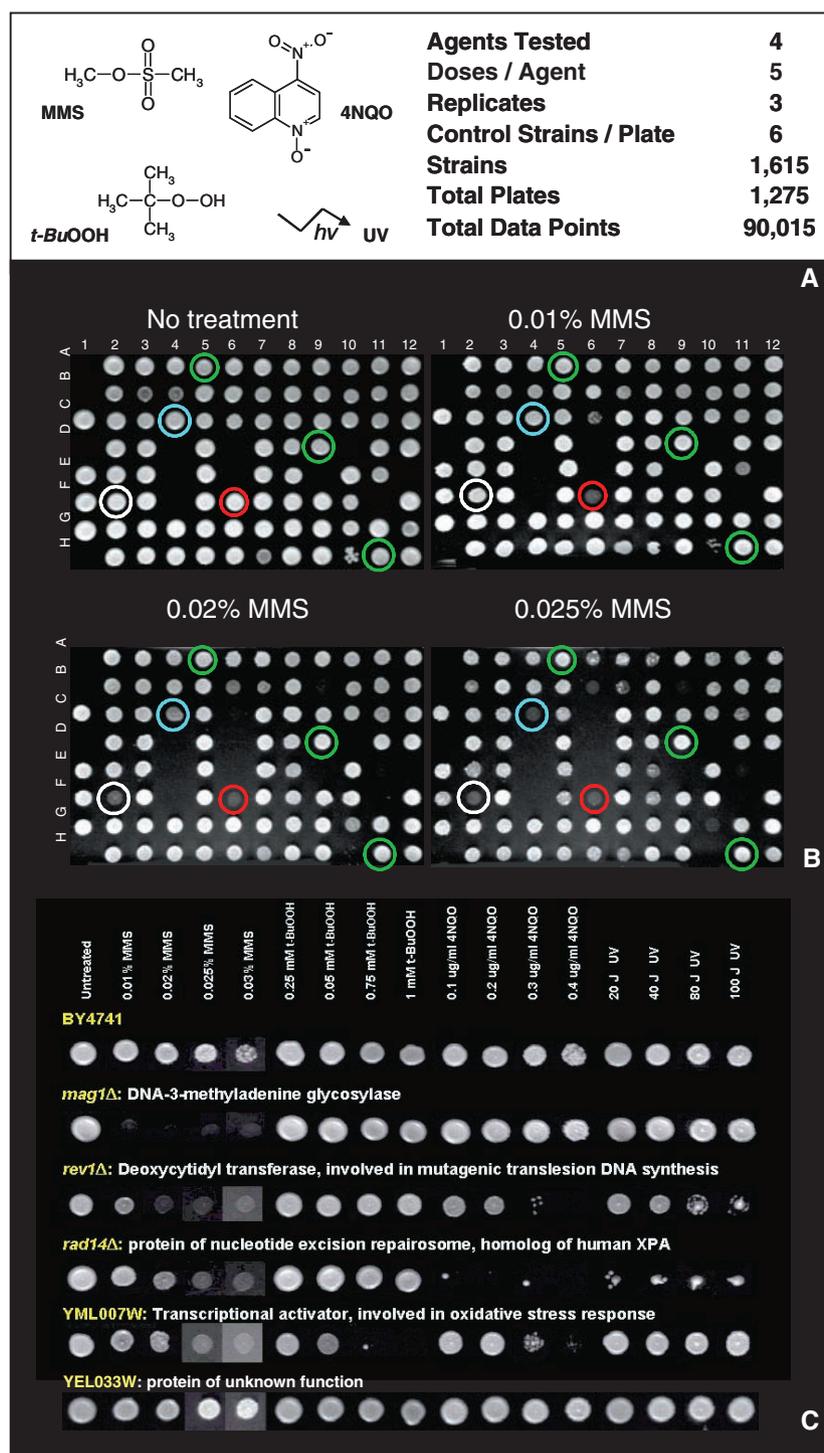
Grant support: NIH Grants RO1-CA-55042 and P30-ES02109; NIH Training Grant ES07155 and National Research Service Award F32-ES11733 (to T.J.B.).

Requests for reprints: Leona D. Samson, Biological Engineering Division and Center for Environmental Health Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139. Phone: (617) 258-7813; Fax: (617) 253-8099. E-mail: lsamson@mit.edu

Copyright © 2002 American Association for Cancer Research.

³A Baltimore Whitehead Fellow.

⁴Ellison American Cancer Society Research Professor.



might be expected for forthcoming mutant mammalian cell libraries) produces informative data. Strains were robotically spotted from saturated cultures grown in 96-well format onto agar plates with and without chemical mutagen in the agar, or with and without subsequent UV exposure. Plates were spotted with \sim 80 deletion strains, plus 6 control strains, namely the BY4741 parent (in triplicate), its MMS-sensitive *mag1Δ*

derivative, and its MMS-, 4NQO-, and UV-sensitive *rev1Δ* and *rad14Δ* derivatives (14–16). Although a *t*-BuOOH-sensitive control strain was not included among the controls, numerous *t*-BuOOH-sensitive strains were detected in the screen (see below). Following a 60-h incubation, recovery was recorded by digital imaging of colony growth, and each strain was scored for sensitivity or resistance (compared to wild

type) using image analysis and a simple algorithm. In essence, after adjusting for varying growth rates in the absence of exposure, treated strains displaying <67% of wild-type growth in at least two of the three replicates were scored as sensitive, and strains displaying >150% of wild-type growth in at least two replicates were scored as resistant; these parameters were chosen to minimize false positives while optimizing detection of true positives and were based on extensive reconstruction experiments (Fig. 2). Moreover, all of the sensitivity and resistance calls made by the simple algorithm were confirmed by visual inspection of the digitally imaged colonies.

Typical colony growth images are shown in Fig. 1B for exposure of 77 deletion strains (plus 6 control strains) to 0%, 0.01%, 0.02%, and 0.025% MMS. All three MMS-sensitive control strains showed diminished recovery at lower doses than did wild type, and analogous results were observed for 4NQO and UV (data not shown). At 0.03% MMS, virtually all strains failed to recover viability, including the wild-type strain (not shown), thus allowing us to identify resistant strains (an example, *yel033w* is shown in Fig. 1C).

Generation of a Genomic Phenotyping Database

To facilitate independent visual confirmation of the computational assignment of sensitive and resistant strains, we compiled the imaged data generated from this study (*i.e.*,

>90,000 digital images of the outgrowth from each spotted strain) into a genomic phenotyping database that is publicly accessible at <http://GenomicPhenotyping.mit.edu> (System Requirements: PC running Internet Explorer). In this database, the imaged colonies for each spotted strain are arranged from left to right with increasing dose of agent, generating visual killing curves for MMS, *t*-BuOOH, 4NQO, and UV. Examples are shown in Fig. 1C for the parent (BY4741), three controls (*mag1Δ*, *rev1Δ*, *rad14Δ*), a strain scored as sensitive to MMS, *t*-BuOOH, and 4NQO (*yml007w*), and a strain scored as MMS resistant (*yel033w*). Individual web pages were compiled for each of the 1615 deletion strains tested, displaying their triplicate visual killing curves for all four agents aligned with triplicates for the parental strain.

