Hot Spots for Modulating Toxicity Identified by Genomic Phenotyping and Localization Mapping

Thomas J. Begley,[1,3] Ari S. Rosenbach,[1]
Trey Ideker,[2] and Leona D. Samson[1,3,*]
[1]Biological Engineering Division and
Center for Environmental Health Sciences
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, Massachusetts 02139
[2]Department of Bioengineering
University of California, San Diego
9500 Gilman Drive
La Jolla, California 92093

Summary

DNA repair and checkpoint pathways protect against carcinogen-induced toxicity. Here, we describe additional, equally protective pathways discovered by interrogating 4,733 yeast proteins for their ability to diminish toxicity induced by four known carcinogens. A computational mapping strategy for global phenotypic data was developed to build a systems toxicology model detailing recovery from carcinogen exposure and identifying protein complexes that modulate toxicity. Global phenotypic data were merged with global subcellular localization and protein interactome data to generate an integrated picture of cellular recovery after carcinogen exposure. Statistically validated results from this systems-wide integration demonstrate that, in addition to the nucleus, subnetworks of toxicity-modulating proteins were overrepresented in the vacuolar membrane, endosome, endoplasmic reticulum, and mitochondrion. In addition, we show that many proteins associated with RNA polymerase II, macromolecular trafficking, and vacuole function can now be counted among the many proteins that modulate carcinogen-induced toxicity.

Introduction

Genomic analyses of *Saccharomyces cerevisiae* have included transcriptional profiling (Hughes et al., 2000), two hybrid screens (Ito et al., 2001; Uetz et al., 2000), transcription factor binding analysis (Lee et al., 2002; Ren et al., 2000), synthetic genetic analysis (Tong et al., 2001), subcellular localization (Huh et al., 2003), and phenotypic studies (Bennett et al., 2001; Chang et al., 2002; Ross-Macdonald et al., 1999; Winzeler et al., 1999). Each approach has yielded catalogs of genes and proteins with defined characteristics and required computational methods to provide biological insights, such as the identification of promoter-specific transcriptional modules for coregulated genes (Jelinsky et al., 2000; Lee et al., 2002; Ren et al., 2000), the demonstration that a large

number of redundant activities exist in the cell (Tong et al., 2004), and the illustration of connectivity between cellular organelles (Huh et al., 2003).

Exposure to environmental DNA-damaging agents can cause mutations and cell death and can promote transformations to diseased states. Fortunately, cells have a battery of DNA repair pathways that contribute to recovery after exposure (Friedberg et al., 1995). *S. cerevisiae* has ~150 known DNA repair proteins that handle damage produced by a wide range of DNA-damaging agents (Friedberg et al., 1995); these and other proteins that regulate DNA repair and activate cell cycle checkpoints play an important role in promoting viable recovery after damage (Elledge, 1996; Jelinsky et al., 2000; Marenstein et al., 2001; Navas et al., 1996; Zhou and Elledge, 1993). However, recent studies have begun to show that a wide range of other cellular activities (e.g., lipid, protein, and RNA metabolic processes) (Begley et al., 2002; Bennett et al., 2001; Ross-Macdonald et al., 1999) can also alleviate the toxic effects of DNA-damaging agents, making it clear that our current models of how DNA-damaging agents induce cytotoxicity are inadequate.

To help address this inadequacy, we have completed a genome-wide screen to assess the role of all nonessential *S. cerevisiae* proteins in modulating toxicity after exposure to four different DNA-damaging agents, each of which is a known carcinogen. Such genome-wide phenotypic screens are termed "genomic phenotyping" (Begley et al., 2002) and have been used to identify "toxicity-modulating" proteins, which may include proteins that prevent or repair cellular damage and proteins that generally affect growth advantage under stressful environmental conditions. Further, we have developed data-mapping techniques to globally integrate phenotypic information and develop a picture of the cell after carcinogen exposure. The systems toxicology method described below demonstrates that phenotypic data can be globally integrated with other information and used to help identify pathways that modulate toxicity.

Results and Discussion

We individually exposed 4,733 haploid *S. cerevisiae* gene deletion strains (the entire set of nonessential genes for this organism) to 4 DNA-damaging agents: the simple alkylating 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 agents represent a cross-section of the damaging agents that we are exposed to from environmental sources. Cells were exposed, in triplicate, to five doses of each agent, and the ability to recover (relative to similarly exposed wild-type) was monitored by digital imaging and computational analysis (Begley et al., 2002). The genomic phenotyping described here is distinct from most other published methods (Chang et al., 2002; Ross-Macdonald et al., 1999) in that strain growth is monitored individually

A

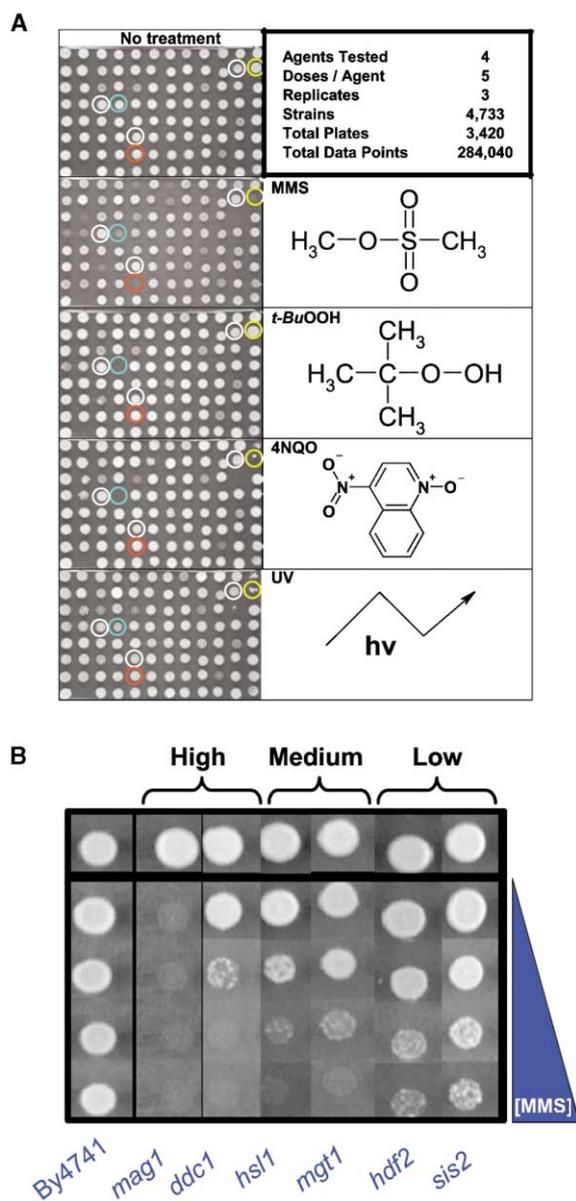|	
Agents Tested	4
Doses / Agent	5
Replicates	3
Strains	4,733
Total Plates	3,420
Total Data Points	284,040

B

Figure 1. Genomic Phenotyping with a Set of Yeast Gene Deletion Strains

(A) Imaged plates, damaging agents used, and study design. Five imaged plates that show the growth of different strains after no, MMS, *t*-BuOOH, 4NQO, or UV treatment are shown. Initially, 96-well plates were supplemented at 3 empty positions with the BY4741 parental strain (white circles) and at 3 positions with damage-sensitive controls, *mag1Δ* (red circles), *rad14* (yellow circles), and *erg6* (blue circles). After growth to stationary phase, 1 μl of each culture was robotically spotted onto agar plates and mock or agent treated; 60 hr later, growth on the plates was imaged. Experiments were performed at 5 doses per agent and in triplicate for 4,733 yeast gene deletion strains.

