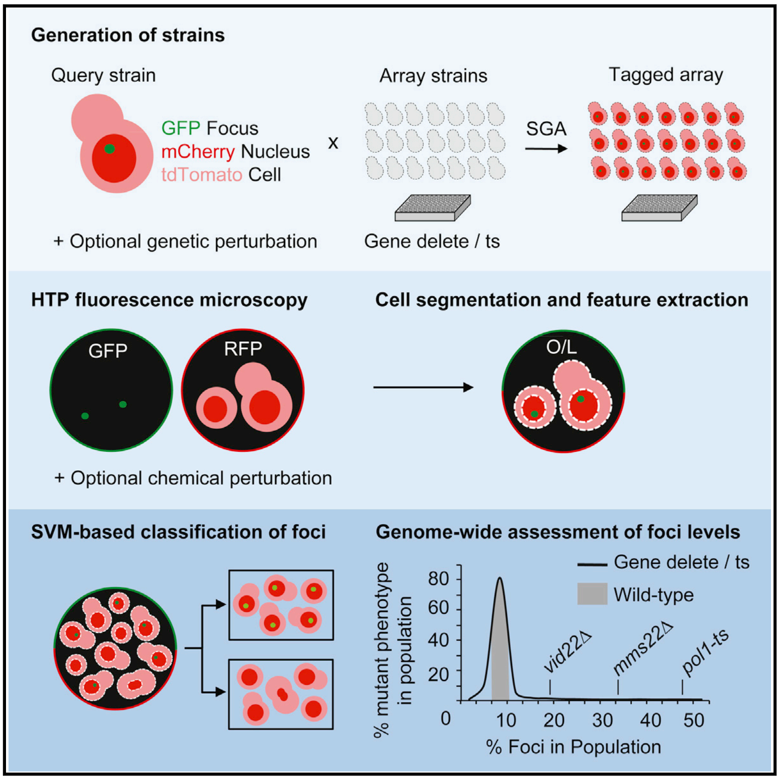
Cell Systems

Article

Exploring Quantitative Yeast Phenomics with Single-Cell Analysis of DNA Damage Foci

Graphical Abstract



Authors

Erin B. Styles, Karen J. Founk, Lee A. Zamparo, ..., Zhaolei Zhang, Charles Boone, Brenda J. Andrews

Correspondence

charlie.boone@utoronto.ca (C.B.), brenda.andrews@utoronto.ca (B.J.A.)

In Brief

Using synthetic genetic array technology, high-throughput imaging, and automated analysis, we identified 345 deletion and temperature-sensitive yeast mutants with elevated levels of DNA damage foci, as single mutants or in combination with chemical or genetic perturbations.

Highlights

- Automated genetics screen via SGA, HTP imaging, and machine learning classification
- Genome-wide screening in the context of genetic and/or chemical perturbations
- Roster of 345 mutants with elevated levels of DNA damage foci
- Sgs1-dependent function for *VID22* at DSBs and separate role at G-quadruplex regions

Styles et al., 2016, Cell Systems *3*, 264–277
September 28, 2016 © 2016 Elsevier Inc.
http://dx.doi.org/10.1016/j.cels.2016.08.008

Cell Systems
Article

Exploring Quantitative Yeast Phenomics with Single-Cell Analysis of DNA Damage Foci

Erin B. Styles,[1,2] Karen J. Founk,[1,2] Lee A. Zamparo,[1,2,3] Tina L. Sing,[1,4] Dogus Altintas,[5] Cyril Ribeyre,[5,11] Virginie Ribaud,[5,12] Jacques Rougemont,[6] David Mayhew,[7] Michael Costanzo,[1] Matej Usaj,[1] Adrian J. Verster,[1] Elizabeth N. Koch,[8] Daniele Novarina,[9] Marco Graf,[10] Brian Luke,[10] Marco Muzi-Falconi,[9] Chad L. Myers,[8] Robi David Mitra,[7] David Shore,[5] Grant W. Brown,[1,4] Zhaolei Zhang,[1,2] Charles Boone,[1,2,*] and Brenda J. Andrews[1,2,13,*]
[1]The Donnelly Centre
[2]Department of Molecular Genetics
[3]Department of Computer Sciences
[4]Department of Biochemistry
University of Toronto, Toronto, ON M5S 3E1, Canada
[5]Department of Molecular Biology, NCCR Program "Frontiers in Genetics", Institute of Genetics, Genomics, Geneva (iGE3), University of Geneva, 30, quai Ernest-Ansermet, 1211 Geneva 4, Switzerland
[6]Laboratory of Computational Systems Biology, Ecole Polytéchnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland
[7]Department of Genetics and Center for Genome Sciences and Systems Biology, Washington University School of Medicine in St. Louis, St. Louis, MO 63108, USA
[8]Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN 55455, USA
[9]Dipartimento di Bioscienze, Universita' degli Studi di Milano, 20122 Milano, Italy
[10]Institute of Molecular Biology (IMB), Ackermannweg 4, Mainz 55128, Germany
[11]Present address: Institute of Human Genetics, CNRS UPR 1142, 141 rue de la Cardonille, 34396 Montpellier Cedex 5, France
[12]Present address: Centre Medicale Universitaire, University of Geneva, rue Michel-Servet 1, 1211 Geneva 4, Switzerland
[13]Lead Contact
*Correspondence: charlie.boone@utoronto.ca (C.B.), brenda.andrews@utoronto.ca (B.J.A.)
http://dx.doi.org/10.1016/j.cels.2016.08.008

SUMMARY

A significant challenge of functional genomics is to develop methods for genome-scale acquisition and analysis of cell biological data. Here, we present an integrated method that combines genome-wide genetic perturbation of *Saccharomyces cerevisiae* with high-content screening to facilitate the genetic description of sub-cellular structures and compartment morphology. As proof of principle, we used a Rad52-GFP marker to examine DNA damage foci in ∼20 million single cells from ∼5,000 different mutant backgrounds in the context of selected genetic or chemical perturbations. Phenotypes were classified using a machine learning-based automated image analysis pipeline. 345 mutants were identified that had elevated numbers of DNA damage foci, almost half of which were identified only in sensitized backgrounds. Subsequent analysis of Vid22, a protein implicated in the DNA damage response, revealed that it acts together with the Sgs1 helicase at sites of DNA damage and preferentially binds G-quadruplex regions of the genome. This approach is extensible to numerous other cell biological markers and experimental systems.

INTRODUCTION

A fundamental goal of functional genomics is to systematically define gene function and cellular pathways. In the budding yeast *Saccharomyces cerevisiae*, genome-wide collections of haploid viable deletion mutants (Giaever et al., 2002; Winzeler et al., 1999) and mutant strains carrying conditional alleles of essential genes (Kofoed et al., 2015; Li et al., 2011; Mnaimneh et al., 2004) enable systematic genetic analysis. Due to ease of measurement and amenability to high-throughput (HTP) applications, most genome-scale studies have focused on cell fitness as a phenotypic readout (Baryshnikova et al., 2010; Collins et al., 2007). Notably, colony size, the ultimate consequence of repeated cell growth and division, has been used to examine the fitness phenotype of millions of double-mutant gene pairs to produce a global yeast genetic interaction network (Costanzo et al., 2010). Despite the information-rich nature of fitness assays, it is clear that the analysis of more subtle and specific phenotypes will yield important new functional information. For example, while ∼10% of the nonessential yeast deletion mutants show a clear fitness defect, nearly 50% exhibit a number of different morphological defects (Ohya et al., 2005). Thus, comprehensive understanding of gene function and genetic interaction networks will require further analysis of more complex phenotypes in a variety of conditions to reveal a complete functional wiring diagram of the cell.

In the past decade, systematic assessment of subcellular spatiotemporal phenotypes using high-content screening

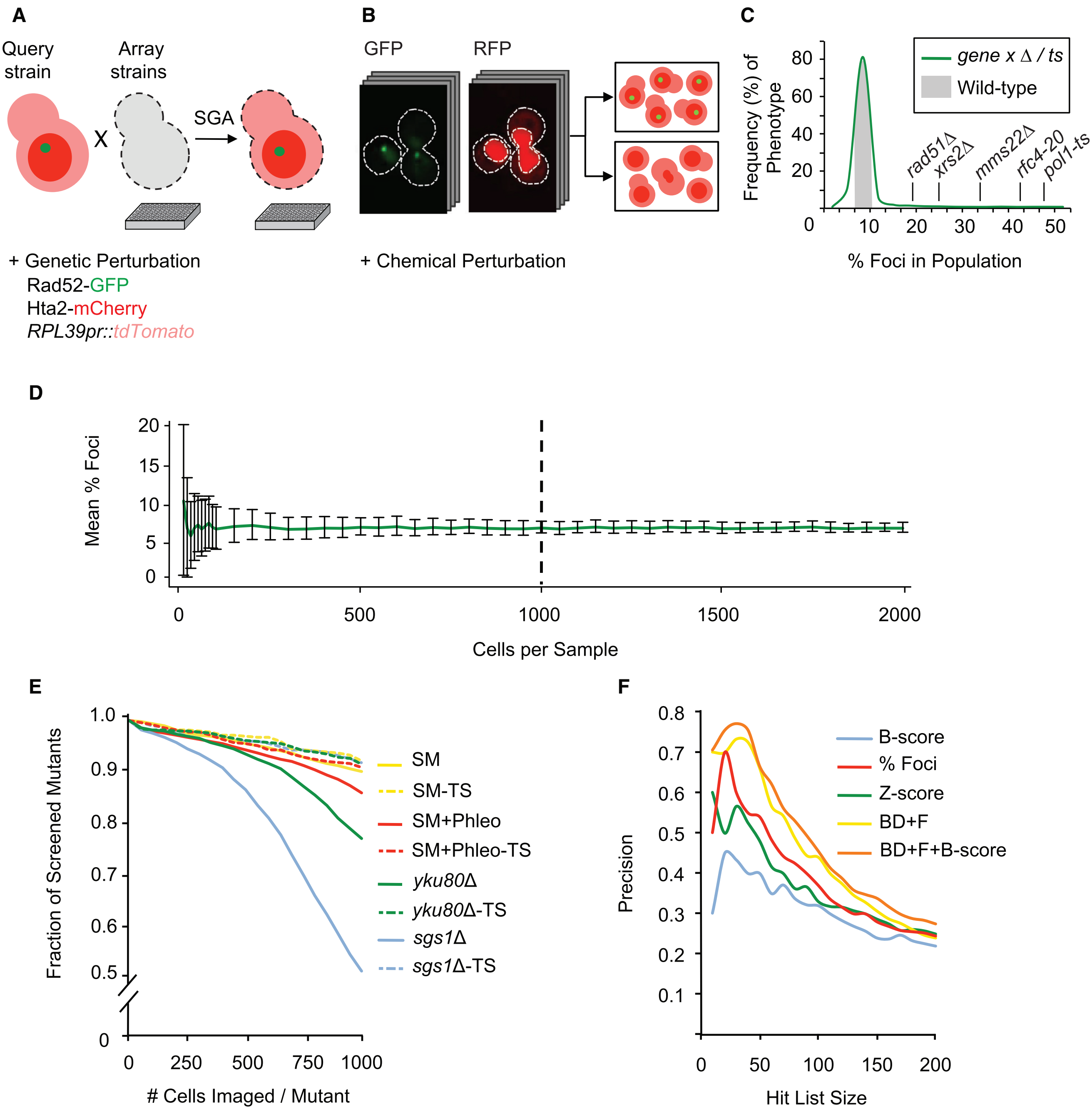


Figure 1. Synthetic Genetic Array-High Content Screening Strategy for Identifying Cell Populations with Elevated Levels of DNA Damage Foci

(A) Diagram illustrating array construction strategy for automated image analysis of fluorescent proteins marking specific compartments within the cell. A *RAD52-GFP* fusion gene product marks DNA damage foci (green dot), while nuclear (*HTA2-mCherry*; dark red) and cytoplasmic (*RPL39*pr-*tdTomato*; light red) signals provide spatial and cell-cycle context. A sensitizing gene deletion can be introduced into the query strain at this stage. The synthetic genetic array (SGA) method is used to introduce reporters and mutations of interest into the essential TS mutant and nonessential gene deletion collections via automated replica-pinning.

(B) High-throughput (HTP) preparation of cells for automated imaging. Cells are transferred to liquid medium or liquid medium containing drug to provide a chemically sensitized background. Objects in micrographs are segmented in CellProfiler, and an SVM-based classification is used to separate cells that contain a DNA-damage-induced focus from those that do not.

(C) Illustration of strategy for identifying hits in SGA-HCS screens of DNA damage foci. A distribution is displayed in which the average frequency of foci in all single gene deletion and TS mutant populations across 11 biological replicates ($n = 4.8 \times 10^4$) is scored, and the wild-type average distribution is highlighted in gray. Five positive controls are indicated with tick marks in the outlier set.

(D) Bootstrapping approach to select an optimal minimum cell count for analysis. The black dashed line indicates the SD in foci levels at the selected sample size minimum (1,000 cells/mutant SD = 0.82%). The error bars indicate SD in the number of Rad52-GFP foci detected across 100 replicates at each sample size, and the green line indicates the mean number of foci across replicates in each sample size.