The genomic phenotyping database was used to visually confirm all the phenotype assignments, and these are categorized in Table 1. We identified a total of 416 MMS-sensitive, 67 *t*-BuOOH-sensitive, 149 4NQO-sensitive, 44 UV-sensitive, 23 MMS-resistant, 6 *t*-BuOOH-resistant, 39 4NQO-resistant, and zero UV-resistant strains. The data were compared to those reported in other screens (17, 18) and, among the limited number of strains and treatments in common, 71% of the assignments were in agreement; considering differences in exposure conditions, the agreement is very good. It is important to point out that for the other screens (17, 18), a different range of mutagens was used and only single-exposure doses tested. In

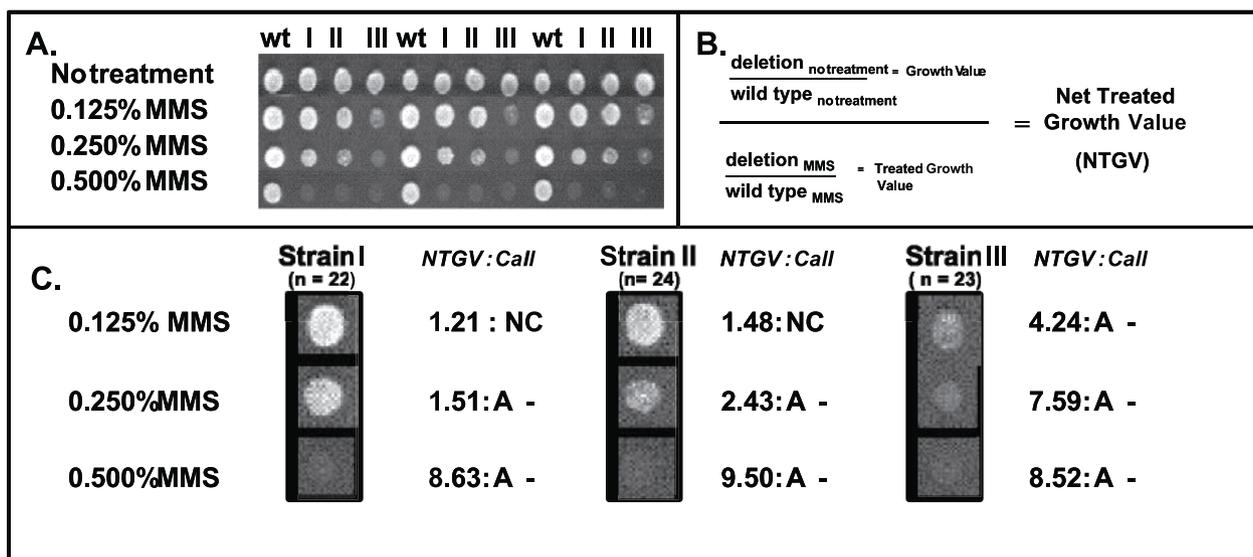


FIGURE 2. Detection standards used to establish the sensitivity of the screen. **A.** The control strains used for establishing MMS exposure conditions include wild type (*wt*), I (*mag1Δ*), II (*mec3Δmag1Δ*), and III (*rad9Δ mec3Δmag1Δ*). The colonies for each MMS dose are taken from separate plates, with this figure being the composite of four plates showing three representative replicates of each strain. **B.** The formula used for determining growth of deletion strains after damage, relative to wild-type growth. Deletion strains on the untreated plates were normalized to wild type on the same plate, to generate an untreated Growth Value. Deletion strains on the treatment plates were also normalized to wild type on the same plate to generate a Treated Growth Value. The Growth Value was then divided by the Treated Growth Value to yield a Net Treated Growth Value (NTGV). This algorithm was used to adjust for slow-growing strains and allowed for the identification of strains that grew significantly worse or better than wild type at each concentration agent tested. **C.** NTGV averaged from 22 to 24 replicates. To establish NTGV values that indicate sensitivity, we carried out multiple replicates, and examined the results visually. Strain II at 0.125% MMS (NTGV 1.48) and strain I at 0.250% MMS (NTGV 1.51) mark the bottom and top edge of the cutoff for detecting a sensitive strain. Thus, strains scoring an NTGV of 1.5 and above were called by the algorithm as sensitive, or A- meaning affected negatively. The average NTGV values of 22–24 replicates for each strain are shown for each treatment; there was less than 10% standard error for each (not shown). A call is made from the algorithm with assignments of NC = no classification, A- = growth affected negatively, *i.e.*, sensitive, and A+ = growth affected positively, *i.e.*, resistant. (There are no resistant calls in this figure.)

Table 1. Cellular Process Associated With Sensitive and Resistant Phenotypes^a

Process	Total Tested	Total MMS (sens/res)	Total <i>t</i> -BuOOH (sens/res)	Total 4NQO (sens/res)	Total UV (sens/res)	Found in subnetwork number ^b
Amino acid metabolism	40	12/0	4/0	0/4	1/0	2, 3, 6
ATP synthesis	4	4/0	0/0	1/0	0/0	-
Cell cycle	25	12/0	2/0	7/0	3/0	1, 5
Chromatin structure	9	5/0	1/0	3/1	1/0	1, 2, 5
Cytoskeleton	16	8/0	1/0	6/0	1/0	1, 4, 5
DNA repair	22	14/1	1/0	8/2	5/0	1, 5, 6
Glycolysis	12	4/0	0/0	1/0	0/0	2, 4
Mating	10	4/0	0/1	1/0	0/0	2
Meiosis	18	10/0	1/0	1/0	0/0	2, 3, 4
mRNA metabolism	22	5/1	1/0	2/1	0/0	1, 2, 5
Protein degradation	29	11/0	2/0	2/2	1/0	1, 2, 3, 4
Protein processing	7	4/0	1/0	0/0	0/0	3, 5
Protein synthesis	61	10/2	5/2	11/3	2/0	1, 3, 5, 6
Protein targeting	25	12/0	4/0	2/1	1/0	2, 4, 6
Secretion	26	8/0	1/0	3/1	0/0	4
Signaling	19	8/0	2/0	2/0	0/0	4
Silencing	6	4/0	0/0	2/0	0/0	5
Sterol metabolism	9	5/0	3/0	4/0	1/0	4
Stress response	26	9/0	2/0	5/0	0/0	3, 6
Transcription	25	8/0	2/1	4/1	2/0	1, 3, 5, 6
Transport	39	7/0	1/0	4/0	1/0	-
Vacuole function	5	4/0	2/0	5/0	1/0	2
All other processes combined ^a	240	54/1	7/0	18/6	6/0	1, 2, 3, 4, 5, 6
Unknown	919	194/17	24/1	56/16	17/0	1, 2, 3, 4, 5, 6
Total	1615	416/23	67/6	149/39	44/0	