(B) Selected examples of strain-specific results after exposure to five increasing doses of MMS. Relative to BY4741 wild-type, some DNA repair (*mag1Δ*, *mgt1Δ*, and *hdf2Δ*) and cell cycle (*ddc1Δ*, *hsl1Δ*, and *sis2Δ*) mutants were classified as having either a high, medium, or low sensitivity to MMS.

over a wide range of doses and with multiple biological replicates (Figure 1). We thus set out to identify any and all proteins that modulate carcinogen-induced toxicity

in *S. cerevisiae*. It should be noted, though, that genomic phenotyping with single gene deletions could fail to identify a toxicity-modulating protein whose activity can be replaced by that of another protein (i.e., proteins with redundant function).

A total of 3,420 agar plates were spotted with ~96 different strains, and the strains were exposed to increasing concentrations of damaging agent. Every plate was spotted with three wild-type replicates and three known sensitive strains (*mag1Δ*, *rad14Δ*, and *erg6Δ*) in addition to the ~90 gene deletion strains to be tested; plates were imaged after 60 hr at 30°C (Figure 1A). The pixel intensity for each strain was quantified, and a simple metric identified strains that were more damage sensitive than wild-type on the same plate (Begley et al., 2002). The computed phenotypic classifications for each gene deletion strain were verified by visual inspection of the imaged plates. To facilitate visual analysis, plate images were cut into 96 pieces in silico and recompiled into virtual killing curves (as in Figure 1B) on 4,733 strain-specific web pages at genomicphenotyping. mit.edu/newpages/complete.html.

Analysis of our results indicates that 2,042 gene deletion strains recovered less well than wild-type following exposure to at least 1 of the 4 damaging agents used (Figure 2). Based on the number of sensitive strains, it is clear that MMS recovery was affected by the largest number of proteins (1,441), followed by 4NQO (819), *t*-BuOOH (447), and UV (288). Many strains are sensitive to more that one agent, but, interestingly, relatively few (27) are sensitive to all 4 agents; the overlap in agent sensitivity among the strains is shown in Figure 2A, and the identity of the genes is shown in Supplemental Table S1 (see the Supplemental Data). Remarkably, there are no DNA repair proteins among the 27 proteins that protect against all 4 agents; there are only 2 cell cycle checkpoint proteins, and the remaining proteins play roles in unknown pathways or in pathways not traditionally considered to be involved in the response of cells to DNA-damaging agents. Many of these pathways will be discussed below.

Note that by using a range of exposure doses in our genomic phenotyping screen, we identified strains with high, medium, and low sensitivity (Figures 1B and 2A). This may be why our screen turns out to be much more sensitive than others (Chang et al., 2002; Ross-Macdonald et al., 1999). Importantly, previously known damage-sensitive strains (e.g., DNA repair and checkpoint mutants) are found with high, medium, or low damage sensitivity, discounting the notion that only highly sensitive strains are of interest. In addition, we included in the current screen a repeat analysis of the 1,615 strains that were previously assessed for damage sensitivity (Begley et al., 2002); this analysis allowed us to determine that there is excellent reproducibility for the screen at ~90% (for details, see the Experimental Procedures). Our results also showed good overlap with other studies (Bennett et al., 2001; Chang et al., 2002; Ross-Macdonald et al., 1999); although, they found fewer sensitive strains, ~85% of the ones they did find were represented in our screen.

We have used data-mapping techniques coupled with statistical analysis to build a cell-based model detailing toxicity-modulating mechanisms after exposure to carcinogenic DNA-damaging agents. For this, we used in-

A

MMS t-BuOOH 4NQO UV

B

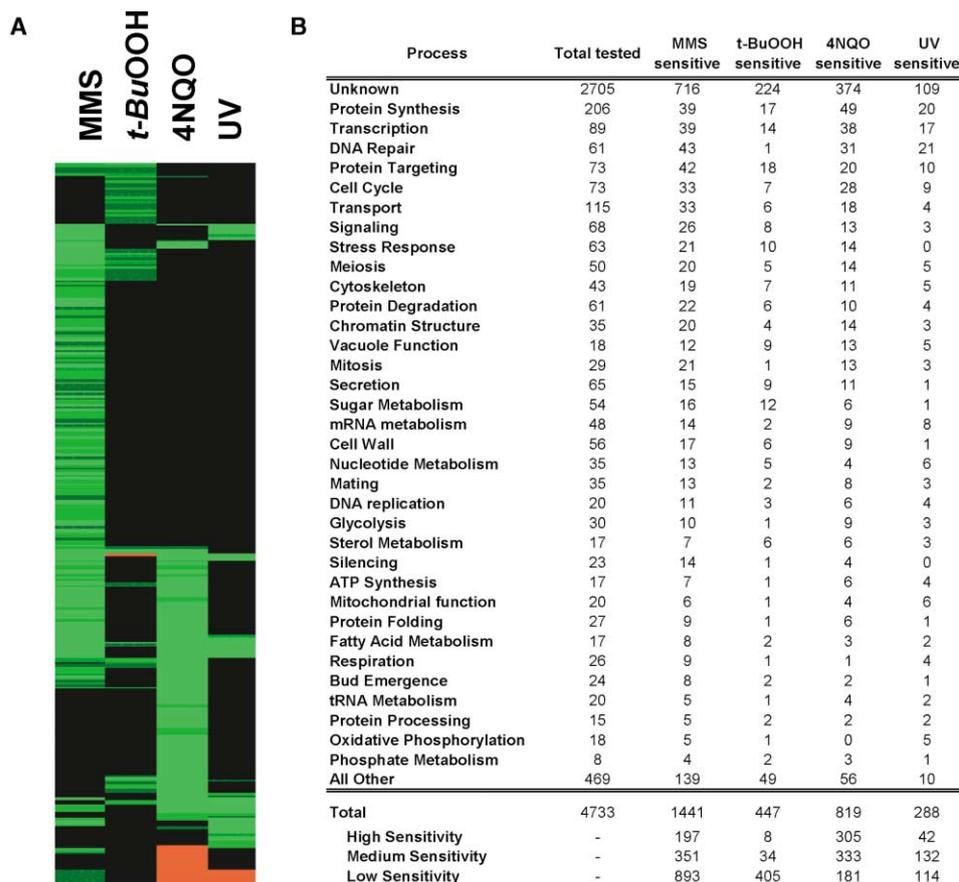| Process	Total tested	MMS sensitive	t-BuOOH sensitive	4NQO sensitive	UV sensitive
Unknown	2705	716	224	374	109
Protein Synthesis	206	39	17	49	20
Transcription	89	39	14	38	17
DNA Repair	61	43	1	31	21
Protein Targeting	73	42	18	20	10
Cell Cycle	73	33	7	28	9
Transport	115	33	6	18	4
Signaling	68	26	8	13	3
Stress Response	63	21	10	14	0
Meiosis	50	20	5	14	5
Cytoskeleton	43	19	7	11	5
Protein Degradation	61	22	6	10	4
Chromatin Structure	35	20	4	14	3
Vacuole Function	18	12	9	13	5
Mitosis	29	21	1	13	3
Secretion	65	15	9	11	1
Sugar Metabolism	54	16	12	6	1
mRNA metabolism	48	14	2	9	8
Cell Wall	56	17	6	9	1
Nucleotide Metabolism	35	13	5	4	6
Mating	35	13	2	8	3
DNA replication	20	11	3	6	4
Glycolysis	30	10	1	9	3
Sterol Metabolism	17	7	6	6	3
Silencing	23	14	1	4	0
ATP Synthesis	17	7	1	6	4
Mitochondrial function	20	6	1	4	6
Protein Folding	27	9	1	6	1
Fatty Acid Metabolism	17	8	2	3	2
Respiration	26	9	1	1	4
Bud Emergence	24	8	2	2	1
tRNA Metabolism	20	5	1	4	2
Protein Processing	15	5	2	2	2
Oxidative Phosphorylation	18	5	1	0	5
Phosphate Metabolism	8	4	2	3	1
All Other	469	139	49	56	10
Total	4733	1441	447	819	288
High Sensitivity	-	197	8	305	42
Medium Sensitivity	-	351	34	333	132
Low Sensitivity	-	893	405	181	114