(legend continued on next page)

approaches has emerged as a powerful approach for functional analysis (Carpenter et al., 2006). While several studies have examined subcellular morphology systematically in yeast (Alvaro et al., 2007; Breker et al., 2013; Chong et al., 2015; Huh et al., 2003; Tkach et al., 2012), a global characterization of mutant phenotypes remains a major challenge. Here, we describe an HTP pipeline for quantifying mutant phenotypes by combining two automated platforms: synthetic genetic array (SGA) analysis, which automates yeast genetics (Tong et al., 2001), and high-content screening (HCS), which enables quantitative cell biological analysis at the single-cell level (Chong et al., 2015; Li et al., 2011; Vizeacoumar et al., 2010). Unlike previous approaches, our method provides a fully scalable image-based approach for systematic analysis of yeast cells, enabling the detection of subcellular morphological defects in response to thousands of genetic or other perturbations in a quantitative and statistically robust manner.

As a case study, we monitored the presence of a transient, gigadalton-sized assembly of proteins referred to as a DNA-damage-induced focus. This subnuclear complex arises in response to double-stranded DNA breaks and acts as a recombination center for DNA repair (Lisby et al., 2001). The DNA damage focus is an appealing compartment for development of an automated imaging pipeline for several reasons. First, key proteins that form and influence the focus are highly conserved, and many genes with potential roles in focus formation or regulation have been identified through manual image inspection, providing useful positive controls (Alvaro et al., 2007). Second, the focus is a relatively simple shape, is usually found as a single entity in the cell, and has a substantial half life (~5 min; Lisby et al., 2003), facilitating the development of useful statistical approaches and automated imaging protocols. We used the DNA damage focus marker Rad52-GFP (Alvaro et al., 2007) to score foci formation in thousands of nonessential gene deletion mutants (Giaever et al., 2002; Winzeler et al., 1999) and conditional temperature-sensitive (TS) alleles of essential genes (Li et al., 2011), both in the presence and the absence of environmental and genetic perturbations. Our general approach is readily adaptable to other cell biological markers and experimental systems, and it enables systematic and quantitative analysis of genes influencing subcellular compartment morphology or pathway activity.

RESULTS

Designing a Robust SGA-HCS Pipeline for Identification of Cell Populations with Elevated Levels of DNA Damage Foci

Our strategy for systematic phenotypic analysis of the DNA damage focus required development of a completely automated pipeline for imaging and scoring DNA damage foci phenotypes in thousands of different yeast mutants. The first component of the pipeline involved assembly of yeast mutant arrays compatible with HTP image acquisition and analysis. We constructed SGA-compatible yeast strains containing phenotypically neutral markers for fluorescently labeled DNA damage foci (*RAD52-GFP*; Figures 1A, S1A, and S1B), nuclei (*HTA2-mCherry*; Figures 1A and S1B), and the cytoplasm (*RPL39*pr-*tdTomato;* Figure 1A; Table S1). SGA and HCS tools were then used to visualize fluorescent proteins in the yeast nonessential gene deletion collection, as well as a collection of mutants carrying TS alleles of essential genes (Figures 1A and 1B; Li et al., 2011). Three yeast mutant arrays were constructed using this protocol: (1) a single mutant array, which was assessed in the presence and absence of the DNA-damaging agent phleomycin; (2) a double-mutant array in which each strain carried a deletion allele of *SGS1*, which encodes a nonessential DNA helicase; and (3) a double-mutant array deleted for *YKU80*, which encodes a nonessential protein involved in non-homologous end-joining (NHEJ) and telomere maintenance. We reasoned that mutation of *SGS1* or *YKU80* would sensitize the cell to defects in the DNA damage response (DDR) in different ways, thus expanding our ability to discover a diverse set of functionally relevant genes.

The second component of our SGA-HCS pipeline involved HTP microscopy and automated image analysis and pattern classification through machine learning. Cell boundaries and nuclei were identified in the red channel, and 470 features were extracted from both the red and green channels for each cell using CellProfiler (Carpenter et al., 2006). We then coupled our feature selection with support vector machine (SVM)-based machine learning to generate a classifier capable of distinguishing nuclei with at least one focus from nuclei lacking any foci (STAR Methods; Figures S1C–S1E).

The final component of our method exploits the systematic and automated nature of the screening pipeline to address critical statistical considerations that could not be addressed in previous studies. Specifically, we defined three parameters to determine a cutoff at which biologically relevant hits could be identified: (1) a minimal cell sample size for reliable measurement, (2) a score to identify mutants that differ significantly from wild-type cells, and (3) a normalization strategy to remove screening bias. First, to determine the sample size required for reliable scoring of yeast mutants for foci detection, staggered sizes of populations were randomly selected from a pool of ~170,000 Rad52-GFP cells and scored for the presence of foci (Figures 1D and S1F). This analysis revealed that an imaging sample size of 1,000 cells/mutant allowed for reliable measurements of foci frequency (SD = 0.82%), a sample size that is difficult to achieve using manual assessment. The sample size criterion was met for ~80% of mutants examined in our experimental

(E) Graph illustrating fraction of mutant strains for which at least 1,000 cells were imaged. SM, single-mutant nonessential deletion mutants; SM-TS, single-mutant TS alleles of essential genes; SM+Phleo, nonessential deletion mutants with phleomycin; SM+Phleo-TS, TS allele array plus phleomycin; *yku80Δ*, nonessential deletion mutants lacking *YKU80*; *yku80Δ*-TS, TS allele array lacking *YKU80*; *sgs1Δ*, nonessential deletion mutants lacking *SGS1*; *sgs1Δ*-TS, TS allele array lacking *SGS1*.

(F) Graph showing precision of five scoring methods. Single mutants were scored for frequency of DNA damage foci using several methods. Precision was scored on ranked mutants using as a standard all genes annotated to the DNA replication/repair/cohesion functional category in Costanzo et al. (2010). BD, binomial distribution; F, Fisher's score.

See also Figure S1.

pipeline (Figure 1E); mutants with a severe fitness defect often failed this step, and strains for which fewer than 250 cells were observed across all biological replicates were excluded from further analysis.

We developed a score to reliably identify mutants that accounts for batch effects and other experimental biases that are typical of large-scale screens. We first scored and ranked mutants using a binomial distribution, a statistical test that determines the likelihood of a mutant having the same fraction of cells with foci as wild-type in the context of sample size, and replicates were combined to generate a single score for each mutant using Fisher's method (Elston, 1991; Skellam, 1948). To eliminate plate-specific effects, scores were normalized to the average fraction of foci on each plate, rather than a global average. The B-score, a non-parametric measure of deviation analogous to the well-known *Z* score, allowed us to filter out hits that contained bias as a result of positional effects within single plates (Malo et al., 2006). A combination of the binomial test and B-score outperformed the binomial test alone based on functional enrichment analysis of rank-ordered single mutants (Figure 1F; sample size interval = 10; maximum sample set = 200). Top-ranking single mutants scored using this method were more enriched for genes involved in DNA repair, homologous recombination, and cohesion than single mutants scored using raw foci percentage or *Z* score.

Applying SGA-HCS to Map Networks of DNA Damage Response Genes

We used our optimized SGA-HCS pipeline to observe ~1,000 yeast cells in each of ~5,000 different mutant backgrounds, in the context of four separate genetic and/or chemical perturbations for a total of ~24,000 different mutant populations and ~20 million single cells. On average, ~7% of the individual cells within a wild-type population exhibited a single focus, likely due to stalled replication forks and other endogenous sources of DNA damage (Figures 1C and 1D); however, application of our scoring criteria identified 345 loss-of-function mutants that had elevated levels of DNA damage foci, either as single mutants or as double mutants when combined with the deletion of *SGS1*, *YKU80*, or following treatment with the DNA-damaging agent phleomycin (Figure 2; Table S2). The sensitized backgrounds were chosen to illustrate particular aspects of the DDR. For instance, analysis of focus formation in the absence of *SGS1* allowed assessment of the DDR in the context of disruption to the repair machinery (Jessop and Lichten, 2008; Mimitou and Symington, 2008), while deletion of *YKU80* perturbed the NHEJ pathway (Boulton and Jackson, 1996). Almost half of the mutants (48%) were identified only in a sensitized background or condition, consistent with previous work illustrating the importance of considering genetic or chemical-genetic interactions for optimal exploration of yeast pathways (Costanzo et al., 2010; Vizeacoumar et al., 2010), and each chemical or genetic sensitization experiment identified a distinct set of mutants with elevated levels of DNA damage foci. For example, mutants identified in the absence of Yku80 were uniquely enriched for those with abnormal telomere size (log odds ratio [LOD] = 0.97; p value = 2.9×10^{-9}, Fisher's exact test; Figure 3A; Askree et al., 2004), consistent with known functions for Yku80. In contrast, enzymes involved in DNA metabolism, including the exonuclease Exo1, the endonuclease Mus81, and the helicase Srs2, were uniquely detected in the *sgs1Δ* double mutant screen (Figure 2), which contained numerous genes that show an *SGS1* genetic interaction or whose products are known to interact physically with Sgs1 (32/44; Figure S2A). A subset of these hits (14/44 nonessential mutants) showed increased sensitivity to hydroxyurea (HU) or methyl methanesulfonate (MMS) in the *sgs1Δ* background, supporting a combined role with Sgs1 in the DDR (Figures S2B and S2C). In some cases, genes were identified only in single mutant screens. This generally occurred for one of two reasons: (1) synthetic lethality of the gene with *SGS1* or *YKU80* or extreme sensitivity of a mutant to phleomycin, leading to few or no live cells being imaged; or (2) an elevated basal level of foci in the sensitized query strain without a corresponding increase in the double mutant, leading to mutants being removed from the hit list by the application of normalizing statistics (Figure 2). Notably 51% of the hits identified in the *sgs1Δ* screens were not previously reported to show a genetic interaction, despite extensive SGA genetic analysis with an *sgs1Δ* query strain (Figure 2; Costanzo et al., 2010); this demonstrates that our phenotypic analysis provides new information that would be missed in fitness-based genetic interaction studies.

We validated the results of our primary screens with two secondary assays, one using an independent deletion mutant collection to score Rad52 foci and another involving a test of plasmid-based gene complementation (Ho et al., 2009). Although we were not able to test all primary hits in the secondary assays due to technical issues (e.g., if a specific plasmid or strain was not available), we confirmed 152 mutants of 230 tested in at least one validation assay, which is suggestive of an upper boundary for the false positive rate of ~30% (Figures S3A–S3E; Costanzo et al., 2010). False positives may reflect discrepancies between biological replicates or aberrant cell segmentation issues with the original screen (Figure S1). We estimated a similar false negative rate of ~30% by assessing discrepancies among Rad52 focus phenotype in strains mutated for genes encoding members of the same protein complex, which should behave similarly in general (Figure S3F; Table S3).

We used several comparative analyses to explore and validate our primary screen results. First, we identified many genes with known roles in the DDR in the primary hit list from our screens. For example, genes important for double-strand break (DSB) repair were detected, including all tested members of the Rad52 epistasis group, DNA replication genes, and genes important for activating the DNA damage checkpoint (Figure 2). Second, genes identified in our screens showed enrichment for DNA replication, DNA repair, homologous recombination, and cohesion functions (Figure 3A, left; Table S4A). We also saw some screen-specific enrichments; for example, mutants defective in ER-to-Golgi trafficking were uniquely enriched in our phleomycin screen, likely due to the abrogation of drug-clearing mechanisms in the absence of specific transport genes. Third, hits from our nonessential mutant screens showed significant overlap with published screens that assayed single deletion mutants for sensitivity to DNA-damaging agents, higher frequencies of chromosomal loss, changes in telomere length, and increased DNA mutation rates (Figure 3A, right; Tables S4A and S4B). Similarly, nonessential deletion mutants identified in our screens were enriched for genes with genetic interaction profiles