^aTwenty-two processes are individually listed here; the others are listed as “all other processes combined,” but can be viewed individually on the website. Compiled from YPD and SGD (35, 36) (<http://GenomicPhenotyping.mit.edu/pages/papertable2.html>). On the website, each number in Table 1 is linked to the relevant strain list, ordered from the strongest to the weakest phenotype (this was only relevant for sensitive strains), and each strain on the list links to the web page displaying its triplicate screening results. ^bRefers to Fig. 5B.

addition, a third study claimed to identify more than 100 genes that affect recovery after exposure to four damaging agents (nonoverlapping with those used here), but in their report only provided the identities of 20 of those genes (19). Here we have tested a wide range of mutagen doses and all of our results, including the primary data, are publicly available.

Many strains were uniquely sensitive or resistant to a single agent (particularly MMS), whereas others scored as sensitive or resistant to more than one agent. The pattern of sensitive and resistant strains was organized by hierarchical clustering (20) to show the distinct signatures for each agent (Fig. 3A). Similar to a recent report (19), we found that the deletion of transcriptionally responsive genes (3, 4) was no more likely to generate mutagen sensitivity than was the deletion of transcriptionally nonresponsive genes. For instance, for the 1615 strains analyzed, 26% of strains deleted for MMS-responsive genes, and 26% of strains deleted for MMS-nonresponsive genes, displayed an MMS-sensitive phenotype (Fig. 3B). Similar results were obtained for 4NQO- and *t*-BuOOH-sensitivity *versus* transcriptional responsiveness (Fig. 3B). Finally, like the sensitive strains, the MMS-, 4NQO-, and *t*-BuOOH-resistant strains were equally represented among strains deleted for responsive and nonresponsive genes (data not shown). We conclude that transcriptional responsiveness *per se* is less predictive than genomic phenotyping for pinpointing pathways that influence damage recovery, but this is not to say that transcriptional responsiveness will not correlate to some other phenotypic end point. We propose that genomic phenotyping represents a robust analytical tool for both identifying toxicants

and for distinguishing between toxicant classes. With increased throughput, genomic phenotyping may hold more promise than transcriptional profiling for identifying and eliminating toxic compounds during drug development.

Diverse Pathways Influence Damage Recovery

Table 1 underscores that eukaryotic cells engage a wide range of gene products to combat toxicity induced by damaging agents; on our website, the numbers in Table 1 link to the relevant gene list, primary data and web-based information for each gene. As expected for cells treated with agents that damage DNA, many of the sensitive strains bear deletions in genes for DNA repair and cell cycle control. The phenotypes of these DNA repair and cell cycle control mutants do not just fall into the most sensitive category, but rather span the entire range of high, medium, and modest sensitivity; this is important in the context of judging the relevance of newly identified genes that, when deleted, generate modest sensitivity to a mutagen. In fact, 63% of the agent-sensitive strains could be classified as being modestly sensitive (*i.e.*, only showing a phenotype at the higher exposure doses); 23% were of medium sensitivity and 14% of high sensitivity. Interestingly, >40% of the deleted genes that generate a damage-sensitive phenotype are assigned in public databases as being of unknown function, and many of these are even more sensitive to damaging agents than known DNA repair and cell cycle control mutants.

One of the largest categories of agent-sensitive strains is that comprised of strains deficient in various aspects of protein metabolism, namely protein degradation, processing, synthesis

and targeting, as well as amino acid metabolism. Furthermore, processes involved in trafficking proteins from the endoplasmic reticulum (ER) to the Golgi for secretion and to the vacuole for degradation, appear to profoundly influence cellular recovery,

especially after MMS exposure. At least part of the protein-related responses involves the unfolded protein response (UPR) signaled by the Ire1 endoplasmic reticulum-membrane-bound protein kinase (21), but it also involves protein degradation via