Figure 2. Clustering and Classifications for Toxicity-Modulating Proteins

(A) Phenotypic values for all agent-sensitive or -resistant strains (detailed on website) were assigned based on sensitivity (green), resistance (red), and no phenotype (black); average linked clustering was performed as described (Eisen et al., 1998). The y axis corresponds to individual gene deletion strains, and the x axis indicates treatment.

(B) Cellular process associated with agent-sensitive phenotypes. Categories are based on the *Saccharomyces cerevesiae* genome database (SGD) and yeast protein database (YPD). In addition, the number of gene deletion strains falling into high, medium, and low categories of sensitivity is shown. On the website, each number is linked to the relevant strain list, ordered from the strongest to the weakest phenotype, and each strain on the list links to the webpage displaying its triplicate screening results.

formation from a study that assigned subcellular localizations to about 70% of the yeast proteome (Huh et al., 2003). This global localization study used GFP-tagged proteins and colocalization methods to assign 1 or more of 22 possible subcellular localizations to each of 4,156 proteins. It should be noted that both essential and nonessential proteins are among this group of 4,156, while our phenotypic data only apply to the nonessential proteins. We integrated the phenotypic data with subcellular localization information to determine the distribution of toxicity-modulating proteins among the 22 localizations and identify overpopulated areas. Since the overall fraction of toxicity-modulating proteins is quite high, it was important to determine whether an overpopulation of such proteins to a subcellular localization was significantly higher than that expected by chance. For each subcellular localization containing N nonessential proteins, we performed ~2000 random samplings of N proteins from the complete set of nonessential proteins to determine the average number of MMS, *t*-BuOOH, 4NQO, and UV toxicity-modulating proteins that would appear in each subcellular localization simply by chance. The toxicity-modulating proteins among the randomly col-

lected sets of proteins approximated a normal distribution (Supplemental Figure S1; see the Supplemental Data). The determination of whether the observed number of toxicity-modulating proteins in each subcellular localization was significantly higher than the mean number of such proteins found by random sampling was based on the Z-score and a one-tailed hypothesis test to specify a p-value. Toxicity-modulating proteins that were significantly overrepresented ($p \leq 0.05$) in a subcellular localization were then mapped en mass, generating cell-based images for hot spots of activity after toxicant exposure (Figure 3). For mapping, we set the p-value cutoff at 0.05, but it should be noted that only one overpopulated localization (nuclear periphery, 4NQO) was at the cutoff, and the rest were between 6.9E-14 and 0.045. In addition, Figures 3B–3F have been color coded (yellow) to show all the localizations that have p-values between 1.0E-4 and 0.05. To additionally corroborate significant results, the binomial distribution probability of all overpopulated localizations was determined (see the Experimental Procedures) (Devore, 2004).

This localization mapping strategy was first validated by using information about essential proteins (Figures
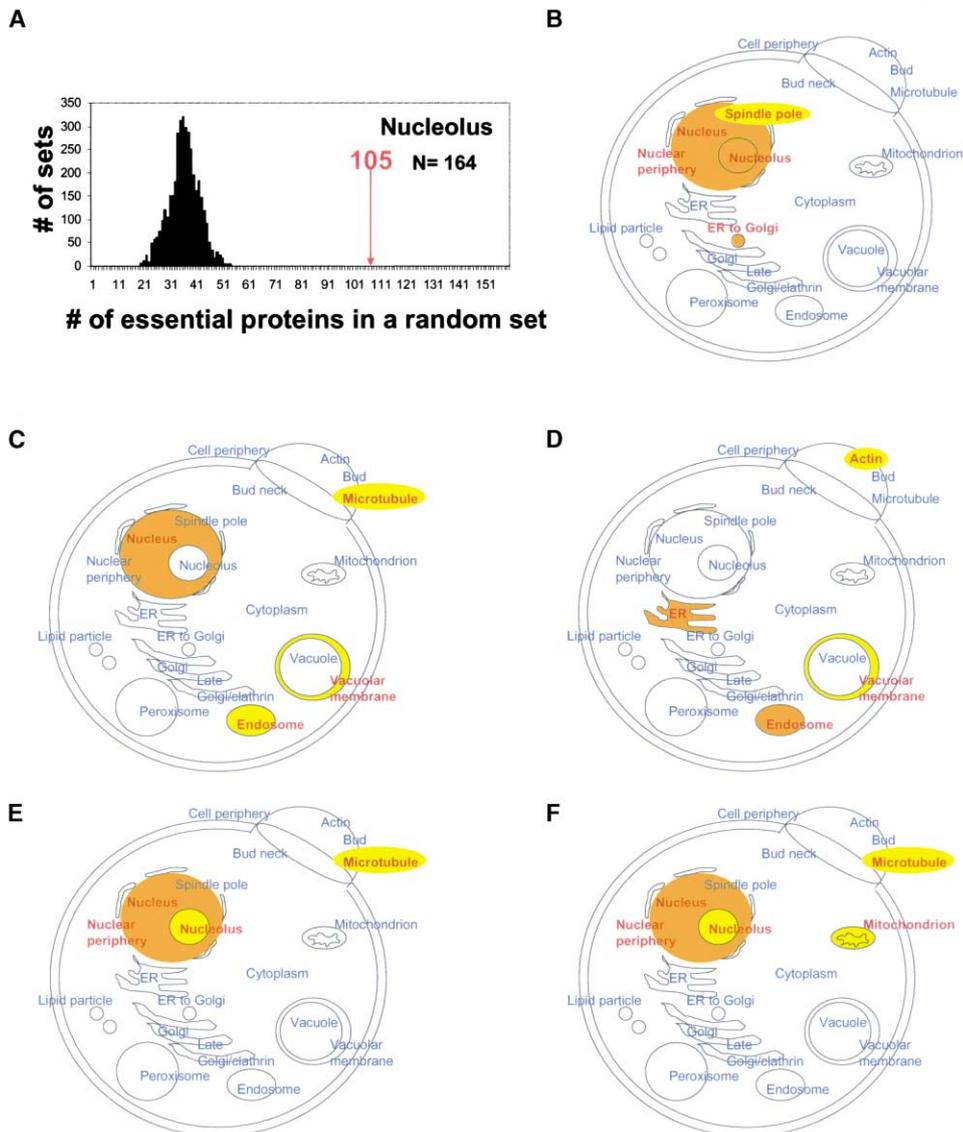
Figure 3. Hot Spots for Essential and Toxicity-Modulating Proteins