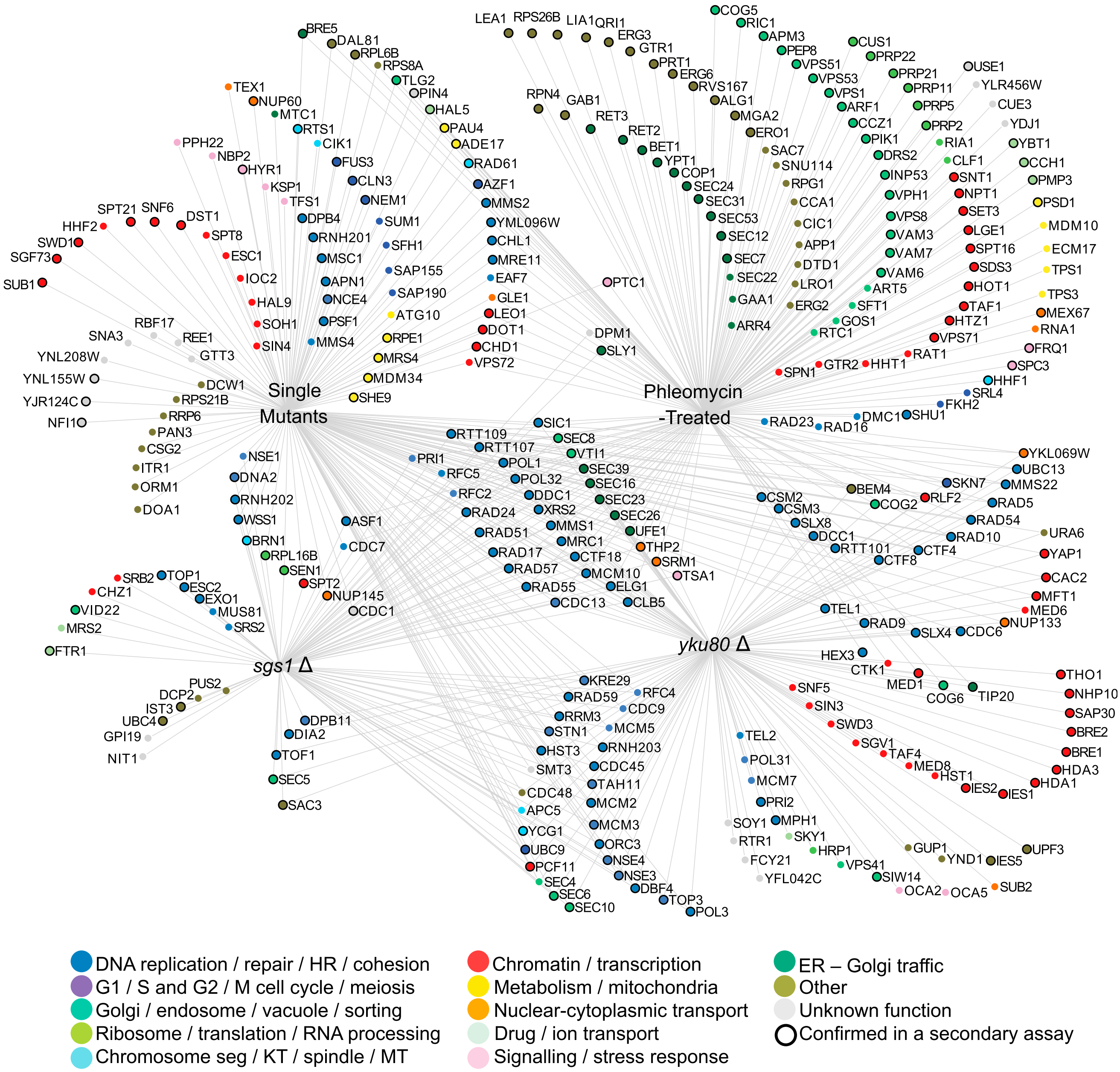


Figure 2. Mutants with Elevated Levels of Rad52-GFP Foci

Summary network of mutants with elevated levels of Rad52-GFP foci. The network diagram summarizes the results of all screens performed. Hub nodes indicate the screening condition (single mutants, BY4879; phleomycin treated, BY4879; *sgs1*Δ, BY4880; *yku80*Δ, BY4881), and edges connect these conditions to the hit genes whose deletion or conditional mutation is implicated in an elevated DNA damage foci phenotype. All hit nodes are color-coded according to functional category (legend below network), and those that confirmed in a secondary assay are outlined in black. Total hits = 345. See also Figures S2 and S3 and Tables S2 and S3.

resembling those annotated to functional categories affiliated with the DDR (Costanzo et al., 2010), indicating that these genes also have a DDR role (Figure 3B). Fourth, we performed a direct comparison between the single deletion mutants identified in this study and those found in a previous manual screen for increased Rad52-YFP foci (Alvaro et al., 2007); our study detected 31% of previously identified genes as well as 101 unique hits (Figure S4A; Table S4B). Importantly the genes uniquely identified in our study are enriched for DDR-associated functions (LOD = 1.04, p value = 4.38×10^{-16}; Fisher's exact test), indicating that this pool is likely enriched for true positives, which may have been missed due to sample size issues and biases associated with manual image assessment. Finally, the number of foci observed in populations of nonessential mutants showed statistically significant associations with several physiological and evolutionary properties of yeast genes. Notably, we observed a strong correlation between

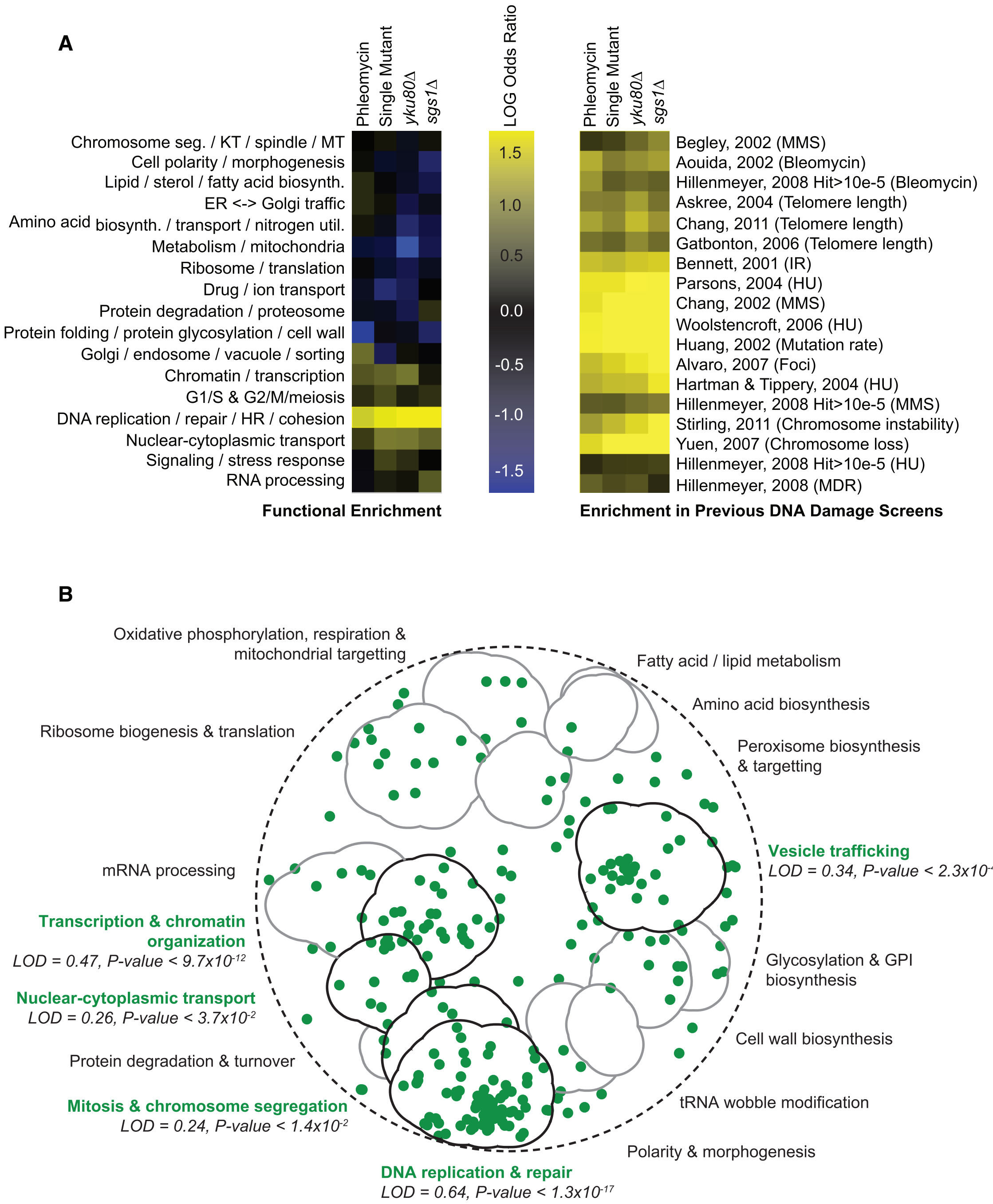


Figure 3. Functional Enrichments in Screens of Nonessential Gene Deletion Mutants for Elevated Levels of DNA Damage Foci

(A) Functional enrichment of hits from SGA-HCS screens of single deletion mutant array. Left: functional categories are derived from Costanzo et al. (2010) and are listed to the left of the heatmap. Yellow indicates a positive log odds ratio (LOD), or enrichment, and blue indicates a negative LOD, or an underrepresentation (scale bar between panels). KT, kinetochore; MT, microtubules; HR, homologous recombination. Right: enrichment of hits in 16 genome-wide datasets, each assessing an aspect of the DNA damage response pathway. Yellow indicates a positive LOD ratio, or enrichment, and blue indicates a negative LOD, or an underrepresentation. MMS, methyl methanesulfonate; IR, ionizing radiation; HU, hydroxyurea; MDR, multidrug resistance genes, in both homozygous and heterozygous deletion sets.

(B) Overlay of mutants exhibiting elevated levels of foci onto the yeast genetic interaction correlation network (Costanzo et al., 2010). The genetic interaction network described in Costanzo et al. (2010) is shown with the locations of 18 prominent bioprocess annotations outlined (solid lines). Nonessential genes identified as hits in our screens are overlaid on this network (green nodes), and the five bioprocesses in which these hits were most highly enriched in Costanzo et al. (2010) are annotated in green, with LOD and p values indicated in italics (black outlines).

See also Figure S4 and Tables S4A and S4B.

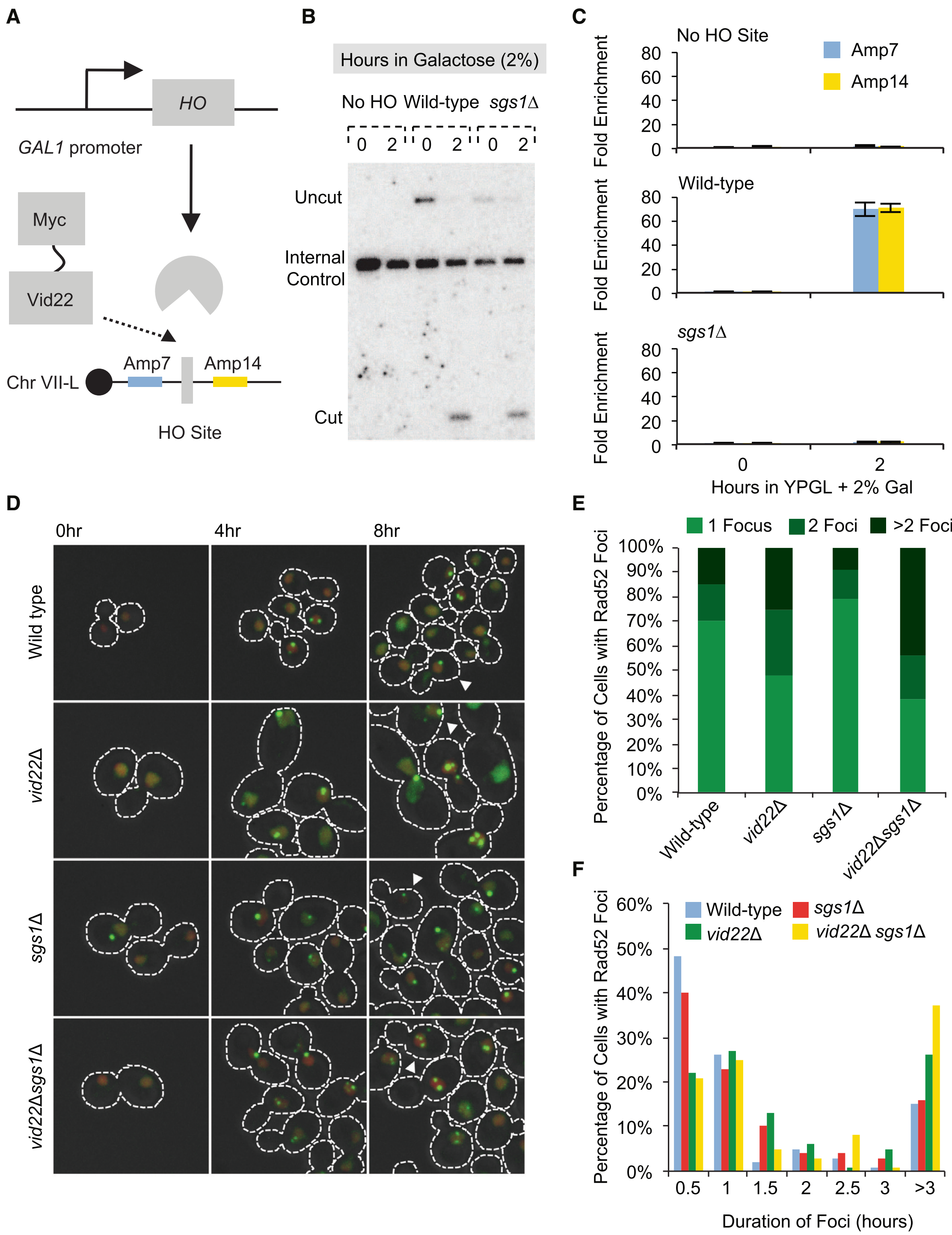


Figure 4. Localization of Vid22 to an Induced DSB and DNA Damage Focus Kinetics in the Absence of *VID22* and *SGS1*

(A) Schematic of a strain designed to query Vid22-Myc recruitment to an induced *HO* break. An *HO* cut site is integrated to the left of the centromere on chromosome VII, which is acted upon by the galactose-inducible *HO* endonuclease. Two probes (Amp7, blue; Amp14, yellow) adjacent to the DSB are used to assess Vid22-Myc binding.