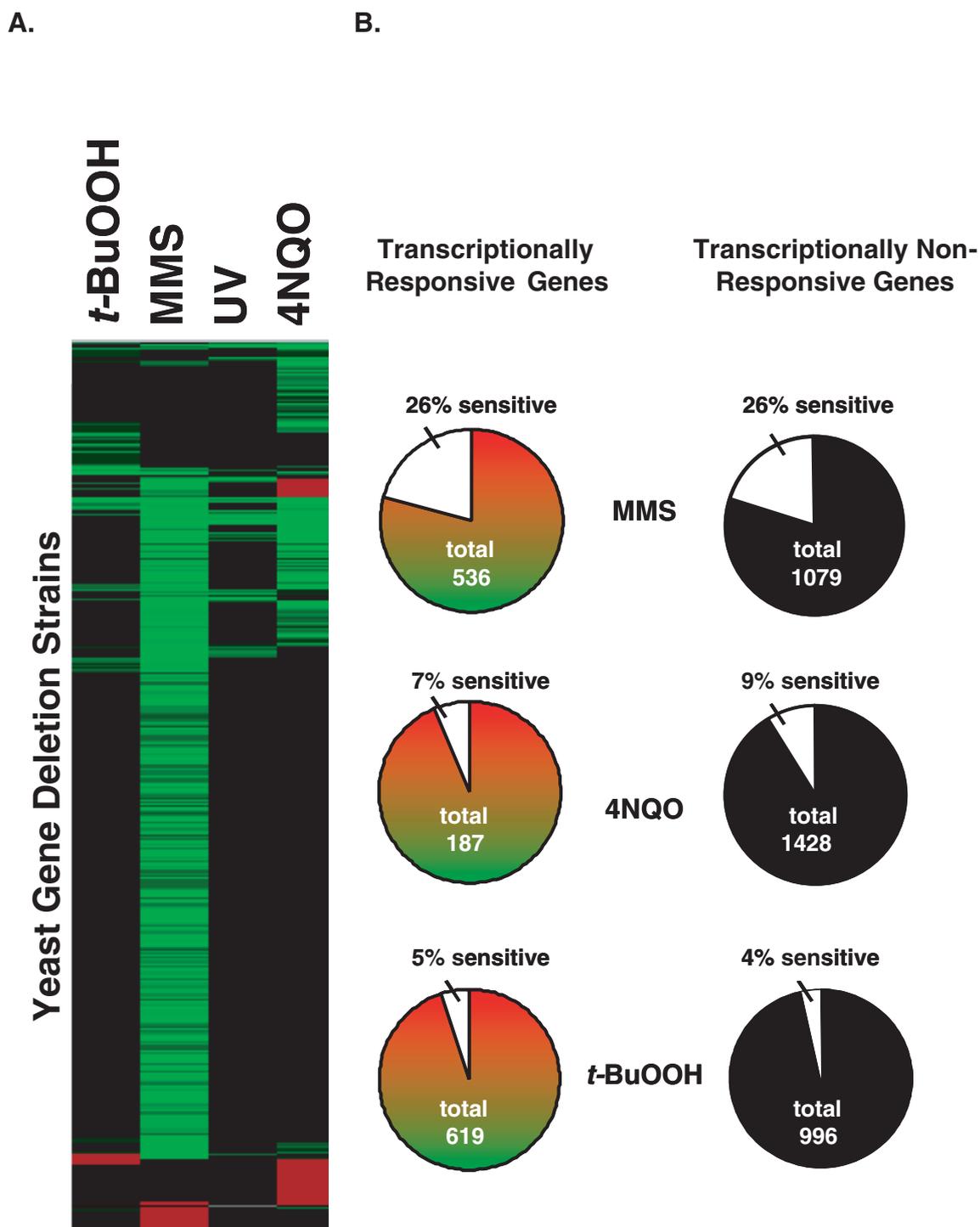


FIGURE 3. Hierarchical clustering of genomic phenotyping data and integration with transcriptional responsiveness. **A.** Phenotypic values for all 536 agent-sensitive or -resistant strains were assigned based on sensitivity (*green*), resistance (*red*), and no phenotype (*black*); average linked clustering was performed as described (20). The *Y* axis corresponds to individual gene deletion strains and the *X* axis indicates treatment. **B.** The 1615 genes tested in this study were divided into transcriptionally responsive (*red/green*) or nonresponsive (*black*) to MMS, 4NQO, and *t*-BuOOH (3, 4), and for each group, the fraction of genes that when deleted generate an MMS-, 4NQO- or *t*-BuOOH-sensitive phenotype is shown in white.

both the proteasome and the vacuole. From these phenotypic data, and from previous analyses of global transcriptional responses to MMS exposure (3–6), we propose that eukaryotic cells mount a concerted response to the induction of protein damage on exposure to these so-called DNA damaging agents and that in the absence of these responses, cells are more likely to die.

Several gene deletions affecting chromatin assembly, structure, and silencing were classified as MMS sensitive (*msi1*, *spt8*, *htz1*, *yor023c*, *hho1*, *san1*, *ydr363w*, *yhr154w*, *hst3*, *hfi1*, *Ycl010c*) and 4NQO sensitive (*msi1*, *spt3*, *spt8*, *ydr363w*, *yhr154w*). These gene products affect chromatin by influencing nucleosome assembly and by modulating histone acetylation (22, 23). Indeed, the Spt3, Spt8, Hfi1, and Ycl010c proteins comprise part of the Spt-Ada-Gcn5-histone acetyltransferase transcriptional activation complex (SAGA), and it seems that SAGA disruption produces a damage-sensitive phenotype. Proteins involved in histone acetylation may affect damage sensitivity by influencing the expression of genes needed for recovery. However, it is also possible that overall changes in acetylation of *S. cerevisiae* histones could both affect the reactivity of DNA with damaging agents and affect access for DNA repair proteins, analogous to affecting access for transcriptional factors (24, 25). Alternatively, in wild-type cells, it may be important to specifically modify chromatin structure adjacent to sites of DNA damage to provide targeted access for DNA repair proteins. Chromatin structural proteins can also influence agent toxicity, because deletion of the histone proteins, Hho1 and Htz1, generates MMS sensitivity. Again, we infer that such structural changes might influence the transcription of genes important for recovery, influence the reactivity of DNA with MMS, or influence the access of DNA repair proteins to damaged DNA. Whatever the mechanisms, it is clear that chromatin structure has a strong influence on damage recovery.

Some of the most sensitive mutants identified in this screen turn out to be deficient in the synthesis of ergosterol, a constituent of the cell membrane thought to contribute to membrane permeability and fluidity (*erg3*, 4, 5, 6, and 24). Increased permeability can readily explain increased sensitivity to chemical damaging agents, but we were surprised to find that *erg6* cells are also modestly UV sensitive. It is possible that, in addition to affecting permeability, ergosterol deficiency diminishes membrane-mediated signal transduction events needed for damage recovery. Indeed, UV light initiates membrane receptor-mediated signaling in mammalian cells (26–29), and *erg6* mutants display altered receptor-mediated signaling in *S. cerevisiae* (30).

Manual Interactome Mapping of Genomic Phenotyping Data

Navigating lists of genes affecting particular phenotypes becomes unwieldy at the genomic level, in particular when trying to integrate such information with other genomic data sets. Nevertheless, we set out to analyze our genomic phenotyping data in the framework of 12,232 protein-protein and protein-DNA interactions comprising the known yeast interactome. Initial integration of our data with the yeast interactome was facilitated using the Curagen protein-protein

interaction database (<http://www.curagen.com>). Manual searching of interaction maps for the proteins the absence of which produced a phenotype, revealed a number of interesting multiprotein interaction hubs, one of which is illustrated in Fig. 4. From amongst the top 40 MMS-sensitive strains, we identified three gene products that lie in the same interaction network, namely the Ymr032w/Cyk2 protein involved in cytokinesis, and two proteins of hitherto unknown function (Yhl006c and Ydr078c). Also included in this network are the Hda1 histone deacetylase, the Hpr5/Sgs2 DNA helicase, plus another protein of unknown function, Ylr423c, which were not represented in the 1615 strains tested. Because of their association with other proteins that contribute to MMS resistance, we tested Hda1-, Hpr5/Sgs2-, and Ylr423c-deficient strains for their MMS sensitivity. All three were very sensitive (Fig. 4, data not shown), and so collectively, these six proteins contribute to a hitherto unknown damage recovery pathway, now ripe for further investigation.