(A) A histogram shows the number of times X essential proteins were identified after randomly sampling ∼2000 groups of 164 proteins and indicates that nucleolus-localized proteins are overrepresented with essential proteins. A total of 105 proteins (the position in the histogram is shown with a red arrow) found in the nucleolus are essential, which is significantly increased (p < 6.9E-14) when compared to a random sampling of yeast proteins.

(B–F) Localizations colored in orange (p < 1.0 E-4) or yellow (p < 0.05) and labeled in red are overpopulated with essential or toxicity-modulating proteins and classified as hot spots. (B) Hot spots for essential proteins include the nucleolus, nucleus (458 proteins, p < 6.9E-14), spindle pole (39 proteins, p < 1.0E-11), ER to Golgi (5 proteins, p < 1.3E-03), and nuclear periphery (22 proteins, p < 8.2E-3). (C) Hot spots for MMS toxicity-modulating proteins include the nucleus (431 proteins, p < 1.1E-08), vacuolar membrane (30 proteins, p < 4.7E-03), microtubule (9 proteins, p < 3.6E-02), and the endosome (24 proteins, p < 4.5E-02). (D) Hot spots for t-BuOOH toxicity-modulating proteins include the endosome (17 proteins, p < 6.0E-09), ER (42 proteins, p < 5.4E-06), actin (6 proteins, p < 3.6E-02), and vacuolar membrane (10 proteins, p < 3.6E-03). (E) Hot spots for 4NQO toxicity-modulating proteins include the nucleus (256 proteins, p < 4.8E-05), microtubule (7 proteins, p < 1.0E-02), nucleolus (20 proteins, p < 1.4E-02), and nuclear periphery (13 proteins, p < 5.0E-02). (F) Hot spots for UV toxicity-modulating proteins include the nucleus (106 proteins, p < 7.9E-07), mitochondrion (49 proteins, p < 1.4E-02), microtubule (3 proteins, p < 2.9E-02), nuclear periphery (6 proteins, p < 2.8E-02), and the nucleolus (8 proteins, p < 3.6E-02).

3A and 3B). Essential genes/proteins were identified by the *S. cerevisiae* gene deletion consortium (Giaever et al., 2002), and it was established that roughly 20% of *S. cerevisiae* genes are essential. Huh et al. briefly noted that essential proteins are overrepresented in the nucleolus (Huh et al., 2003). Here, we confirm that essential proteins are indeed significantly overrepresented in the nucleolus (p < 6.9E–14) and also find that they are overrepresented in the nucleus, nuclear periphery, spindle pole, and "endoplasmic reticulum (ER) to Golgi," with p-values ranging from p < 0.008 to p < 6.9E–14 (Supplemental Table S2). Such localization of essential proteins was, for the most part, expected since all genetic transactions take place in the nucleus and nucleolus and

because the spindle pole helps accurately segregate chromosomes. The nuclear periphery is home to the nuclear pore complex that orchestrates the import and export of proteins and RNA to and from the nucleus, clearly important events for cell function. Surprisingly, the "ER to Golgi" localization contains a higher than expected number of essential proteins, including proteins required for functioning of the CopII complex (Dolinski et al., 2004), which orchestrates vesicle budding and the proper transport of cargo molecules to the Golgi (Haucke, 2003). Knowing the identity of cargo molecules would surely shed light on the essential "ER to Golgi" function. It is important to note that essential proteins reside in at least 20 of the 22 subcellular localizations and that hot spots simply highlight localizations that have a higher density of essential proteins than would be expected by chance. The same is true for the hot spots for toxicity-modulating proteins described below.

Localization mapping of the toxicity-modulating proteins identified by genomic phenotyping was used to help build an image of the important cellular responses that alleviate toxicity induced by carcinogenic DNA-damaging agents; such mapping identified expected and unexpected hot spots for toxicity modulation, and it is immediately obvious that the hot spot locations vary according to the type of damaging agent analyzed (Figures 3C–3F, Supplemental Tables S3–S6). DNA is a major target for the four DNA-damaging agents (Friedberg et al., 1995). Accordingly, the nucleus was overpopulated with MMS, 4NQO, and UV toxicity-modulating proteins, as one might expect for DNA repair and DNA damage-sensing proteins. However, it was not a hot spot for t-BuOOH toxicity-modulating proteins. This is not to say that no nuclear proteins provide t-BuOOH resistance, but rather that the nucleus is not a hot spot for such proteins. Since there is ample evidence that t-BuOOH can induce DNA damage (Mukherjee et al., 1995; Ochi, 1989), these results suggest either that DNA damage induced by this oxidizing agent is not cytotoxic, or that redundant DNA repair activities exist. Indeed, it is clear from $E.$ $coli$, humans, and mice that multiple redundant pathways exist for protecting against DNA oxidation (Michaels et al., 1992; Mo et al., 1992; Rosenquist et al., 1997; Slupska et al., 1996).

Toxicity-modulating hot spots were also identified in other localizations. Microtubule-associated proteins were overpopulated with MMS, 4NQO, and UV toxicity-modulating proteins, presumably because of their role in mitosis (Dolinski et al., 2004). The nucleolus and nuclear periphery were overpopulated by 4NQO and UV toxicity-modulating proteins, suggesting that nucleocytoplasmic transport of RNA and proteins, plus the biogenesis of new ribosomes, is relatively important for toxicity modulation. Finally, localization mapping analysis of toxicity-modulating proteins also classified the vacuolar membrane (for MMS and t-BuOOH), the endosome (MMS and t-BuOOH), actin (t-BuOOH), the ER (t-BuOOH), and mitochondria (UV) as being hot spots (Figures 3C–3F).