(B) Southern blot analysis indicating HO endonuclease efficiently cleaves an integrated HO cut site. DNA from strains carrying a unique cut site for the HO endonuclease with (wild-type, BY5495; or *sgs1*Δ, BY5496) or without (no HO, BY5508) an integrated *GAL-HO* gene was digested with EcoRV. The blot was probed with both a ^{32}P-radiolabeled *ADE2* DNA fragment and a ^{32}P-radiolabeled *NMD5* fragment, and the uncut DNA, cut DNA, and an internal control (*SNR52*) are indicated.

(legend continued on next page)

an elevated foci phenotype and single mutant fitness, wherein mutants with increasingly severe foci phenotypes tend to have severe fitness defects, as well as correlations between a number of gene and protein attributes of biological networks and the elevated foci phenotype, consistent with previous observations that genes involved in DNA maintenance and organization have broad phenotypic impact (Figure S4B; Levy and Siegal, 2008).

Roles for Vid22 in Promoter Binding, the DNA Damage Response, and G-Quadruplex DNA

To explore DDR biology in our network, we decided to focus on Vid22, whose deletion caused a dramatic elevated focus phenotype only in the absence of Sgs1 (Table S2). Vid22 was recently linked to the DDR (Bonetti et al., 2013), but a potential functional relationship between Vid22 and Sgs1 was largely unexplored. Vid22 contains a BED (BEAF and DREF; boundary element-associated factor and DNA replication-regulated element binding factor, respectively)-finger domain, consistent with a function involving DNA binding (Aravind, 2000). Vid22 physically interacts with both Tbf1, an essential Myb domain telomere binding protein, and Env11, another BED-finger domain protein and paralog of Vid22 (Preti et al., 2010). These three proteins form a stable complex in which Vid22 and Env11 serve to stabilize the chromatin association of Tbf1 (Preti et al., 2010; Ribaud et al., 2012) and promote nucleosome rearrangements around promoters (Badis et al., 2008; Preti et al., 2010) and DSBs (Bonetti et al., 2013).

To investigate the biochemical connection between Sgs1 and Vid22, we used chromatin immunoprecipitation (ChIP) to assay recruitment of a Myc-tagged version of Vid22 to a unique, induced DNA DSB (Figures 4A and 4B; Ribeyre and Shore, 2012). Vid22-Myc was strongly recruited (~60-fold enriched over background) to both sides of an induced DSB, consistent with previous work (Bonetti et al., 2013). Recruitment of Vid22 was entirely dependent on *SGS1* (Figure 4C). Because we also observed a genetic interaction between *VID22* and *SGS1*, these two proteins may work in concert to control critical DNA repair functions during normal cell growth. We next performed a kinetic analysis to examine focus formation evident in *vid22Δ sgs1Δ* mutants, tracking cells over the course of 8 hr (Figure 4D). While the wild-type and *sgs1Δ* cells formed either one or two large foci, both *vid22Δ* and *vid22Δ sgs1Δ* cells formed multiple smaller foci in each cell (Figure 4E; Movies S1, S2, S3, and S4), with the phenotype being more pronounced in the double mutant. We also observed a higher frequency of long-lasting, unresolved foci in these mutant strains (foci lasting >3 hr; Figure 4F), consistent with a pronounced defect in DNA damage repair (Lukas et al., 2011).

We next explored aspects of genome integrity and DNA repair mechanisms in the mutant cells. First, we assayed the effect of *VID22* and/or *SGS1* deletion on the integrity of the rDNA cassette, a series of 9.2-kb repeat sequences of chromosome XII that are sequestered into the nucleolus; Rad52 is normally excluded from the nucleolus to prevent recombination between the repeats (Torres-Rosell et al., 2007). We employed an *ADE2* reporter system to assay unequal sister chromatid exchange (USCE) within the rDNA cassette and saw an increase in marker loss in *vid22Δ sgs1Δ* cells, beyond that observed in wild-type or single-mutant strains (Figures S5A and S5B). Furthermore, we saw elevated levels of extrachromosomal circles (ERCs) caused by intra-chromosomal recombination involving the rDNA in *vid22Δ*, *sgs1Δ*, and *vid22Δ sgs1Δ* double-mutant strains (Figures S5C and S5D). Consistent with these phenotypes, subnucleolar Rad52-GFP foci were increased in *vid22Δ sgs1Δ* cells (Figure S5E), which may also explain the elevated levels of DNA-damage-induced foci identified in our primary screens. We also assessed DNA damage phenotypes in *vid22Δ* mutants using a series of strains featuring galactose-inducible homothallic (HO) breaks in different genetic contexts, each of which query a different aspect of the DNA repair pathway. Growth of mutant strains was assessed using serial spot dilutions and revealed that *vid22Δ* mutants exhibited decreased fitness in a strain featuring an NHEJ repair-dependent HO break (Figure S6), consistent with previous results linking Vid22 to this DNA repair mechanism (Bonetti et al., 2013).

As noted above, Vid22 and its paralog, Env11, localize to promoter sites throughout the genome, together with the general regulatory factor Tbf1 (Preti et al., 2010). We wondered about the relationship between the Sgs1-dependent role for Vid22 in the DDR and the function(s) for Vid22 at promoter regions. We also explored the sites of Vid22 binding throughout the genome using a calling card assay, which assesses the frequency and location of Vid22-mediated Ty5 transposon integration on chromosomal DNA in vivo (Wang et al., 2011). Consistent with our ChIP-sequencing (ChIP-seq) analysis (Figure 5A; 67.5-fold increase over expected, p value = 2.9^{-56}, hypergeometric test; Preti et al., 2010), Vid22 localized specifically in promoter regions of 161 genes (Table S5), including those involved in DNA replication, repair, and recombination (e.g., *RAD14*, *NSE1*, and *HUG1*)

(C) Chromatin immunoprecipitation (ChIP) of Vid22 to an HO-induced DNA double-stranded break. Vid22 recruitment was assessed using probes to two sites by ChIP of Vid22-Myc before (0 on x axis) and after (2 on x axis) induction of HO (wild-type; BY5495 and *sgs1Δ*; BY5496). A strain with no HO site was used as a control (top; BY5508). Error bars represent the standard deviation for three replicate qPCR reactions.

(D) Kinetic analysis of Rad52-GFP focus formation. Wild-type (BY4879), *vid22Δ* (BY5418), *sgs1Δ* (BY4880), and *vid22Δ sgs1Δ* (BY5433) cells expressing Rad52-GFP (green) and Hta2-RFP (red) in logarithmic growth phase were imaged every 30 min for 8 hr. Merged projections of the differential interference contrast (DIC), green, and red channels are shown for the 0, 4, and 8 hr time points, and representative cells containing DNA damage foci are highlighted in the right panel with white arrowheads.

(E) Maximum number of Rad52-GFP foci per cell. Spontaneous Rad52 foci were counted in 100 wild-type, *vid22Δ*, *sgs1Δ*, and *vid22Δ sgs1Δ* cells as indicated. Individual cells were tracked for the duration of the 8-hr time course (one focus/cell, light green; two foci/cell, darker green; more than two foci/cell, darkest green). This assay was performed in a single biological replicate.

(F) Quantification of the duration of Rad52-GFP foci. Spontaneous Rad52-GFP foci were assessed every 30 min for more than 3 hr as indicated (x axis). Foci were followed in 100 wild-type (blue), *vid22Δ* (green), *sgs1Δ* (red), and *vid22Δ sgs1Δ* (yellow) cells. The percentage of cells with a persistent focus at each time point is shown.

See also Figures S5 and S6 and Movies S1, S2, S3, and S4.

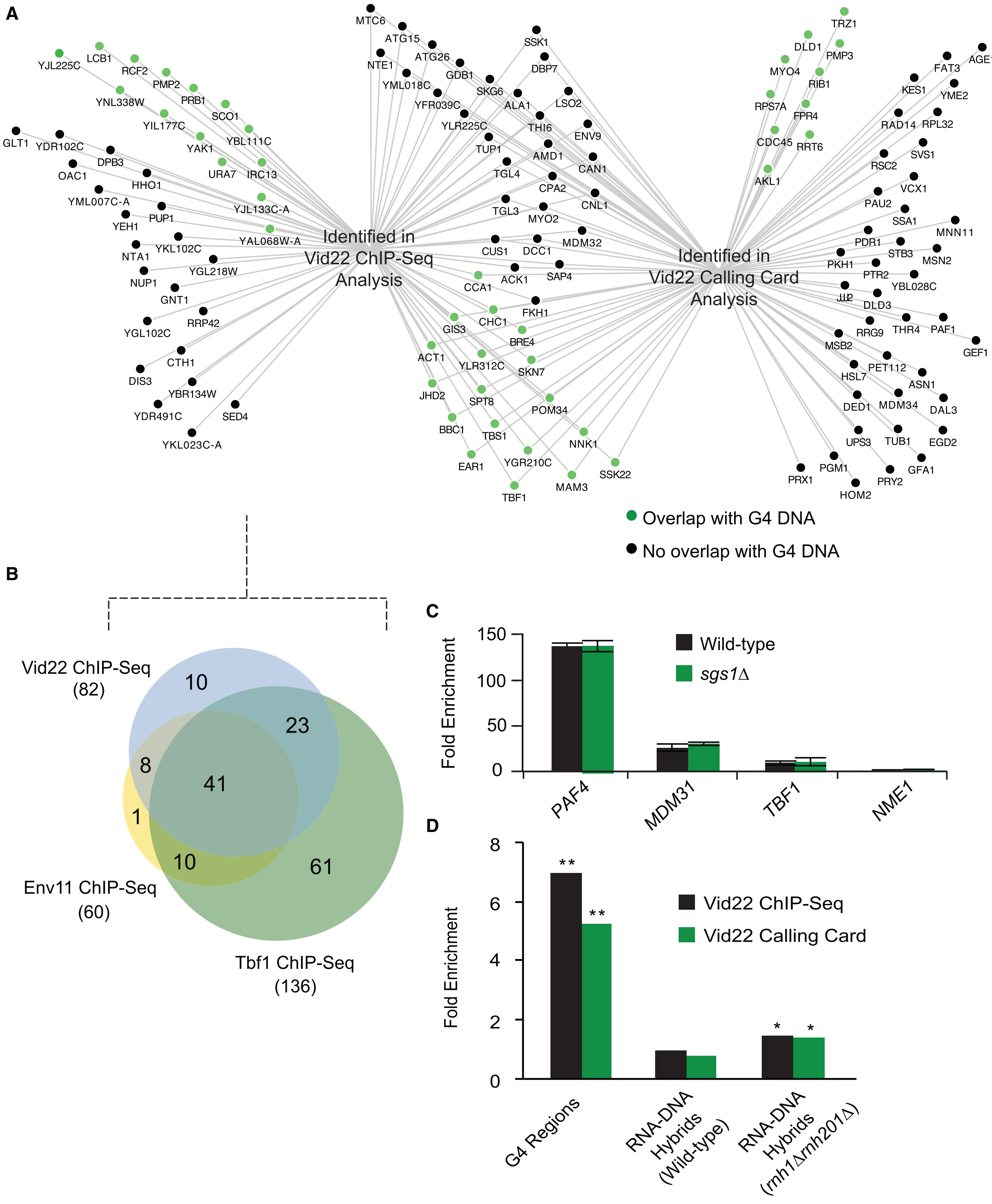


Figure 5. Localization of Vid22 to Gene Promoters

(A) Summary network of all loci identified as Vid22 binding sites by ChIP-seq (BY5493) and calling card (BY5487) analyses. Green nodes indicate a ChIP or calling card site that overlaps a region of G4 DNA (Capra et al., 2010), and black nodes represent those that do not overlap G4 regions.

(legend continued on next page)

and genes involved in ER-to-Golgi trafficking (e.g., *SED4* and *FRT2*), consistent with work suggesting a secretory defect in *vid22*Δ mutants (Brown et al., 2001; 2002). Although the Tbf1, Vid22 and Env11 ChIP-seq sites at non-small nucleolar RNA (non-snoRNA) promoters show high overlap (Figure 5B and Table S5; Preti et al., 2010), Vid22 recruitment to representative promoters was not affected by *SGS1* deletion (Figure 5C). Thus, the role of Vid22 and Sgs1 in the DDR appears distinct from the function of Vid22 in gene regulation.