Computational Interactome Mapping of Genomic Phenotyping Data

Although our initial attempts at interactome mapping were quite successful, it soon became apparent that manual searching of the protein interaction network was extremely inefficient and likely to be error-prone. Accordingly, we performed automatic

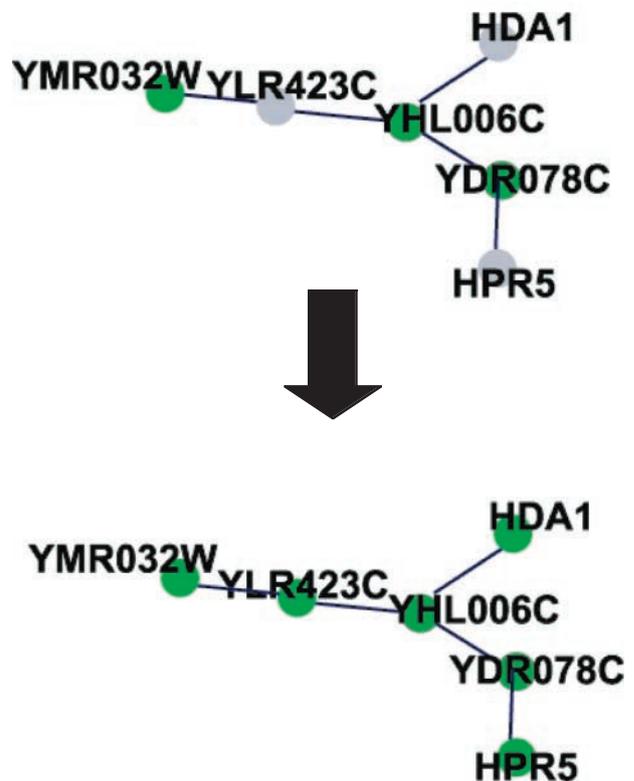


FIGURE 4. Interacting protein network identified manually from the Curagen protein interaction database. *Green nodes* represent proteins identified as being required for MMS resistance; *gray nodes* represent proteins that were not tested as part of the 1615 strains used in this study; *blue lines* represent protein-protein interactions. The untested gene deletion strains (*yhr423c*, *hda1*, and *hpr5*) were subsequently tested for MMS sensitivity; all were found to be sensitive (*bottom*).

network searches using methods provided in the newly developed software package *Cytoscape* (31). Using this software, we performed a comprehensive search of the full 12,232-interaction network in which 4,025 proteins are currently woven into the same connected circuit component, meaning that each of these proteins can connect to any other through some chain of protein-protein or protein-DNA interactions. Our goal was to identify and visualize *subnetworks* within the interactome, *i.e.*, connected groups of proteins in which a significant proportion were associated with sensitivity or resistance to a damaging agent in our genomic phenotyping database. To do this, the full network of 12,232 molecular interactions (Fig. 5A, [i]) was filtered to exclude proteins that were more than one interaction away from proteins the deletions of which conferred sensitive or resistant phenotypes. Within this filtered network (Fig. 5A, [ii]), we performed an automated search to identify specific regions with a higher-than-expected proportion of proteins associated with sensitivity (an example is shown in Fig. 5A, [iii]). In searching for particular subnetworks, it is thus possible to reduce the immense complexity of the entire interaction network by pinpointing just those network regions relevant to a particular end point, in this case a sensitive or resistant phenotype. In addition, false-positive protein interactions are unlikely to correlate with phenotype and thus are generally removed from further analysis. The remaining subnetworks provide hypotheses as to the structural, signaling, and regulatory pathways contributing to viable recovery after exposure to damaging agents.

The highest scoring subnetworks are shown in Fig. 5B. Subnetworks (1)–(4) are associated with MMS sensitivity and subnetworks (5) and (6) are associated with sensitivity to any one of the other three agents. These six large subnetworks comprised of between 18 and 29 proteins emerged from the *Cytoscape* analysis with *P* values ranging from 0.0013 to 0.014. Almost all of the proteins in each network contribute to damage recovery (indicated by the green notation). Importantly, the group of interacting proteins described in Fig. 4 was identified as a component of subnetwork (2) in Fig. 5B. In addition to the six large networks shown, several other smaller protein networks required for recovery after damage were also identified (shown at <http://GenomicPhenotyping.mit.edu>). Interestingly, searches for subnetworks of connected proteins the gene deletions of which lead to damage resistance, did not reveal any significant clusters within the interaction network.

Like the information in Table 1, Fig. 5B illustrates the engagement of many different cellular processes to aid the recovery of *S. cerevisiae* from damage; the processes embraced by each subnetwork are indicated in Table 1. Subnetwork (1) in Fig. 5B contains a group of DNA damage response proteins (Mec1, Rad17, Swi6, and Cln2) that are indirectly linked to the SAGAA chromatin remodeling complex proteins (Spt8, Spt7, Hfl1, and Ycl010c), via two proteins involved in transcriptional regulation (Yap1 and Med2). That a group of DNA damage response proteins should be physically linked to transcription factors, one of which is known to activate stress response genes (Yap1) (32), as well as to the SAGAA chromatin remodeling complex, supports the idea that the SAGAA complex may provide damage resistance by participating in transcriptional activation of genes whose products aid recovery. Subnetwork

(2) is predominated by gene products involved in the degradation of proteins (Vsp9, Aap1, Pep3, Vsp41, Vma22, Vma6, Pep12, and Snf7), but in addition, as mentioned, contains the interacting hub of proteins found by brute force screening and shown in Fig. 4. Subnetwork (2) also contains several proteins of unknown function, and proteins involved in chromatin silencing (Spp1, Swd3, and Bre2), RNA processing (Npl3, Mud13, and Snu7), and amino acid metabolism (Asn2 and Lys14). Subnetworks (3) and (4) contain a large number of proteins involved in chromosome segregation and the associated changes in cytoskeleton required to achieve this; these include Rif2, Cin8, Tpd3, Bem1, Spo11, Mck1, Gbp2, Spo16, Rec104 for subnetwork (3) and Ssp1, Tem1, Hof1, Ste50, Nip100, and Arp1 for subnetwork (4). In addition, subnetworks (3) and (4) contain a number of proteins involved in protein synthesis and protein trafficking, namely Sse1, Spo7, His7, Rpl5, Ard1, Ssf1 for subnetwork (3) and Spc2, Tlg2 and Rsm27 for subnetwork (4). The absence of virtually any one of the proteins in each subnetwork renders cells less able to recover from MMS-induced damage.

Fig. 5B also shows two subnetworks, (5) and (6), for proteins identified as contributing to recovery from *t*-BuOOH, 4NQO, or UV exposure (phenotypic data for all three agents were combined). Both networks contain a number of proteins involved in protein synthesis and protein trafficking (Yke2, Ylr119w, Rpl18b, Rps17a, Ard1, Rpl8A, Rpl7A, Tef4, Sse1, Ilv1, Mrp4, Tlg2, Rsm27). Subnetwork (5), like subnetwork (1), contains a hub of proteins surrounding the Mec3 DNA damage response protein. However, in this case, the Mec3 hub does not link to the SAGAA histone acetylation complex, but rather links to a group of proteins involved in telomere maintenance and chromosome segregation (Rap1, Rif1, Cin8, and Top1). We infer that connection of the Mec3 DNA damage response hub of proteins to a different part of the interactome reflects the different pathways that are engaged for recovery from *t*-BuOOH, 4NQO, or UV *versus* recovery from MMS.