To better understand what happens at hot spots for toxicity modulation, we performed another layer of data integration to identify interacting proteins involved in preventing cell death after treatment with a carcinogen. Such interacting proteins were assembled from publicly available molecular interaction data (Lee et al., 2002;

Xenarios et al., 2002) (with both protein-protein and protein-DNA interactions) for all of the proteins located in subcellular localizations that emerged as hot spots for toxicity modulation (Figures 3C–3F); i.e., the actin, endosome, ER, microtubule, mitochondria, nuclear periphery, nucleolus, nucleus, and vacuolar membrane domains. The use of only molecular interaction between similarly localized proteins has the added benefit of reducing noise inherent to protein interaction data sets. The localization-specific molecular interaction information was merged with phenotypic data, generating phenotypically annotated localization-specific protein networks.

An example of localization mapping followed by protein subnetwork analysis for the nucleus is shown in Figure 4A (Supplemental Figures S2–S5). The nucleus was identified as a hot spot based on localization mapping. The 1,454 nuclear proteins were assembled into a network of 1,248 proteins connected by 2,918 protein-protein interactions and 1,079 protein-DNA interactions. Next, all essential and no-phenotype proteins were removed from the subnetworks found in the nucleus, via a filtering step, and we derived a large, connected subnetwork of 252 proteins, all of which confer resistance to MMS-induced toxicity. While this was the largest subnetwork to be identified, similar subnetworks were sought for each hot spot in Figures 3C–3F; filtered nuclear subnetworks are displayed in Figures 4A–4C, and all phenotypically annotated localization-specific protein networks are found in Supplemental Figures S2–S5, with annotated proteins.

Nuclear subnetworks for MMS, 4NQO, and UV (Figures 4A–4C) were comprised, respectively, of 272, 147, and 41 proteins connected to at least 1 other toxicity-modulating protein. Many of the proteins in these subnetworks are actually involved in damage recovery at the DNA level. For each agent, 1 large, connected subnetwork emerged containing 252 proteins for MMS recovery, 131 for 4NQO recovery, and 22 for UV –recovery, all with p $<$ 1.0E-5. The nucleus contains DNA repair pathways to process damage, and so it was no surprise to find such proteins represented in all 3 toxicity-modulating nuclear subnetworks; 27 were found for MMS, 26 were found for NQO, and 11 were found for UV. However, it was surprising to find roughly the same number of proteins involved in transcription embedded in these subnetworks; 27 were found for MMS, 26 were found for 4NQO, and 9 were found for UV. None of these transcription proteins are known to participate in transcription-coupled DNA repair; i.e., they are not part of the TFIIH RNA Polymerase II initiation complex (Friedberg et al., 1995). Instead, these toxicity-modulating proteins comprise components of the RNA polymerase II mediator complex and the Swi/Snf global transcription activator complex; plus, several others are also associated with RNA polymerase II-associated proteins. It should be noted that many of the strains missing RNA polymerase II-associated proteins had high sensitivity to specific agents (see Figure 1B), and some were sensitive to all four DNA-damaging agents (Snf6 and Rpb4). We have thus uncovered hitherto unknown transcription-related pathways that play an enormous role in helping cells recover from the damage inflicted by a variety of carcinogenic agents.
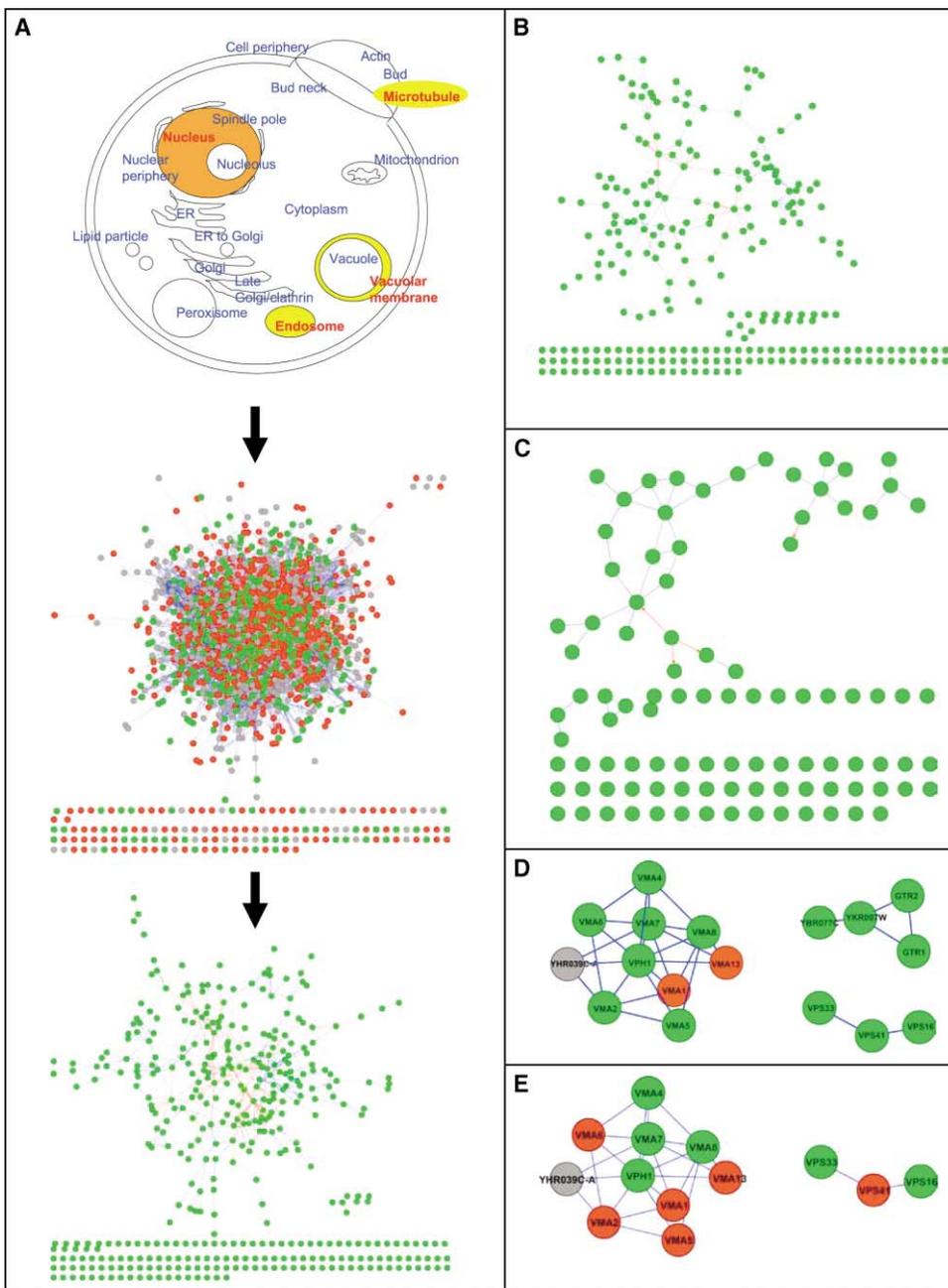
Figure 4. Localization-Based Protein Subnetworks that Modulate Toxicity

(A) Subcellular localizations that were overpopulated with toxicity-modulating proteins were identified (top), molecular interactions connecting proteins belonging to this localization were compiled (middle), and the resulting protein subnetwork (bottom) was filtered to identify connected groups of toxicity-modulating proteins. Visualization was performed by using the program Cytoscape (Shannon et al., 2003) (www.cytoscape.org). The example shown represents the analysis of MMS toxicity-modulating proteins (green circles) found in the nucleus ($p < 2.6E-12$). Protein-protein interactions are represented as blue lines, and protein-DNA interactions are represented as red arrows.