Detailed analysis of our global Vid22 calling card and ChIP-seq analyses revealed an intriguing enrichment for predicted G-quadruplex (G4) DNA regions at Vid22 binding sites (Figure 5D; 6.9-fold increase over expected, p value = 1.6×10^{-19}, and 5.2-fold increase over expected, p value = 1.7×10^{-13}, hypergeometric test, respectively; Capra et al., 2010). G4 DNA regions have the potential to form four-stranded G4 quadruplex structures that are predicted to result from the opening of the DNA helix during either replication or transcription, and are resolved by a number of helicases, including Sgs1 (reviewed in Maizels and Gray, 2013). G4 DNA that is formed on the non-template strand during transcription is associated with a stable co-transcriptional RNA-DNA hybrid on the template strand, and is highly susceptible to DNA damage. Interestingly, the Vid22 ChIP and calling card binding sites were slightly enriched at loci known to be susceptible to RNA-DNA hybrids in an *rnh1*Δ *rnh201*Δ background (Figure 5D; 1.5-fold increase, p value = 0.02, and 1.4-fold increase, p value = 0.05, hypergeometric test, respectively; Chan et al., 2014). This suggests a possible functional overlap between Vid22 and members of the RNase HI and/or RNase HII complexes, which remove RNA-DNA hybrids by degrading RNA (reviewed in Aguilera and García-Muse, 2012).

Since Sgs1 can function as a G4 helicase (Sun et al., 1999), it is possible that Vid22 facilitates the unwinding of G4 structures or assists in the removal or prevention of the stable RNA-DNA heteroduplex at G4 sites. To investigate these possible functions, we examined genetic interaction data to ask which mutant strains share genetic interactions in common with a *vid22*Δ strain, a phenotype that is typical of genes that function in similar biological processes and pathways (Costanzo et al., 2010). As expected, *VID22* shared many genetic interactions with genes involved in the DDR, including negative genetic interactions with the structural maintenance of chromosomes (SMC) complex (Figure 6A), which has key roles in DNA repair and the segregation of repetitive DNA regions (Torres-Rosell et al., 2005). Consistent with a possible relationship between Vid22 and RNA-DNA hybrids, the same set of genetic interactions was also seen in strains mutated for genes involved in removal of RNA primers from DNA, including *RNH201*, *RNH202*, and *RNH203,* which encode the members of the RNase HII complex, as well as *DNA2,* which encodes a helicase and tracking protein for flap cleavage during Okazaki fragment maturation (Figure 6A). Notably, the apparent functional relationship between RNase HII and Vid22 that is suggested by these genetic interaction profiles was recapitulated in our SGA-HCS analysis, because we observed that RNase HII is required for genome integrity in an *sgs1*Δ mutant background. Consistent with the genetic data, we observed localization of RNase HII to induced DSBs (Figure 6B), and RNHII mutant strains had increased Rad52 foci in an *sgs1*Δ mutant background (Figure 2). These results are suggestive of a possible shared but complementary role for Vid22 and the RNase HII complex at sites of DNA damage.

DISCUSSION

In this study, we describe development of an optimized pipeline for combining SGA analysis with high content screening to identify budding yeast mutants with aberrant subcellular morphology, using the DDR pathway as a case study. We focused on developing methodology for identifying significant mutant phenotypes in cell images. As for other functional genomics screens (Baryshnikova et al., 2010), normalization of batch effects, including plate-specific results and spatial effects within a microtiter plate that influence the fraction of DNA-damage-induced foci in a given population, was critical for a statistically robust measurement of the focus phenotype of each mutant. SGA and liquid handling for HCS are automated separately in our pipeline, making these experimental factors somewhat sporadic and not amenable to classical normalization approaches, which treat all data with a single correction factor (Malo et al., 2006). Filtering data based on a site-specific correction factor that takes into account the relative spatial effect incurred at each location within the context of all other mutants on each plate was key for distinguishing biological effects as opposed to experimental anomalies. Our normalization protocol requires consolidation of data from multiple biological replicates, and the identification of a minimum number of cells to be counted for statistical reliability, which is easily achieved with computational image analysis.

We chose the DDR for optimization of our integrated SGA-HCS pipeline, since the core biology of DNA repair is well studied and conserved (Lisby et al., 2004), yet recent efforts to explore the DDR using unbiased genome-scale screens consistently reveal new biology (Alvaro et al., 2007; Tkach et al., 2012). Indeed, our screens identified 105 genes with well-established roles in the DDR, and 240 genes with poorly understood or previously unappreciated phenotypes associated with DDR defects. There are two important features of our experimental

(B) Overlap of Vid22 binding sites identified by ChIP-seq with Vid22, Env11, and Tbf1 ChIP-seq analysis (Vid22 ChIP-seq = blue, BY5493; Env11 ChIP-seq = yellow, BY5494; Tbf1 ChIP-seq = green, BY5507; Preti et al., 2010).

(C) Effect of *SGS1* deletion on Vid22 recruitment to promoter regions. Association of Vid22-Myc with promoters of known target genes (*PAF1*, *MDM31*, and *TBF1*) and a negative control gene (*NME1*) was assessed using ChIP as described in the legend of Figure 4 (wild-type, black; *sgs1*Δ, green). Error bars represent the SD between three replicate qPCR reactions.

(D) Enrichment of Vid22 binding sites at regions that overlap G4 DNA structures. Fold enrichment over background of Vid22 binding at predicted G4 DNA regions in Vid22 calling card (green) and ChIP-seq data (black) is shown (*, p value < 0.03; **, p value $\leq 1.7 \times 10^{-13}$). Less significant enrichment is seen at regions that are predisposed to elevated levels of RNA-DNA heteroduplex formation in wild-type (RNA-DNA hybrids wild-type) and an RNase HI and HII mutant strain (rnh1Δ rnh201Δ; Chan et al., 2014). Vid22 ChIP-seq, black; Vid22 calling card, green.
See also Table S5.

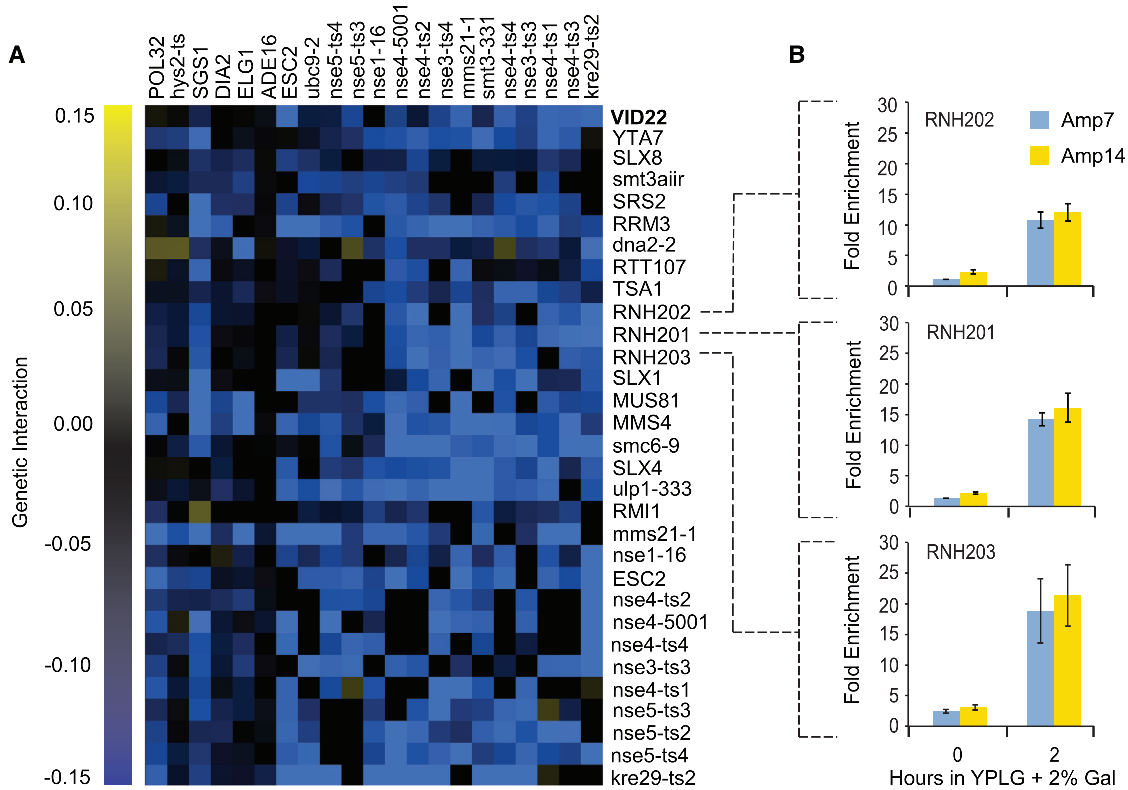


Figure 6. Two-Dimensional Hierarchical Clustering of Synthetic Genetic Interactions Associated with *VID22*

(A) Component of a large cluster-gram of genetic interactions involving deletion mutants of nonessential genes and TS allele mutants of essential genes (unpublished data available at http://andrewslab.ccbr.utoronto.ca/supplement/styles2015/; Costanzo et al., 2010). Array genes (x axis) and query genes (y axis) are hierarchically clustered based on genetic interaction score (yellow, positive GI; blue, negative GI; black, no GI; Baryshnikova et al., 2010). Uppercase gene names indicate nonessential genes screened as deletion mutants (*VID22* in bold), and lowercase gene names are associated with TS alleles of essential genes (different alleles are indicated by a unique allele number or designation).

(B) Association of Rnh202-Myc (top; BY5501), Rnh201-Myc (middle; BY5498), and Rnh203-Myc (bottom; BY5504) with a DNA double-strand break site. ChIP was performed after 0-hr and 2-hr induction of *HO* endonuclease as described in the legend of Figure 4. RNH recruitment was assessed using probes to two sites (Amp7, blue; Amp14, yellow). Error bars represent the SD between three replicate qPCR reactions.

pipeline that enabled discovery of this collection of potential new participants in the DDR. First, many genes were only linked to the DDR by screening in chemically or genetically sensitized backgrounds, consistent with previous systematic exploration of genetic interactions causing growth defects (Costanzo et al., 2010) and highlighting the importance of the automated genetics component of our method. Second, automated image analysis facilitated accurate measurement of a detailed cell biological phenotype (in this case, the DNA damage focus) that provides a highly sensitive assay for defects in the DDR. Although defects in the DDR often translate into cell growth defects, we identified 164 mutants in our cell biological screens that were not identified using fitness-based assays, either in standard growth conditions or in the presence of DNA damaging agents (Alvaro et al., 2007; Aouida et al., 2004; Begley et al., 2002; Bennett et al., 2001; Chang et al., 2002; Costanzo et al., 2010; Hartman and Tippery, 2004; Hillenmeyer et al., 2008; Parsons et al., 2004; Woolstencroft et al., 2006). Our method is readily extensible to other fluorescent markers covering fundamental subcellular compartments or structures, as well as markers of important phenotypes, such as aging and cell death, although marker-specific classifiers would need to be developed.

In an effort to understand DNA damage focus phenotypes that were only evident in a sensitized background, we focused on *VID22*, which was recently shown to be involved in the DDR but whose relationship to *SGS1* was largely unexplored (Bonetti et al., 2013). Our experiments revealed shared roles for *VID22* and *SGS1* in minimally two facets of genome integrity maintenance. First, several of our phenotypic tests suggest a prominent role for Sgs1 and Vid22 in rDNA integrity. Both unequal sister chromatid exchange and hyper-recombination of the rDNA repeats were elevated in *vid22*Δ *sgs1*Δ mutant populations, mirroring phenotypes seen in Bloom syndrome, a disease caused by mutations in the mammalian homolog of *SGS1* (*BLM*; Grierson et al., 2013; Langlois et al., 1989; Wang et al., 2003). Also, *vid22*Δ mutants were sensitive to DNA breaks that can only be repaired by NHEJ. These observations are consistent with previous work showing that Vid22 is required for recruitment of the DNA ligase Dnl4 to DSBs, which is necessary for DNA repair by NHEJ (Bonetti et al., 2013; Grierson et al., 2013; Wilson et al., 1997). NHEJ is the preferred method of break repair at the rDNA locus (Torres-Rosell et al., 2007), and in the absence of *SGS1*, Dnl4 has a special role in DSB repair as a result of collapsed replication forks at replication fork barriers in the rDNA cassette. Together, these observations implicate both Vid22 and Sgs1 in the repair of breaks via NHEJ, which is required for rDNA stability.

Our experiments also suggest a second role for Vid22 and Sgs1 in maintaining genome stability. We discovered an enrichment of Vid22 binding sites at predicted G4 DNA regions in the genome, implicating Vid22 in the processing, prevention, or removal of RNA-DNA heteroduplex structures, which are associated with G4 DNA. Consistent with this possibility, for both *VID22* and genes encoding members of the RNase HII complex, which processes RNA-DNA hybrids, we observed strong negative genetic interactions with genes encoding components of the Smc5-6 complex, which is known to be required for the removal of X-shaped DNA structures that arise between sister chromatids during DNA repair (Bermúdez-López et al., 2010). These observations suggest that RNA-DNA hybrids may accumulate in the absence of either Vid22 or the RNase HII proteins, causing replication fork stalls and collapses that require the SMC5-6 complex to resolve. A role for Vid22 in dealing with RNA-DNA hybrids may also involve Sgs1, because this helicase is important for removing the complicated DNA structures that result from collapsed replication forks, a phenotype that is exaggerated at the rDNA locus, which is especially prone to stalled and collapsed forks and elevated levels of RNA-DNA hybrids (Torres-Rosell et al., 2007). The mammalian homolog of *SGS1* (*BLM*) has been implicated in the unwinding of RNA-DNA hybrids (Grierson et al., 2013) and is a G4 helicase (Sun et al., 1999). Our automated imaging pipeline implemented in yeast cells may have identified a conserved pathway involving BED domain family proteins such as Vid22 and RecQ helicases (Sgs1-Blm1) in the maintenance of genome integrity through resolution of aberrant DNA structures linked to RNA-DNA hybrids.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures, five tables, and four movies and can be found with this article online at http://dx.doi.org/10.1016/j.cels.2016.08.008.