It is important to point out that the 23 gray nodes included in the protein subnetworks shown in Figs. 4 and 5B represent proteins which were not initially tested, simply because they were not represented in the set of 1615 deletion strains screened in this study. We therefore determined whether the inclusion of an untested protein in a subnetwork heavily populated by proteins known to contribute to a damage resistance phenotype, predicts whether that protein also contributes to damage resistance. We therefore tested the mutagen sensitivity of 16 of 23 proteins represented by gray nodes (the other 7 were encoded by essential genes and could not be tested). Eleven of the 16 scored as mutagen sensitive, representing a significant enrichment (with a *P* value of <0.0035) compared to the overall frequency of mutagen sensitivity observed in this study (535/1615). Thus, combining genomic phenotyping with interactome mapping can provide a predictive component to an analysis with incomplete genome coverage. Such predictive capability is important because it will likely be some years until entire mammalian genomes are represented in libraries of mutant mammalian cell lines, which will presumably be generated by systematic application of RNAi technology (33, 34).

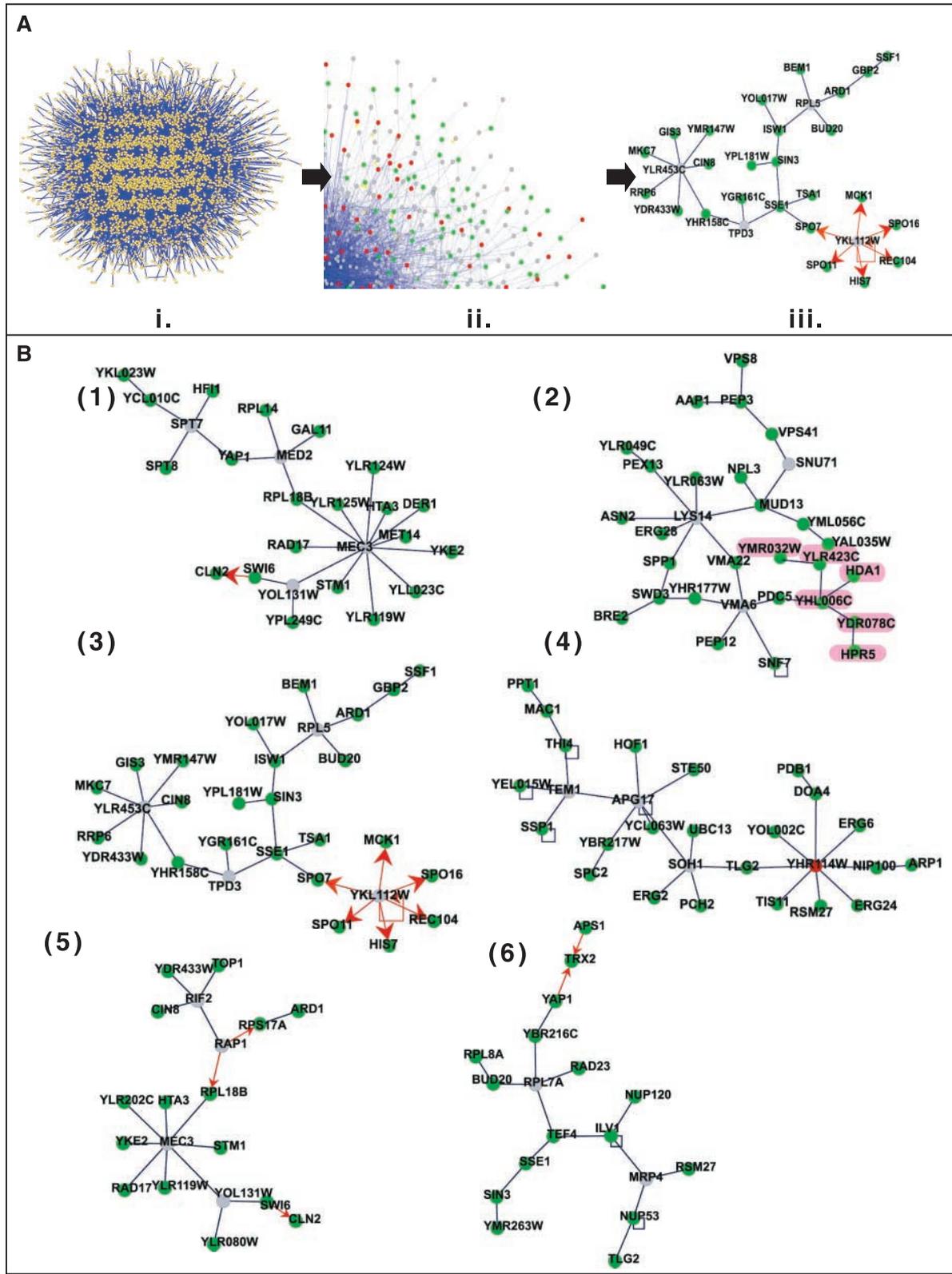


FIGURE 5. Interacting protein networks identified computationally using the *Cytoscape* software. **A.** The interactome [i] was filtered [ii] and searched [iii]. In [iii] and the rest of the figure, green nodes represent proteins associated with sensitivity, red nodes represent proteins the deletions of which are not sensitive, and gray nodes represent deletions that were not initially tested. Blue lines represent protein-protein interactions, and red arrows represent protein-DNA interactions. **B.** High-scoring MMS recovery pathways (1)–(4) with respective *P* values 31: (1) $P < 0.0062$; (2) $P < 0.0013$ (proteins highlighted in pink represent the interacting network shown in Fig. 3); (3) $P < 0.0013$; and (4) $P < 0.0062$. The high-scoring subnetworks (5) $P < 0.014$ and (6) $P < 0.013$ are combined for the *t*-BuOOH, 4NQO, and UV recovery pathways.

Concluding Remarks

Global analyses of biological systems are steadily unveiling new roles for old proteins and uncovering hitherto unsuspected connections between diverse cellular pathways. Indeed, the connectedness of so many proteins in the yeast interactome, albeit dynamic and certain to change depending on environmental cues, has fundamentally changed the way one must think about both the proximal and distal effects of eliminating proteins from the network. The present study combines genomic phenotyping with a newly developed computational method for merging genomics databases (*Cytoscape*) and demonstrates how systematic phenotypic assays may be directly linked to underlying molecular mechanisms. We have uncovered an extremely diverse set of mechanisms for cellular recovery on mutagen exposure, mechanisms that may now be viewed as influencing the accumulation of mutations in cell populations. Our next goal is to accelerate throughput for the biological screen to generate genomic phenotyping profiles for very large numbers of agents, exposures, and growth conditions, for both yeast and mammalian cells. Hence, we will be able to address the dynamics of how cells deploy interconnected protein hubs embedded in their interactome, viewing the biological system as a whole.

Materials and Methods

Strains and Plasmids

S. cerevisiae, strain BY4741 and deletion mutant derivatives were supplied by Research Genetics, Carlsbad, CA. Parental strain BY4741 was transformed with plasmid pYE13g (American Type Culture Collection, Manassas, VA) and selected for on YPD (10 g yeast extract, 20 g peptone, 20 g dextrose, 20 g agar/liter) containing 200 µg/ml G418.