(B–D) Nuclear subnetworks of (B) 4NQO toxicity-modulating ($p < 1.3E-5$) and (C) UV toxicity-modulating proteins ($p < 9.9E-8$) were also identified. Significantly enriched subnetworks found in the vacuolar membrane representing (D) MMS toxicity-modulating ($p < 0.02$, $p < 0.02$, and $p < 0.05$ for the 10, 4, and 3 protein subnetworks, respectively) and (E) t-BuOOH toxicity-modulating proteins ($p < 0.002$ and $p < 0.05$ for the 10 and 3 protein subnetworks, respectively). All protein names can be found in the Supplemental Figures. Green circles represent toxicity-modulating proteins, gray circles represent essential proteins, and red circles represent proteins whose corresponding gene deletion strains display no phenotype. Blue lines represent protein-protein interactions, and red arrows represent protein-DNA interactions.

In the nuclear protein subnetworks for MMS, 4NQO, and UV (Figures 4A–4C), proteins for DNA repair, cell cycle checkpoints (Friedberg et al., 1995), and transcription (Dolinski et al., 2004) (among others) are linked to one another (Supplemental Figures S2A–S5A), and each one contributes to toxicity modulation. Juxtaposition of such proteins is even seen for the smallest subnetwork, representing UV toxicity-modulating proteins (Figure 4C

and Supplemental Figure S5A), where the cell cycle-specific Swi6 transcription factor interacts with the DNA damage-inducible Dun1 kinase, which in turn interacts with DNA repair and cell cycle checkpoint proteins (Rad9, Rad24, and Rad1). Thus, DNA repair pathways are linked to the signal transduction apparatus, suggesting that integrated response modules allow the processing of DNA damage to signal to the rest of the cell via kinase cascades and altered transcription. Indeed, proteins involved in remodeling chromatin structure are abundant among the nuclear toxicity-modulating subnetworks, with 16 such proteins for MMS, 10 for 4NQO, and 3 for UV (Figures 4A–4C and Supplemental Figures S2A and S4A–S5A). The chromatin assembly proteins, histone acetyltransferases, and deacetylases represented here may participate in reprogramming gene expression in response to damage exposure.

We also explored the non-nuclear protein complexes that modulate toxicity. For the vacuolar membrane, we see connections between 19 of the 60 proteins localized to this compartment, via 31 protein-protein interactions (Supplemental Figures S2B and S3A). We thus identified significant ($p < 0.02$) connected groups of MMS and t-BuOOH toxicity-modulating proteins that comprise part of a vacuolar H+-ATPase (Vma2, Vma4, Vma5, Vma6, Vma7, Vma8, and Vph1) (Dolinski et al., 2004; Forgac, 1999), that comprise part of a complex that plays a role in the Ran/Gsp1p GTPase cycle (Gtr1 and Gtr2; $p < 0.02$), and that comprise a group of proteins involved in vacuolar protein sorting (Vps33, Vps41, and Vps16; $p < 0.05$) (Figures 4D and 4E; Supplemental Figures S2 and S3) (Dolinski et al., 2004); each protein is important for vacuolar function. The H+-ATPase acidifies vacuoles and facilitates the transport of ions and small metabolites into the vacuole, and the acidic environment promotes proteolysis by vacuolar peptidases (Dolinski et al., 2004; Stevens and Forgac, 1997) as well as the breakdown of other macromolecules to be removed from the cell. Models for vacuole function after MMS and t-BuOOH treatment include: (i) degradation and disposal of damaged macromolecules to avoid cellular dysfunction; and (ii) sequestration of damaging agents to lower the effective exposure dose. These models are not mutually exclusive, and either one could account for the importance of vacuolar membrane proteins in preventing carcinogen-induced cell death.

The endosome also plays a role in modulating MMS and t-BuOOH toxicity (Figures 3C and 3D; Supplemental Figures S2 and S3). Note that identified endosomal toxicity-modulating proteins (Vps35, Vps5, Pep8, Pep3, Vps8, Vps27, and Yhl002w) are directly linked to the vacuole (Dolinski et al., 2004), and it appears that the vacuole and endosome preferentially interact with each other (Huh et al., 2003). Some of the endosome toxicity-modulating subnetworks help to sort and deliver proteins to the vacuole (Supplemental Figures S2C and S3B). Identified protein subnetworks ($p < 0.06$) in the endosome include those defined by Pep3, Vps8, and Pep5 in addition to Vps5, Vps7, Vps35, and Pep8. Strains deficient in Pep3, Pep8, Vps5, Vps8, or Vps35 have sorting defects for delivery of degredative enzymes to the vacuole (Chen and Stevens, 1996; Dolinski et al., 2004; Paravicini et al., 1992). Again, the inability to degrade damaged cellular constituents decreased viability after

treatment with carcinogens, and we propose that the vacuole is more important in the response of cells to carcinogenic agents than previously realized.

We have used genomic phenotyping in combination with mapping techniques to identify localization hot spots and protein subnetworks that modulate agent toxicity. We have, for the most part, interpreted these results to suggest that cellular responses represented in hot spots and subnetworks play an important role in either preventing or repairing cellular damage induced by the applied agents. Another plausible interpretation is that some of the identified proteins are playing a more indirect role, by optimizing cell growth under conditions of stress. It could be argued that most genes have a role in optimizing cell growth based on the principals of evolutionary selection. The growth defect observed in some of the damage-sensitive strains could certainly be attributed to fitness deficiencies under stressful conditions, as opposed to damage recovery. We suggest that fitness deficiencies are likely to be more prevalent among gene deletion strains that are sensitive to multiple agents.

In conclusion, we have developed a localization mapping method that uses three diverse global data sets, namely, genomic phenotyping to identify toxicity-modulating proteins, subcellular localization of proteins, and molecular interaction data to generate global models of cellular responses to carcinogenic agents. The integrated data identified statistically validated toxicity-modulating hot spots in the cell and identified a number of interacting protein subnetworks residing in these hot spot locations. The reported analysis used subcellular localization data obtained under basal conditions. However, it should be noted that reshuffling of proteins may occur upon treatment with a damaging agent, and this is currently under investigation. In all, though, the data integration method reported has highlighted a number of unexpected pathways that play important roles in modulating cellular toxicity after treatment with a damaging agent. The approach is now ripe for screening clinically relevant pharmaceuticals and has the potential to provide insight into the mechanism of action for many different compounds.