AUTHOR CONTRIBUTIONS

B.J.A., C.B., B.L., M.M.-F., C.L.M., R.D.M, D.S., G.W.B., and Z.Z. designed and supervised the project. E.B.S., K.J.F., T.L.S., D.A., C.R., V.R., D.M., D.N., and M.G carried out and analyzed experiments. E.B.S., K.J.F., and L.A.Z. performed large-scale analysis and interpretation. J.R., M.C., M.U., A.J.V., and E.N.K. performed additional data analysis. E.B.S., L.A.Z., T.L.S., D.A., E.N.K., and D.N. created figures. B.J.A., C.B., and E.B.S. wrote the manuscript.

ACKNOWLEDGMENTS

We thank Helena Friesen, Michael Cox, Yolanda Chong, Elena Kuzmin, Edith Cheng, Zhen-Yuan Lin, and Anne-Claude Gingras for technical assistance and advice. We thank Harsha Garadi Suresh for comments on the manuscript. This work was supported primarily by grant MOP-97939 and Foundation grants FDN-143264 and FDN-143265 from the Canadian Institutes for Health Research (to B.A. and C.B.) and by a grant from the National Institutes of Health (R01HG005853, to C.B., B.A., and C.L.M.), Contributions from the G.W.B. lab were supported by Impact grant 702310 from the Canadian Cancer Society Research Institute and NSERC Discovery Grant RGPIN 326897-12. E.B.S. was supported by a scholarship from the Canadian Institute of Health Research. Infrastructure for high-content imaging was acquired with funds from the Canadian Foundation for Innovation and the Ministry of Research and Innovation (Ontario; grant 21745, to B.J.A. and C.B.). B.J.A. and C.B. are senior fellows and co-director (C.B.) of the Genetic Networks program of the Canadian Institute for Advanced Research.

Received: March 29, 2016
Revised: May 27, 2016
Accepted: August 11, 2016
Published: September 8, 2016

SUPPORTING CITATIONS

The following references appear in the Supplemental Information: Cox et al. (2015); Gietz (2014); Sinclair et al., 1997.

REFERENCES

Aguilera, A., and García-Muse, T. (2012). R loops: from transcription byproducts to threats to genome stability. Mol. Cell *46*, 115–124.

Alvaro, D., Lisby, M., and Rothstein, R. (2007). Genome-wide analysis of Rad52 foci reveals diverse mechanisms impacting recombination. PLoS Genet. *3*, e228.

Aouida, M., Pagé, N., Leduc, A., Peter, M., and Ramotar, D. (2004). A genome-wide screen in Saccharomyces cerevisiae reveals altered transport as a mechanism of resistance to the anticancer drug bleomycin. Cancer Res. *64*, 1102–1109.

Aravind, L. (2000). The BED finger, a novel DNA-binding domain in chromatin-boundary-element-binding proteins and transposases. Trends Biochem. Sci. *25*, 421–423.

Askree, S.H., Yehuda, T., Smolikov, S., Gurevich, R., Hawk, J., Coker, C., Krauskopf, A., Kupiec, M., and McEachern, M.J. (2004). A genome-wide screen for Saccharomyces cerevisiae deletion mutants that affect telomere length. Proc. Natl. Acad. Sci. USA *101*, 8658–8663.

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

Baryshnikova, A., Costanzo, M., Kim, Y., Ding, H., Koh, J., Toufighi, K., Youn, J.-Y., Ou, J., San Luis, B.-J., Bandyopadhyay, S., et al. (2010). Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. Nat. Methods *7*, 1017–1024.

Begley, T.J., Rosenbach, A.S., Ideker, T., and Samson, L.D. (2002). Damage recovery pathways in Saccharomyces 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.

Bermúdez-López, M., Ceschia, A., de Piccoli, G., Colomina, N., Pasero, P., Aragón, L., and Torres-Rosell, J. (2010). The Smc5/6 complex is required for dissolution of DNA-mediated sister chromatid linkages. Nucleic Acids Res. *38*, 6502–6512.

Berriz, G.F., King, O.D., Bryant, B., Sander, C., and Roth, F.P. (2003). Characterizing gene sets with FuncAssociate. Bioinformatics *19*, 2502–2504.

Bonetti, D., Anbalagan, S., Lucchini, G., Clerici, M., and Longhese, M.P. (2013). Tbf1 and Vid22 promote resection and non-homologous end joining of DNA double-strand break ends. EMBO J. *32*, 275–289.

Boulton, S.J., and Jackson, S.P. (1996). Identification of a Saccharomyces cerevisiae Ku80 homologue: roles in DNA double strand break rejoining and in telomeric maintenance. Nucleic Acids Res. *24*, 4639–4648.

Breker, M., Gymrek, M., and Schuldiner, M. (2013). A novel single-cell screening platform reveals proteome plasticity during yeast stress responses. J. Cell Biol. *200*, 839–850.

Brown, C.R., Cui, D.Y., Hung, G.G., and Chiang, H.L. (2001). Cyclophilin A mediates Vid22p function in the import of fructose-1,6-bisphosphatase into Vid vesicles. J. Biol. Chem. *276*, 48017–48026.

Brown, C.R., McCann, J.A., Hung, G.G.-C., Elco, C.P., and Chiang, H.-L. (2002). Vid22p, a novel plasma membrane protein, is required for the fructose-1,6-bisphosphatase degradation pathway. J. Cell Sci. *115*, 655–666.

Byrne, K.P., and Wolfe, K.H. (2005). The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. Genome Res. *15*, 1456–1461.

Capra, J.A., Paeschke, K., Singh, M., and Zakian, V.A. (2010). G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in Saccharomyces cerevisiae. PLoS Comput. Biol. *6*, e1000861.

Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J., et al. (2006). CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol. *7*, R100.

Chan, Y.A., Aristizabal, M.J., Lu, P.Y.T., Luo, Z., Hamza, A., Kobor, M.S., Stirling, P.C., and Hieter, P. (2014). Genome-wide profiling of yeast DNA:RNA hybrid prone sites with DRIP-chip. PLoS Genet. *10*, e1004288.

Chang, C.-C., and Lin, C.-J. (2011). LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. *2*, 27.

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.

Chang, H.-Y., Lawless, C., Addinall, S.G., Oexle, S., Taschuk, M., Wipat, A., Wilkinson, D.J., and Lydall, D. (2011). Genome-wide analysis to identify pathways affecting telomere-initiated senescence in budding yeast. G3 (Bethesda) *1*, 197–208.

Chong, Y.T., Koh, J.L.Y., Friesen, H., Duffy, S.K., Cox, M.J., Moses, A., Moffat, J., Boone, C., and Andrews, B.J. (2015). Yeast Proteome Dynamics from Single Cell Imaging and Automated Analysis. Cell *161*, 1413–1424.

Collins, S.R., Miller, K.M., Maas, N.L., Roguev, A., Fillingham, J., Chu, C.S., Schuldiner, M., Gebbia, M., Recht, J., Shales, M., et al. (2007). Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. Nature *446*, 806–810.

Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E.D., Sevier, C.S., Ding, H., Koh, J.L.Y., Toufighi, K., Mostafavi, S., et al. (2010). The genetic landscape of a cell. Science *327*, 425–431.

Cox, M.J., Chong, Y.T., Boone, C., and Andrews, B. (2015). Liquid growth of arrayed fluorescently tagged *Saccharomyces cerevisiae* strains for live-cell high-throughput microscopy screens. In Budding Yeast: A Laboratory Manual, B. Andrews, C. Boone, T.N. Davis, and S. Fields, eds. (Cold Spring Harbor Laboratory Press).

Elston, R.C. (1991). On Fisher's method of combining p-values. Biomed. J. *33*, 339–345.

Gatbonton, T., Imbesi, M., Nelson, M., Akey, J.M., Ruderfer, D.M., Kruglyak, L., Simon, J.A., and Bedalov, A. (2006). Telomere length as a quantitative trait: genome-wide survey and genetic mapping of telomere length-control genes in yeast. PLoS Genet. *2*, e35.

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

Gietz, R.D. (2014). Yeast transformation by the LiAc/SS carrier DNA/PEG method. Methods Mol. Biol. *1205*, 1–12.

Grierson, P.M., Acharya, S., and Groden, J. (2013). Collaborating functions of BLM and DNA topoisomerase I in regulating human rDNA transcription. Mutat. Res. *743-744*, 89–96.

Gu, Z., Cavalcanti, A., Chen, F.-C., Bouman, P., and Li, W.-H. (2002). Extent of gene duplication in the genomes of Drosophila, nematode, and yeast. Mol. Biol. Evol. *19*, 256–262.

Haber, J.E. (2002). Uses and abuses of HO endonuclease. Methods Enzymol. *350*, 141–164.

Hartman, J.L., 4th, and Tippery, N.P. (2004). Systematic quantification of gene interactions by phenotypic array analysis. Genome Biol. *5*, R49.

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

Ho, C.H., Magtanong, L., Barker, S.L., Gresham, D., Nishimura, S., Natarajan, P., Koh, J.L.Y., Porter, J., Gray, C.A., Andersen, R.J., et al. (2009). A molecular barcoded yeast ORF library enables mode-of-action analysis of bioactive compounds. Nat. Biotechnol. *27*, 369–377.

Holstege, F.C., Jennings, E.G., Wyrick, J.J., Lee, T.I., Hengartner, C.J., Green, M.R., Golub, T.R., Lander, E.S., and Young, R.A. (1998). Dissecting the regulatory circuitry of a eukaryotic genome. Cell *95*, 717–728.

Huang, D., Moffat, J., and Andrews, B. (2002). Dissection of a complex phenotype by functional genomics reveals roles for the yeast cyclin-dependent protein kinase Pho85 in stress adaptation and cell integrity. Mol. Cell. Biol. *22*, 5076–5088.

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.

Jessop, L., and Lichten, M. (2008). Mus81/Mms4 endonuclease and Sgs1 helicase collaborate to ensure proper recombination intermediate metabolism during meiosis. Mol. Cell *31*, 313–323.

Jones, T.R., Kang, I.H., Wheeler, D.B., Lindquist, R.A., Papallo, A., Sabatini, D.M., Golland, P., and Carpenter, A.E. (2008). CellProfiler Analyst: data exploration and analysis software for complex image-based screens. BMC Bioinformatics *9*, 482.

Kaeberlein, M., McVey, M., and Guarente, L. (1999). The SIR2/3/4 complex and SIR2 alone promote longevity in Saccharomyces cerevisiae by two different mechanisms. Genes Dev. *13*, 2570–2580.

Kofoed, M., Milbury, K.L., Chiang, J.H., Sinha, S., Ben-Aroya, S., Giaever, G., Nislow, C., Hieter, P., and Stirling, P.C. (2015). An updated collection of sequence barcoded temperature-sensitive alleles of yeast essential genes. G3 (Bethesda) *5*, 1879–1887.

Langlois, R.G., Bigbee, W.L., Jensen, R.H., and German, J. (1989). Evidence for increased in vivo mutation and somatic recombination in Bloom's syndrome. Proc. Natl. Acad. Sci. USA *86*, 670–674.

Levy, S.F., and Siegal, M.L. (2008). Network hubs buffer environmental variation in Saccharomyces cerevisiae. PLoS Biol. *6*, e264.

Li, Z., Vizeacoumar, F.J., Bahr, S., Li, J., Warringer, J., Vizeacoumar, F.S., Min, R., Vandersluis, B., Bellay, J., Devit, M., et al. (2011). Systematic exploration of essential yeast gene function with temperature-sensitive mutants. Nat. Biotechnol. *29*, 361–367.

Lisby, M., Rothstein, R., and Mortensen, U.H. (2001). Rad52 forms DNA repair and recombination centers during S phase. Proc. Natl. Acad. Sci. USA *98*, 8276–8282.

Lisby, M., Mortensen, U.H., and Rothstein, R. (2003). Colocalization of multiple DNA double-strand breaks at a single Rad52 repair centre. Nat. Cell Biol. *5*, 572–577.

Lisby, M., Barlow, J.H., Burgess, R.C., and Rothstein, R. (2004). Choreography of the DNA damage response: spatiotemporal relationships among checkpoint and repair proteins. Cell *118*, 699–713.

Lukas, C., Savic, V., Bekker-Jensen, S., Doil, C., Neumann, B., Pedersen, R.S., Grøfte, M., Chan, K.L., Hickson, I.D., Bartek, J., and Lukas, J. (2011). 53BP1 nuclear bodies form around DNA lesions generated by mitotic transmission of chromosomes under replication stress. Nat. Cell Biol. *13*, 243–253.

Maizels, N., and Gray, L.T. (2013). The G4 genome. PLoS Genet. *9*, e1003468.