Growth Conditions for Liquid Cultures

Ninety-six-well master plates containing individual deletion strains were thawed and a 96-pin tool was used to transfer them to a sub-master plate (Corning, Corning, NY) containing 150 µl of YPD supplemented with G418 at 200 µg/ml. There were at least six empty wells on each master plate. The wild-type controls were spotted into empty wells in the plate along with three damage-sensitive control strains, *mag1Δ*, *rad14Δ*, and *rev1Δ*. Strains in sub-master plates were grown for 2 days at 30°C. Settled yeast cells in the sub-master plate were mixed using the 96-pin tool and replicated to another 96-well plate containing 150 µl of YPD supplemented with G418 at 200 µg/ml and grown for 36 h at 30°C.

Growth Conditions for Spotted Strains

Settled cells in each position of the 96-well plate were resuspended with 60 µl bursts of forced air from a Hydra liquid handling apparatus (Robbins Scientific, Sunnyvale, CA), which on average provided a suspension of 10^7 cells/ml. Thirty microliters of cell suspensions were drawn into each syringe of the Hydra and ninety-six 1-µl samples were spotted simultaneously onto an agar-containing plate. MMS, *t*-BuOOH, and 4NQO were purchased from Aldrich, St. Louis, MO. UV radiation (254 nm) was supplied from a UV Stratalinker 2400

(Stratagene, La Jolla, CA). Plates containing up to 96 strains were tested using the following conditions: no treatment, 0.01% MMS, 0.02% MMS, 0.025% MMS, 0.03% MMS, 0.25 mM *t*-BuOOH, 0.50 mM *t*-BuOOH, 0.75 mM *t*-BuOOH, 1 mM *t*-BuOOH, 0.1 µg/ml 4NQO, 0.2 µg/ml 4NQO, 0.3 µg/ml 4NQO, 0.4 µg/ml 4NQO, 25 J/m² UV, 50 J/m² UV, 75 J/m² UV, and 100 J/m² UV. Chemicals were added to cooled agar on plate preparation. UV treatments were performed after strains had been spotted and dried on the plates. Strains were grown for 60 h at 30°C and imaged using a Gel Doc 1000 from BioRad (Hercules, CA) running Quantity One software. Images were analyzed using ScanAlyze software to quantitate the pixel intensity of each spotted colony. All screens were performed in triplicate using fresh liquid cultures.

Database Construction

At website <http://GenomicPhenotyping.mit.edu>, we have compiled images from 1275 plates and batch processed them using the software program U Lead Smart Saver Pro, version 3.0. This program divided each plate into 96 separate plate pieces, each containing an image of a single strain; these were recompiled into strain-specific rows in a hierarchical website containing >1700 web pages using in-house-prepared visual basic scripts via Microsoft applications. Table 1, hierarchical clustering, and Cytoscape-generated figures, with links to relevant gene lists, can be found at this address. In addition, lists of gene deletion strains tested and agent-specific phenotype assignments are found here, along with basic and advanced search methods that allow for efficient mining of the database.

Interactome Mapping

A composite interaction network of 4,232 proteins was constructed from 12,232 previously characterized molecular interactions in yeast, including 5,003 two-hybrid interactions catalogued in BIND (11), 6,925 additional protein-protein interactions determined by coimmunoprecipitation studies (12), and 304 protein-DNA interactions recorded in the TRANSFAC database (13). To evaluate phenotypic data on the filtered interaction network, we used a network scoring and search procedure as described by Ideker *et al.* (31). Briefly, each protein was assigned a score of +1.25, -0.25, or +0.25 representing sensitivity, nonsensitivity, or missing data, respectively. Subnetwork scores z_A were computed using the sum $z_A = \frac{1}{\sqrt{k}} \sum_{i=1}^k z_i$ (where k is the number of proteins in the subnetwork and z_i is the score of subnetwork protein i) and were calibrated against a “background” distribution of random sets of k proteins drawn independently of the network. Subnetworks with significantly high scores were identified using a heuristic search algorithm based on simulated annealing. We ran the algorithm with parameters [$N = 10^8$, $T_i = 1$, $T_f = 0.01$, $M = 20$, $d_{\min} = 70$] and required that the identified subnetworks contain less than 30 proteins.

Acknowledgments

We thank Drs. David Hunter, John Cairns, Philip Sharp, Ben van Houten, and Sam Wilson, plus members of the Samson lab, for their critical comments. We also thank Drs. Scott Jelinsky and Deepak Rajpal for some initial experiments.