Experimental Procedures

Genomic Phenotyping and Database Construction

S. cerevisiae strain BY4741 and deletion mutant derivatives were supplied by Research Genetics. Parental strain BY4741 was transformed with plasmid pYE13g (American Type Culture Collection) and selected for on YPD (10 g yeast extract, 20 g peptone, 20 g dextrose, 20 g agar/liter) containing 200 μg/ml G418. Genomic phenotyping was performed as previously described (Begley et al., 2002), with some modifications. Briefly, 96-well master plates containing individual deletion strains were supplemented with the agent-sensitive controls *mag1Δ*, *rad14*, and *erg6* and grown in 150 μl of YPD, containing G418 at 200 μg/ml. 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), and then using the Hydra, 1 μl samples were spotted simultaneously onto an agar-containing plate. MMS, *t*-BuOOH, and 4NQO were purchased from Aldrich. UV radiation (254 nm) was supplied from a UV Stratalinker 2400 (Stratagene). Plates containing up to 96 strains were tested under the following conditions: no treatment, 0.01% MMS, 0.02% MMS, 0.025% MMS, 0.03% MMS, 0.50 mM *t*-BuOOH, 0.75 mM *t*-BuOOH, 1.0 mM *t*-BuOOH, 1.25 mM *t*-BuOOH,

0.2 μg/ml 4NQO, 0.3 μg/ml 4NQO, 04 μg/ml 4NQO, 0.5 μg/ml 4NQO, 40 J/m² UV, 80 J/m² UV, 100 J/m² UV, and 125 J/m² UV. The maximum dose of each agent was selected to induce 10% killing to the wild-type strain. Strains were grown for 60 hr at 30°C and then imaged with a Gel Doc 1000 from BioRad running Quantity One software. Images were analyzed with ScanAlyze software to quantitate the pixel intensity of each spotted colony. All screens were performed in triplicate with fresh liquid cultures. The genomic phenotyping database (genomicphenotyping.mit.edu) was constructed as described (Begley et al., 2002). Representative sensitivity values were generated by using a scoring scheme that allocated values of 4, 3, 2, or 1 depending on the concentration of agent when strain sensitivity was identified; 4 is allocated to the lowest, and 1 is allocated to the highest concentration of damaging agents. These values were allowed to accumulate in each replicate, and then they were summed across all replicates. For example, in replicate 1, strains sensitive to all concentrations of agents received a score of 10 (4 + 3 + 2 + 1), and this was summed over all 3 replicates for a final score of 30 (10 + 10 + 10). Damage-sensitive strains had scores that ranged from 30 (most sensitive) to 2 (least sensitive).

We have also analyzed our data in the context of a partial data set performed previously in our lab containing 1615 overlapping strains, covering 4 identical damaging agents, and containing a total of 6460 (i.e., 1615 × 4) overlapping data points. Upon analysis, we determined that out of the 6460 overlapping data points, 5,615 of the individual comparisons agreed (87%) (both sensitive or both no-phenotype), while 845 differed. Similarly, a comparison to data found in the public domain (Bennett et al., 2001; Chang et al., 2002; Ross-Macdonald et al., 1999) indicates that there are 5,788 overlapping data points (i.e., similar strain and damaging agent, but damaging agent at different concentration). Upon comparison to data reported here, we find that 4160 data points were in agreement, while 1628 differed. Closer inspection of the Chang et al. data, which monitored MMS sensitivity across the same BY4741 strain used in this study, shows that they classified 103 MMS-sensitive strains, of which we found 87 to be MMS sensitive.

Localization Mapping and Subnetwork Analysis
Localization data were generated by O'Shea and colleagues (Huh et al., 2003) and integrated with our genomic phenotyping data. A complete list can be downloaded at the genomic phenotyping web site. The total number (N) of proteins in each subcellular localization was provided, and the number of essential (E_N) and toxicity-modulating (T_N) proteins in each subcellular localization was determined by using our genomic phenotyping data and an exhaustive search algorithm. Proteins without subcellular localization information were removed from the rest of our analysis. Corresponding to each of the 22 subcellular localizations specified in Supplemental Tables S2–S6, a random sampling of N proteins was performed ~2000 times. The random sampling algorithm was written in visual basic, and for each sampling, the number of essential MMS, t-BuOOH, 4NQO, and UV toxicity-modulating proteins was recorded. From our random samplings, the average number of essential (AE_N) or toxicity-modulating (AT_N) for each N group of randomly sampled proteins was determined, along with the standard deviation (for toxicity-modulating proteins [SD_{N-T}] and for essential proteins [SD_{N-E}]). It should be noted that for the identification of toxicity-modulating hot spots, random sampling was restricted to nonessential proteins. In addition, the actual number of toxicity-modulating proteins (shown in red, Supplemental Figure S1) in a subcellular localization and the distribution of random samplings containing a specific number of toxicity-modulating proteins (T_N) are shown (Supplemental Figure S1). It should be noted that the distribution details the number of times X proteins were counted as being toxicity modulating in our random sampling (see Formula 1).

$$Z = \frac{(T_N - AT_N)}{SD_{N-T}} \quad \text{or} \quad \frac{(E_N - AE_N)}{SD_{N-E}} \qquad (1)$$

Associated p-values were determined for all overrepresented localizations by using a one-tailed test and a directional hypothesis that states that each subcellular localization is not overrepresented with toxicity-modulating or essential proteins. In addition, by using Microsoft Excel, the binomial distribution probability for all hot spots

was determined to be significant and was used, as opposed to a hypergeometric distribution, due to sample size.

Protein-protein and protein-DNA interactions were integrated and visualized with our genomic phenotyping data and subcellular localization data by using filtering methods found in the visualization program Cytoscape (found at www.cytoscape.org). First, the localization-specific molecular interaction information was merged with phenotypic data generating phenotypically annotated localization-specific protein networks. For visualization purposes, large networks found in the nucleus, and shown in Figure 4, were filtered to remove all essential and no-phenotype proteins. The significance for toxicity modulation of a localization-specific protein subnetwork containing N nonessential proteins was determined based on a random sampling of N proteins (from the total sample space of 3185 proteins) to determine the number of randomly selected toxicity-modulating proteins. This was performed 2000 times to generate average (AT_N) and standard deviation (SD_{N-T}) values. The number of toxicity-modulating (T_N) proteins found in a localization-specific protein subnetwork was then compared to the average (ADR_N) for a sample of the same size, and a Z-score and p-value were determined as described above.

Supplemental Data
Supplemental Data including Supplemental Figures S1–S5 and Supplemental Tables S1–S6 are available at http://www.molecule.org/cgi/content/full/16/1/117/DC1/.

References

Begley, T.J., Rosenbach, A.S., Ideker, T., and Samson, L.D. (2002). Recovery pathways in *S. cerevisiae* revealed by genomic phenotyping and interactome mapping. Mol. Cancer Res. *1*, 103–112.