Malo, N., Hanley, J.A., Cerquozzi, S., Pelletier, J., and Nadon, R. (2006). Statistical practice in high-throughput screening data analysis. Nat. Biotechnol. *24*, 167–175.

Medvedik, O., and Sinclair, D.A. (2007). Caloric restriction and life span determination of yeast cells. Methods Mol. Biol. *371*, 97–109.

Mimitou, E.P., and Symington, L.S. (2008). Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. Nature *455*, 770–774.

Mnaimneh, S., Davierwala, A.P., Haynes, J., Moffat, J., Peng, W.-T., Zhang, W., Yang, X., Pootoolal, J., Chua, G., Lopez, A., et al. (2004). Exploration of essential gene functions via titratable promoter alleles. Cell *118*, 31–44.

Myers, C.L., Barrett, D.R., Hibbs, M.A., Huttenhower, C., and Troyanskaya, O.G. (2006). Finding function: evaluation methods for functional genomic data. BMC Genomics *7*, 187.

Ohya, Y., Sese, J., Yukawa, M., Sano, F., Nakatani, Y., Saito, T.L., Saka, A., Fukuda, T., Ishihara, S., Oka, S., et al. (2005). High-dimensional and large-scale phenotyping of yeast mutants. Proc. Natl. Acad. Sci. USA *102*, 19015–19020.

Parsons, A.B., Brost, R.L., Ding, H., Li, Z., Zhang, C., Sheikh, B., Brown, G.W., Kane, P.M., Hughes, T.R., and Boone, C. (2004). Integration of chemical-genetic and genetic interaction data links bioactive compounds to cellular target pathways. Nat. Biotechnol. *22*, 62–69.

Preti, M., Ribeyre, C., Pascali, C., Bosio, M.C., Cortelazzi, B., Rougemont, J., Guarnera, E., Naef, F., Shore, D., and Dieci, G. (2010). The telomere-binding protein Tbf1 demarcates snoRNA gene promoters in Saccharomyces cerevisiae. Mol. Cell *38*, 614–620.

Ribaud, V., Ribeyre, C., Damay, P., and Shore, D. (2012). DNA-end capping by the budding yeast transcription factor and subtelomeric binding protein Tbf1. EMBO J. *31*, 138–149.

Ribeyre, C., and Shore, D. (2012). Anticheckpoint pathways at telomeres in yeast. Nat. Struct. Mol. Biol. *19*, 307–313.

Rost, B. (1999). Twilight zone of protein sequence alignments. Protein Eng. *12*, 85–94.

Sinclair, D.A., Mills, K., and Guarente, L. (1997). Accelerated aging and nucleolar fragmentation in yeast sgs1 mutants. Science *277*, 1313–1316.

Skellam, J. (1948). A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials. J. R. Stat. Soc. Series B Stat. Methodol. *10*, 257–261.

Stark, C., Breitkreutz, B.J., Reguly, T., Boucher, L., Breitkreutz, A., and Tyers, M. (2006). BioGRID: a general repository for interaction datasets. Nucleic Acids Res. *34*, D535–D539.

Stirling, P.C., Bloom, M.S., Solanki-Patil, T., Smith, S., Sipahimalani, P., Li, Z., Kofoed, M., Ben-Aroya, S., Myung, K., and Hieter, P. (2011). The complete spectrum of yeast chromosome instability genes identifies candidate CIN cancer genes and functional roles for ASTRA complex components. PLoS Genet. *7*, e1002057.

Sun, H., Bennett, R.J., and Maizels, N. (1999). The Saccharomyces cerevisiae Sgs1 helicase efficiently unwinds G-G paired DNAs. Nucleic Acids Res. *27*, 1978–1984.

Tkach, J.M., Yimit, A., Lee, A.Y., Riffle, M., Costanzo, M., Jaschob, D., Hendry, J.A., Ou, J., Moffat, J., Boone, C., et al. (2012). Dissecting DNA damage response pathways by analysing protein localization and abundance changes during DNA replication stress. Nat. Cell Biol. *14*, 966–976.

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

Torres-Rosell, J., Machín, F., Farmer, S., Jarmuz, A., Eydmann, T., Dalgaard, J.Z., and Aragón, L. (2005). SMC5 and SMC6 genes are required for the segregation of repetitive chromosome regions. Nat. Cell Biol. *7*, 412–419.

Torres-Rosell, J., Sunjevaric, I., De Piccoli, G., Sacher, M., Eckert-Boulet, N., Reid, R., Jentsch, S., Rothstein, R., Aragón, L., and Lisby, M. (2007). The Smc5-Smc6 complex and SUMO modification of Rad52 regulates recombinational repair at the ribosomal gene locus. Nat. Cell Biol. *9*, 923–931.

Vizeacoumar, F.J., van Dyk, N., Vizeacoumar, F.S., Cheung, V., Li, J., Sydorskyy, Y., Case, N., Li, Z., Datti, A., Nislow, C., et al. (2010). Integrating high-throughput genetic interaction mapping and high-content screening to explore yeast spindle morphogenesis. J. Cell Biol. *188*, 69–81.

Wang, W., Seki, M., Narita, Y., Nakagawa, T., Yoshimura, A., Otsuki, M., Kawabe, Y., Tada, S., Yagi, H., Ishii, Y., and Enomoto, T. (2003). Functional relation among RecQ family helicases RecQL1, RecQL5, and BLM in cell growth and sister chromatid exchange formation. Mol. Cell. Biol. *23*, 3527–3535.

Wang, H., Mayhew, D., Chen, X., Johnston, M., and Mitra, R.D. (2011). Calling Cards enable multiplexed identification of the genomic targets of DNA-binding proteins. Genome Res. *21*, 748–755.

Wapinski, I., Pfeffer, A., Friedman, N., and Regev, A. (2007). Natural history and evolutionary principles of gene duplication in fungi. Nature *449*, 54–61.

Wilson, T.E., Grawunder, U., and Lieber, M.R. (1997). Yeast DNA ligase IV mediates non-homologous DNA end joining. Nature *388*, 495–498.

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.

Woolstencroft, R.N., Beilharz, T.H., Cook, M.A., Preiss, T., Durocher, D., and Tyers, M. (2006). Ccr4 contributes to tolerance of replication stress through control of CRT1 mRNA poly(A) tail length. J. Cell Sci. *119*, 5178–5192.

Yuen, K.W., Warren, C.D., Chen, O., Kwok, T., Hieter, P., and Spencer, F.A. (2007). Systematic genome instability screens in yeast and their potential relevance to cancer. Proc Natl Acad Sci U S A *104*, 3925–3930.

CellPress

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal anti-myc 9E10	From culture supernatant, Shore lab	N/A
Dynabeads M-280 sheep anti-mouse IgG	Dynal, ThermoFisher	11202D
Chemicals, Peptides, and Recombinant Proteins		
Phleomycin	InvivoGen	Ant-ph-1
Hydroxyurea	Santa Cruz	Sc-29061A
Methyl methanesulfonate	Aldrich	129925
Canavanine	Sigma	C9758
S-aminoethyl-L-cysteine	Sigma	A2636
Nourseothricin	Werner BioAgents	CAS 96736-11-7
Geneticin	Life Technologies	11811098
Critical Commercial Assays		
RNeasy RNA extraction mini kit	QIAGEN	74104
GoScript cDNA synthesis kit	Promega	A5000
LightCycler 480 SYBR Green I Master Mix	Roche	04707516001
Prime-It II Random Primer Labeling Kit	Agilent Technologies	#300385
Sequencing	Illumina GA	Fasteris, S.A.
Experimental Models: Organisms/Strains		
*MAT***a** *xxx*Δ::*KANMX his3*Δ*1 leu2*Δ*0 ura3*Δ*0 met15*Δ*0*	The Yeast Deletion Collection; Giaever et al., 2002	N/A
*MAT***a** xxx-ts::*KANMX his3*Δ*1 leu2*Δ*0 ura3*Δ*0 met15*Δ*0*	The Yeast Collection of Temperature-sensitive Strains; Li et al., 2011	V 6.0
*MAT*α xxxΔ::*NATMX can1*Δ::*STE2pr-Sp_his5 lyp1*Δ *his3*Δ*1 leu2*Δ*0 ura3*Δ0 *met15*Δ*0*	The Yeast Collection of *MAT*α NATMX-marked Deletion Query Strains; Costanzo et al., 2010	N/A
S. cerevisiae strains derived from the BY4741 and W303 backgrounds, see Table S1	This paper	N/A
Recombinant DNA		
Molecular Barcoded Yeast (MoBY) ORF 1.0 plasmid collection	http://moby.ccbr.utoronto.ca; Ho et al., 2009	N/A
Software and Algorithms		
CellProfiler	http://cellprofiler.org/releases/; Carpenter et al., 2006	V 1.0.5811
CellProfiler Analyst	http://cellprofiler.org/releases/; Jones et al., 2008	N/A
MATLAB and Statistics Toolbox Release 2011b	http://www.mathworks.com/includes_content/domainRedirect/domainRedirect.html?uri=http%3A%2F%2Fwww.mathworks.com%2Fproducts%2Fmatlab%2F	N/A
SGA Genetic Interaction Score	Baryshnikova et al., 2010	N/A
R software package	R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org/	N/A
SVM; libSVM package and libSVM interface to MATLAB	Chang and Lin, 2011; https://www.csie.ntu.edu.tw/~cjlin/libsvm/	N/A
FuncAssociate 2.0: The Gene Set Functionator	http://llama.mshri.on.ca/funcassociate/; Berriz et al., 2003	N/A

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents may be directed to, and will be fulfilled by the corresponding author Brenda J. Andrews (brenda.andrews@utoronto.ca).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Strain Construction and Confirmation that Tags Did Not Impair Protein Function

To visualize DNA damage foci within the cell, we fused a *GFP::HIS3* cassette to the C terminus of the endogenous *RAD52* gene using polymerase chain reaction and lithium acetate transformation (Gietz, 2014). To test for a possible growth defect associated with the fluorescent tag, a saturated Rad52-GFP culture was serially diluted 10-fold five times, spotted on synthetic complete (SC) media and SC media containing 100 mM hydroxyurea (HU), and growth was assessed after 2 and 3 days respectively (Figure S1A; strains BY4394, BY4879, and a *rad52*Δ strain from the Yeast Deletion Collection; Giaever et al., 2002). To provide spatial and cell cycle context, we fused an mCherry::*NatMX* cassette to the endogenous *HTA2* locus to mark the nucleus and *RPL39*pr-tdtomato::*CaURA3* (Figure 1A; Chong et al., 2015) was integrated into the *CAN1* promoter locus to mark the cytoplasm. Fitness of strains in SGA output arrays was also assessed, by quantifying colony size using the SGA score (Baryshnikova et al., 2010) and comparing these values to single mutant fitness scores (Figure S1B; Costanzo et al., 2010). Of 117 previously identified synthetic lethal interactions with *RAD52* and 34 synthetic lethal interactions with *HTA2*, 113 and 31 mutants had no growth defect when combined with the *RAD52-GFP HTA2-RFP* tagged strain. To generate query strains containing deletions of *SGS1* or *YKU80*, the *RPL39*pr-*RFP::CaURA3* cassette was integrated into the *SGS1* or *YKU80* locus instead of the *CAN1* promoter locus (BY4880, BY4881). To provide SGA compatibility, all fluorescent reporters were generated in a BY4741-based SGA query strain background (BY4394; *MAT*α *his3*Δ*1 leu2*Δ*0 ura3*Δ*0 MET15 can1*Δ*::STE2*pr-*LEU2 lyp1*Δ).

S. cerevisiae strains are listed in Table S1. Strains containing *RAD52-GFP, HTA2-mCherry and RPL39*pr-*tdTomato* (with *sgs1*Δ or *yku80*Δ in some experiments) were crossed to the deletion collection (Giaever et al., 2002) and to a collection of mutants carrying TS alleles of essential genes (Figures 1A and 1B; Li et al., 2011) and haploid (mutant) strains carrying the fluorescent protein markers were selected using the SGA method (Tong et al., 2001).

METHOD DETAILS

High Throughput Preparation and Imaging of Yeast Cells

Cells were prepared for imaging as described in detail previously (Cox et al., 2015). Briefly, cells were grown to saturation in 200 μl SD+MSG medium (0.1% monosodium glutamate, 0.17% yeast nitrogen base without amino acids and without ammonium sulfate, 2% glucose, 0.15 g/l methionine) with antibiotic in 96-well beaded microplates, then diluted in 800 μl SD+MSG low fluorescence medium (with 0.17% yeast nitrogen base without amino acids and without ammonium sulfate and without riboflavin and folic acid) plus antibiotic in beaded deep-well blocks and grown overnight to early log phase. Non-essential gene deletion mutants were grown at 30°C and TS mutants in essential genes were grown at 22°C and then incubated for three hours at 37°C prior to imaging. Cells at ~0.2-0.4 OD_{600} /ml were transferred to a 384-well Perkin-Elmer Ultra imaging plate and left to settle for ten minutes before imaging. Four images per well, each containing fifty to a hundred cells, were taken in a single plane using an automated spinning disk confocal microscope (Evotec Opera, PerkinElmer) with a 60x water-immersion objective. Details of the imaging protocol are described in Chong et al. (2015).