References

- Friedberg, E. C., Walker, G. C., and Siede, W. DNA Repair and Mutagenesis. Washington, DC: American Society for Microbiology Press, 1995.
- Wahl, G. M. and Carr, A. M. The evolution of diverse biological responses to DNA damage: insights from yeast and p53. *Nat. Cell Biol.*, 3: E277–E286, 2001.
- Jelinsky, S. A. and Samson, L. D. Global response of *Saccharomyces cerevisiae* to an alkylating agent. *Proc. Natl. Acad. Sci. USA*, 96: 1486–1491, 1999.
- Jelinsky, S. A., Estep, P., Church, G. M., and Samson, L. D. Regulatory networks revealed by transcriptional profiling of damaged *Saccharomyces cerevisiae* cells: rpn4 links base excision repair with proteasomes [In Process Citation]. *Mol. Cell. Biol.*, 20: 8157–8167, 2000.
- Gasch, A. P., Huang, M., Metzner, S., Botstein, D., Elledge, S. J., and Brown, P. O. Genomic expression responses to DNA-damaging agents and the regulatory role of the yeast ATR homolog Mec1p. *Mol. Biol. Cell*, 12: 2987–3003, 2001.
- Causton, H. C., Ren, B., Koh, S. S., Harbison, C. T., Kanin, E., Jennings, E. G., Lee, T. I., True, H. L., Lander, E. S., and Young, R. A. Remodeling of yeast genome expression in response to environmental changes. *Mol. Biol. Cell*, 12: 323–337, 2001.
- Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi-Emili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadmodar, G., Yang, M., Johnston, M., Fields, S., and Rothberg, J. M. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature*, 403: 623–627, 2000.
- Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. USA*, 98: 4569–4574, 2001.
- Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D., Moore, L., Adams, S. L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., Yang, L., Wolting, C., Donaldson, I., Schandorff, S., Shewnarane, J., Vo, M., Taggart, J., Goudeault, M., Muskaf, B., Alfarano, C., Dewar, D., Lin, Z., Michalickova, K., Willems, A. R., Sassi, H., Nielsen, P. A., Rasmussen, K. J., Andersen, J. R., Johansen, L. E., Hansen, L. H., Jepsen, H., Podtelejnikov, A., Nielsen, E., Crawford, J., Poulsen, V., Sorensen, B. D., Matthiesen, J., Hendrickson, R. C., Gleeson, F., Pawson, T., Moran, M. F., Durocher, D., Mann, M., Hogue, C. W., Figeys, D., and Tyers, M. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature*, 415: 180–183, 2002.
- Ren, B., Robert, F., Wyrick, J. J., Aparicio, O., Jennings, E. G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., Volkert, T. L., Wilson, C. J., Bell, S. P., and Young, R. A. Genome-wide location and function of DNA binding proteins. *Science*, 290: 2306–2309, 2000.
- Bader, G. D., Donaldson, I., Wolting, C., Ouellette, B. F., Pawson, T., and Hogue, C. W. BIND—The Biomolecular Interaction Network Database. *Nucleic Acids Res.* 29: 242–245, 2001.
- Gavin, A. C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat, C. M., Remor, M., Hofert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M. A., Copley, R. R., Edelman, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G., and Superti-Furga, G. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, 415: 141–147, 2002.
- Wingender, E., Chen, X., Fricke, E., Geffers, R., Hehl, R., Liebich, I., Krull, M., Matys, V., Michael, H., Ohnhauser, R., Pruss, M., Schacherer, F., Thiele, S., and Urbach, S. The TRANSFAC system on gene expression regulation. *Nucleic Acids Res.*, 29: 281–283, 2001.
- Chen, J., Derfler, B., Maskati, A., and Samson, L. Cloning a eukaryotic DNA glycosylase repair gene by the suppression of a DNA repair defect in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA*, 86: 7961–7965, 1989.
- Bankmann, M., Prakash, L., and Prakash, S. Yeast RAD14 and human xeroderma pigmentosum group A DNA-repair genes encode homologous proteins. *Nature*, 355: 555–558, 1992.
- Lawrence, C. W. and Christensen, R. UV mutagenesis in radiation-sensitive strains of yeast. *Genetics*, 82: 207–232, 1976.
- Ross-Macdonald, P., Coelho, P. S., Roemer, T., Agarwal, S., Kumar, A., Jansen, R., Cheung, K. H., Sheehan, A., Symoniatis, D., Umansky, L., Heidman, M., Nelson, F. K., Iwasaki, H., Hager, K., Gerstein, M., Miller, P., Roeder, G. S., and Snyder, M. Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature*, 402: 413–418, 1999.
- Bennett, C. B., Lewis, L. K., Karthikeyan, G., Lobachev, K. S., Jin, Y. H., Sterling, J. F., Snipe, J. R., and Resnick, M. A. Genes required for ionizing radiation resistance in yeast. *Nat. Genet.*, 29: 426–434, 2001.
- Birrell, G. W., Brown, J. A., Wu, H. I., Giaever, G., Chu, A. M., Davis, R. W., and Brown, J. M. Transcriptional response of *Saccharomyces cerevisiae* to DNA-damaging agents does not identify the genes that protect against these agents. *Proc. Natl. Acad. Sci. USA*, 99: 8778–8783, 2002.
- Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA*, 95: 14863–14868, 1998.
- Kaufman, R. J. Stress signaling from the lumen of the endoplasmic reticulum: coordination of gene transcriptional and translational controls. *Genes Dev.*, 13: 1211–1233, 1999.
- Bernstein, B. E., Tong, J. K., and Schreiber, S. L. Genomewide studies of histone deacetylase function in yeast. *Proc. Natl. Acad. Sci. USA*, 97: 13708–13713, 2000.
- Robyr, D., Suka, Y., Xenarios, I., Kurdistani, S. K., Wang, A., Suka, N., and Grunstein, M. Microarray deacetylation maps determine genome-wide functions for yeast histone deacetylases. *Cell*, 109: 437–446, 2002.
- Lee, D. Y., Hayes, J. J., Pruss, D., and Wolffe, A. P. A positive role for histone acetylation in transcription factor access to nucleosomal DNA. *Cell*, 72: 73–84, 1993.
- Belotserkovskaya, R., Sterner, D. E., Deng, M., Sayre, M. H., Lieberman, P. M., and Berger, S. L. Inhibition of TATA-binding protein function by SAGA subunits Spt3 and Spt8 at Gen4-activated promoters. *Mol. Cell. Biol.*, 20: 634–647, 2000.
- Devary, Y., Gottlieb, R. A., Smeal, T., and Karin, M. The mammalian ultraviolet response is triggered by activation of Src tyrosine kinases. *Cell*, 71: 1081–1091, 1992.
- Devary, Y., Rosette, C., DiDonato, J. A., and Karin, M. NF-kappa B activation by ultraviolet light not dependent on a nuclear signal. *Science*, 261: 1442–1445, 1993.
- Sachsenmaier, C., Radler-Pohl, A., Zinck, R., Nordheim, A., Herrlich, P., and Rahmsdorf, H. J. Involvement of growth factor receptors in the mammalian UVC response. *Cell*, 78: 963–972, 1994.
- Rosette, C. and Karin, M. Ultraviolet light and osmotic stress: activation of the JNK cascade through multiple growth factor and cytokine receptors. *Science*, 274: 1194–1197, 1996.
- Sitcheran, R., Emter, R., Kralli, A., and Yamamoto, K. R. A genetic analysis of glucocorticoid receptor signaling. Identification and characterization of ligand-effect modulators in *Saccharomyces cerevisiae*. *Genetics*, 156: 963–972, 2000.
- Ideker, T., Ozier, O., Schwikowski, B., Siegel, A. Discovering regulatory and signaling circuits in molecular interaction networks. *Bioinformatics*, 18: 233–240, 2002.
- Kuge, S. and Jones, N. YAP1 dependent activation of TRX2 is essential for the response of *Saccharomyces cerevisiae* to oxidative stress by hydroperoxides. *EMBO J.*, 13: 655–664, 1994.
- Sui, G., Soohoo, C., Affar el, B., Gay, F., Shi, Y., and Forrester, W. C. A DNA vector-based RNAi technology to suppress gene expression in mammalian cells. *Proc. Natl. Acad. Sci. USA*, 99: 5515–5520, 2002.
- Brummelkamp, T. R., Bernards, R., and Agami, R. A system for stable expression of short interfering RNAs in mammalian cells. *Science*, 296: 550–553, 2002.
- Costanzo, M. C., Crawford, M. E., Hirschman, J. E., Kranz, J. E., Olsen, P., Robertson, L. S., Skrzypek, M. S., Braun, B. R., Hopkins, K. L., Kondu, P., Lengieza, C., Lew-Smith, J. E., Tillberg, M., and Garrels, J. I. YPD, PombePD and WormPD: model organism volumes of the BioKnowledge library, an integrated resource for protein information. *Nucleic Acids Res.*, 29: 75–79, 2001.
- Cherry, J. M., SGD. "Saccharomyces Genome Database" (<http://genome-www.stanford.edu/Saccharomyces/>).