Bennett, C.B., Lewis, L.K., Karthikeyan, G., Lobachev, K.S., Jin, Y.H., Sterling, J.F., Snipe, J.R., and Resnick, M.A. (2001). Genes required for ionizing radiation resistance in yeast. Nat. Genet. *29*, 426–434.

Chang, M., Bellaoui, M., Boone, C., and Brown, G.W. (2002). A genome-wide screen for methyl methanesulfonate-sensitive mutants reveals genes required for S phase progression in the presence of DNA damage. Proc. Natl. Acad. Sci. USA *99*, 16934–16939.

Chen, Y.J., and Stevens, T.H. (1996). The VPS8 gene is required for localization and trafficking of the CPY sorting receptor in *Saccharomyces cerevisiae*. Eur. J. Cell Biol. *70*, 289–297.

Devore, J.L. (2004). Probability and Statistics, Sixth Edition (Belmont, CA: Brooks/Cole-Thomason Learning).

Dolinski, K., Balakrishnan, R., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R., Fisk, D.G., Hirschman, J.E., Hong, E.L., Issel-Tarver, L., et al. (2004). *Saccharomyces* Genome Database (http://www.yeastgenome.org/).

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA *95*, 14863–14868.

Elledge, S.J. (1996). Cell cycle checkpoints: preventing an identity crisis. Science *274*, 1664–1672.

Forgac, M. (1999). Structure and properties of the vacuolar (H+)-ATPases. J. Biol. Chem. *274*, 12951–12954.

Friedberg, E.C., Walker, G.C., and Siede, W. (1995). DNA Repair and

Mutagenesis (Washington, D.C.: American Society for Microbiology Press).

Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., Andre, B., et al. (2002). Functional profiling of the *Saccharomyces cerevisiae* genome. Nature *418*, 387–391.

Haucke, V. (2003). Vesicle budding: a coat for the COPs. Trends Cell Biol. *13*, 59–60.

Hughes, T.R., Marton, M.J., Jones, A.R., Roberts, C.J., Stoughton, R., Armour, C.D., Bennett, H.A., Coffey, E., Dai, H., He, Y.D., et al. (2000). Functional discovery via a compendium of expression profiles. Cell *102*, 109–126.

Huh, W.K., Falvo, J.V., Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S., and O'Shea, E.K. (2003). Global analysis of protein localization in budding yeast. Nature *425*, 686–691.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc. Natl. Acad. Sci. USA *98*, 4569–4574.

Jelinsky, S.A., Estep, P., Church, G.M., and Samson, L.D. (2000). Regulatory networks revealed by transcriptional profiling of damaged *Saccharomyces cerevisiae* cells: rpn4 links base excision repair with proteasomes. Mol. Cell. Biol. *20*, 8157–8167.

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

Marenstein, D.R., Ocampo, M.T., Chan, M.K., Altamirano, A., Basu, A.K., Boorstein, R.J., Cunningham, R.P., and Teebor, G.W. (2001). Stimulation of human endonuclease III by Y box-binding protein 1 (DNA-binding protein B). Interaction between a base excision repair enzyme and a transcription factor. J. Biol. Chem. *276*, 21242–21249.

Michaels, M.L., Cruz, C., Grollman, A.P., and Miller, J.H. (1992). Evidence that MutY and MutM combine to prevent mutations by an oxidatively damaged form of guanine in DNA. Proc. Natl. Acad. Sci. USA *89*, 7022–7025.

Mo, J.Y., Maki, H., and Sekiguchi, M. (1992). Hydrolytic elimination of a mutagenic nucleotide, 8-oxodGTP, by human 18-kilodalton protein: sanitization of nucleotide pool. Proc. Natl. Acad. Sci. USA *89*, 11021–11025.

Mukherjee, S.K., Yasharel, R., Klaidman, L.K., Hutchin, T.P., and Adams, J.D., Jr. (1995). Apoptosis and DNA fragmentation as induced by tertiary butylhydroperoxide in the brain. Brain Res. Bull. *38*, 595–604.

Navas, T.A., Sanchez, Y., and Elledge, S.J. (1996). RAD9 and DNA polymerase epsilon form parallel sensory branches for transducing the DNA damage checkpoint signal in *Saccharomyces cerevisiae*. Genes Dev. *10*, 2632–2643.

Ochi, T. (1989). Effects of iron chelators and glutathione depletion on the induction and repair of chromosomal aberrations by tert-butyl hydroperoxide in cultured Chinese hamster cells. Mutat. Res. *213*, 243–248.

Paravicini, G., Horazdovsky, B.F., and Emr, S.D. (1992). Alternative pathways for the sorting of soluble vacuolar proteins in yeast: a vps35 null mutant missorts and secretes only a subset of vacuolar hydrolases. Mol. Biol. Cell *3*, 415–427.

Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., et al. (2000). Genome-wide location and function of DNA binding proteins. Science *290*, 2306–2309.

Rosenquist, T.A., Zharkov, D.O., and Grollman, A.P. (1997). Cloning and characterization of a mammalian 8-oxoguanine DNA glycosylase. Proc. Natl. Acad. Sci. USA *94*, 7429–7434.

Ross-Macdonald, P., Coelho, P.S., Roemer, T., Agarwal, S., Kumar, A., Jansen, R., Cheung, K.H., Sheehan, A., Symoniatis, D., Umansky, L., et al. (1999). Large-scale analysis of the yeast genome by transposon tagging and gene disruption. Nature *402*, 413–418.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

Slupska, M.M., Baiklov, C., Luther, W.M., Chiang, J.-h., Wei, Y.-F., and Miller, J.H. (1996). Cloning and sequencing a human homolog (*hMYH*) of the *Escherichia coli mutY* gene whose function is required for the repair of oxidative DNA damage. J. Bacteriol. *178*, 3885–3892.

Stevens, T.H., and Forgac, M. (1997). Structure, function and regulation of the vacuolar (H+)-ATPase. Annu. Rev. Cell Dev. Biol. *13*, 779–808.

Tong, A.H., Evangelista, M., Parsons, A.B., Xu, H., Bader, G.D., Page, N., Robinson, M., Raghibizadeh, S., Hogue, C.W., Bussey, H., et al. (2001). Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science *294*, 2364–2368.

Tong, A.H., Lesage, G., Bader, G.D., Ding, H., Xu, H., Xin, X., Young, J., Berriz, G.F., Brost, R.L., Chang, M., et al. (2004). Global mapping of the yeast genetic interaction network. Science *303*, 808–813.

Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., et al. (2000). A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. Nature *403*, 623–627.

Winzeler, E.A., Shoemaker, D.D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J.D., Bussey, H., et al. (1999). Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. Science *285*, 901–906.

Xenarios, I., Salwinski, L., Duan, X.J., Higney, P., Kim, S.M., and Eisenberg, D. (2002). DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. Nucleic Acids Res. *30*, 303–305.

Zhou, Z., and Elledge, S.J. (1993). DUN1 encodes a protein kinase that controls the DNA damage response in yeast. Cell *75*, 1119–1127.