Secondary Analysis of Mutants in the *NatMX*-Marked Mutant Collection

All mutants identified in primary screens with increased levels of DNA damage-induced foci were compiled into mini-arrays and assessed in a parallel yeast mutant collection marked with the *NatMX* antibiotic resistance cassette. Four mutants that were identified as hits in primary screens could not be assessed in this way, as they were not present in the *NatMX* collection (*GTT3, SNA3, SOH1, PRI1*). Query strains marked with *GFP::KAN*, *RFP::LEU2* and *RFP::CaURA3* (BY5084, BY5085, BY5086) were crossed to a *NatMX*-marked collection of non-essential deletion mutants (Costanzo et al., 2010), as well as a collection of essential TS mutants (Li et al., 2011), both containing a *MAT***a**-specific promoter driving *SpHis5*. SGA, imaging and analysis were performed as described above. In order to identify and rank hits in this collection, a statistical analysis identical to that used in the analysis of primary screens was employed.

Confirmation of Single Mutants with Increased Levels of Foci by Gene Complementation

Single mutants identified with increased levels of DNA damage-induced foci were confirmed using a gene complementation assay. Eighty single gene deletion mutants marked with *GFP::HIS3* and *RFP::NatMX* (constructed in a BY5092 background) were transformed with plasmids from the Molecular Barcoded Yeast (MoBY) ORF collection (Ho et al., 2009). The remaining 48 mutants could not be assessed, as the predicted size of the ORF in the MoBY plasmids was not successfully confirmed via plasmid digest. Each deletion mutant of interest was transformed with a MoBY-ORF plasmid as well as an empty vector (EV) control plasmid. Transformants were grown on SD-U plates, and subsequently replica plated onto SD-UH+N medium. Colonies transformed with MoBY-ORF plasmids as well as EVs were imaged in at least quintuplicate. In order to identify the successful rescue of deletion mutants by gene complementation, the standard deviation of the percentage of foci in the population was calculated between a given deletion mutant in the primary screen and that mutant following transformation with the EV plasmid. Mutants were considered to have been rescued if the percentage of foci in the deletion mutant following transformation with the MoBY-ORF plasmid was 2 standard deviations or higher from the percentage of foci in the mutant following transformation with the EV plasmid.

Assessment of Chemical Sensitivity of *sgs1*Δ Double Mutants

A mini-array featuring the top non-essential hits identified in the *sgs1*Δ screens was created using standard SGA pinning technology (see above), by crossing in either a wild-type query strain or an *sgs1*Δ query strain, in 3 biological replicates. Final selection arrays were pinned onto either SC media with no drug, SC media + 100mM HU, or SC media + 0.01% methyl methanesulfonate (MMS), and colony size was assessed. Colony growth defects were scored and normalized using the SGA score (Baryshnikova et al., 2010; Costanzo et al., 2010) and significant interactions were scored by the following calculation:

$$\text{Interaction Score} = \frac{(\text{mean score of } sgs1\Delta genex\Delta + \text{drug}/\text{mean score of } sgs1\Delta genex\Delta - \text{drug})}{(\text{mean score of } genex\Delta + \text{drug}/\text{mean score of } genex\Delta - \text{drug})}$$

A threshold interaction score of 0.9 was used, and any double mutants with a score less than this threshold were confirmed by performing serial 10-fold dilutions of saturated cultures, and spotting them onto synthetic complete (SC) media, SC media containing 100mM HU, and SC media containing 0.01% MMS, and assessing for sensitivity after 2, 4, and 9 days, respectively.

Chromatin Immunoprecipitation of Vid22

To assess localization of proteins to Gal-inducible HO breaks, cells were grown in YPLG (lactic acid / glycerol) medium for 3 hr, followed by induction of HO endonuclease expression by addition of galactose to the medium (2%). ChIP to breaks (strains BY5508, BY5495, BY5496, BY5498, BY5501 and BY5504) and ChIP-Seq (strains BY5507, BY5493, BY5494 and BY5508) assays were carried out as previously described (Ribeyre and Shore, 2012). To confirm that HO-mediated cleavage was occurring, Southern blots were performed on genomic DNA digested with *Eco*RV (Figure 4B). DNA samples were run on a 0.8% agarose gel, and transferred to Hybond N+ nylon membrane. The blot was probed with both a ^{32}P-radiolabeled *ADE2* DNA fragment and a ^{32}P-radiolabeled *NMD5* fragment. To determine enrichment of promoter regions in the Vid22 ChIP, fold enrichment was calculated after normalization to both the input fraction and an internal control (*SNR52*).

Kinetic Live Cell Imaging

Strains BY4879, BY4880, BY5418 and BY5433 were grown to an optical density at 600 nm of 0.4 in YPD and imaged using a spinning disc confocal system (WaveFX; Quorum) on a Leica DMI 6000B microscope with Velocity 4 software (PerkinElmer). Images were captured at 30 min intervals in microfluidic chambers (CellAsic; Y04C ONIX plates) with constant flow of YPD at room temperature for 8 hr. Each image represents the projection of eleven 0.4 mm z-stacks in the DIC, GFP and RFP channels. Images were merged, GFP and RFP levels were adjusted to optimize foci visualization, and image sequences were made in ImageJ 1.45 s.

Assessment of Sub-nucleolar Rad52-GFP Foci

Wild-type (BY5440), *vid22*Δ (BY5442), *sgs1*Δ (BY5441), and *vid22*Δ *sgs1*Δ (BY5443) strains harboring Rad52-GFP and Nop56-mCherry were grown to mid-log phase in SC +G418 medium and imaged using a spinning disc confocal system (WaveFX; Quorum) on a Leica DMI 6000B microscope with Velocity 4 software (PerkinElmer). A minimum of 850 cells was imaged in each strain background, and foci were quantified manually. The presence of sub-nucleolar foci was assessed via the co-localization of Rad52-GFP and Nop56-mCherry, which localizes to the nucleolus.

rDNA Unequal Sister Chromatid Exchange Assay

Rate of loss of an *ADE2* marker integrated into the rDNA array was used to measure the instability at the rDNA locus (Kaeberlein et al., 1999). Wild-type (BY5481), *vid22*Δ (BY5482), *sgs1*Δ (BY5483), and *vid22*Δ *sgs1*Δ (BY5484) strains were grown overnight and then plated onto solid YPD with 12.5 μg/ml adenine. Colonies were grown 3-4 days at 28°C, and then placed at 4°C for 3 days prior to analysis. The number of half-red/half-white colonies was determined; each was assumed to represent a marker loss event during the first cell division after plating. The number of half-sectored colonies divided by the total number of colonies (excluding entirely red colonies) was reported as the rate of marker loss. About 10,000-15,000 colonies were examined for each strain in each experiment.

Extrachromosomal rDNA Circle Analysis via Southern Blot

Genomic DNA was isolated from strains BY5479, BY5480, BY5401, BY5160 and an *rrm3Δ* strain from the Yeast Deletion Collection (Giaever et al., 2002) as follows (Medvedik and Sinclair, 2007). Cells were incubated for 30 min at 30°C in 500 μl 0.5 mg/ml zymolyase (100T), 1 M sorbitol, 14 mM mercaptoethanol, then 80 μl 10% SDS was added, tubes were inverted to mix, and incubated at 65°C for 20 min. Two hundred μl of 5 M potassium acetate was added and tubes were inverted to mix and left on ice 30 min. Samples were centrifuged for 5 min at high speed and the supernatant was retained, then precipitated with ethanol. Samples were treated with RNase, then extracted with phenol-chloroform, and subsequently reprecipitated with ethanol. At this time, 1 μl of glycogen and 1/10th of the volume of 3M sodium acetate was also added, to aid in efficient DNA precipitation in the absence of tRNA. DNA was digested for three hours at 37°C using BamHI (New England BioLabs, #R0136S), which does not cut within the rDNA cassette, and was analyzed in a 0.7% agarose gel (Certified Megabase Agarose, Bio-Rad; #161-3108). DNA was then transferred to Hybond N+ nylon membrane (Amersham; GE Healthcare Life Sciences; #RPN82N). Plasmid 2484 (originally pNL47; Sinclair et al., 1997) was digested for three hours at 37°C using EcoRI (New England BioLabs, #R0101S) and prepared for use as a ^{32}P-radiolabeled probe to the rDNA repeat using the Prime-It II Random Primer Labeling Kit (Agilent Technologies; #300385) and Spin-*Pure* G-50 Columns (Pure Biotech; #SCD50-50). ^{32}P-radiolabeled *HDA1* DNA was used as a loading control.

Sensitivity to Gal-Inducible Double-Strand Breaks by the Homothallic Endonuclease

Strains Y14220-223, 307-310, 318, 327, 330, 332, 334, 338, 346, and 380, which each carry a galactose (GAL)- inducible allele of the homothallic (HO) endonuclease (Figures S6B–S6G; described in Haber, 2002) were grown overnight in YEPR (yeast extract peptone raffinose). Saturated cultures were serially diluted 10- fold, and spotted onto YEPG (yeast extract peptone galactose) and YEPD. Strain growth was assessed for sensitivity after three days.

Calling Card Analysis of Vid22

Calling card analysis was performed as follows in accordance with Wang et al., 2011. In brief, Vid22 was tagged with the component of the Sir4 protein that physically interacts with the Ty5 integrase, in a *sir4*Δ background (BY5487, with control BY5486), transposition was induced, and genomic insertions of the transposon were selected. The integration sites of Ty5 transposons were then mapped using paired-end DNA sequencing, and are detailed in full in Table S5. In order to compare directly to Vid22 ChIP-seq results and DNA elements, all Vid22-Sir4-directed Ty5 integration peaks for which more than one possible gene target was identified were reduced to a single putative hit. The 96 gene promoters chosen for this reduced list were those that were either closest to the Vid22-Sir4-directed Ty5 integration peak, or had a corresponding hit in the Vid22 ChIP-seq dataset.

QUANTIFICATION AND STATISTICAL ANALYSIS

Bootstrapping to Determine Ideal Cell Count

To estimate the relationship between foci to cell object ratio and sample size, we performed a sampling experiment using the R software package. We sampled from three populations of yeast cells (*his3*Δ, *xrs2*Δ and *rad51*Δ), and in three biological replicates. We sampled on two scales: first on a small scale ranging from 10 cells to 100 cells in increments of 10, and on a larger scale ranging from 150 cells to 2000 cells in increments of 50. For each of the three gene pools, we sampled the designated number of cells randomly without replacement 100 times from populations of ~170,000 cells, and calculated the mean and standard deviation of the ratio of cells with foci to total cells sampled. We then took the average of both mean and standard deviation for foci ratio to estimate the mean and standard deviation for each sample size (Figures 1D and S1F).

Classification of DNA Damage Foci

Classification was used to detect DNA damage foci in cellular objects identified and measured using CellProfiler image analysis software. A training set was constructed using CellProfiler Analyst™ and consisted of ~1000 cells containing at least one DNA damage focus (positive bin) and ~1000 cells that did not contain a DNA damage focus (negative bin). Each of these training objects was associated with approximately 1000 features measuring different aspects of each image. A Wilcoxon rank-sum test was used to select only features that were informative for distinguishing the positive and negative bins (470 features in total). A support vector machine (SVM; libSVM package and libSVM interface to MATLAB) was trained using an SVM training model called svmtrain, in which a linear kernel was specified. Cross-validation training was performed on $1/5^{th}$ of the training set and a receiver-operating characteristic (ROC) curve was generated by calculating a false positive rate (TP/[TP+FP]) and a true positive rate or recall (TP/[TP+FN]); Figure S1C). Following training, the classifier was used to make label predictions for all identified cells within the screen (focus versus non-focus). To validate this approach, a set of 50 images was manually inspected, identifying a good agreement with the automated foci detection (r = 0.96; Figure S1D). The classifier was further validated on a per object basis using a pool of ~1000 cells from each screen with an average false positive rate of 15.5% and an average false negative rate of 1% (Figure S1E).

Calculation of False Negative Rate

False Negative Rate (FNR) was calculated by assessing the division of protein complexes between screens. If any member of a protein complex was identified in any screen (single mutant or sensitized backgrounds), all other screens were assessed for the identification of members of the same protein complex. Any discrepancies in the specific members of the complex identified between screens were labeled as false negatives. If no members of the complex were identified in a given screen background, this was not labeled as a FN hit, but rather discounted from the calculation as uninformative data. FNR was determined by dividing the total number of complex members that were "missed" by the total number of complex members that should have been identified in all screens, giving a FNR of 27% (Table S3). A selection of four mutants identified as FN in the *sgs1*Δ screens were reconstructed (BY5781-84) to confirm that a Rad52-focus phenotype was present, indicating that these mutants were incorrectly assigned as "negative" in the primary screen analysis (Figure S3F).

Scoring Enrichment and Underrepresentation of Non-essential Mutants

Hits were scored for significant enrichment and underrepresentation by inputting hit lists into FuncAssociate 2.0: The Gene Set Functionator (Berriz et al., 2003), available at http://llama.mshri.on.ca/funcassociate/. LOG Odds (LOD) ratios were calculated by comparison to a manually generated associations file, using the algorithm for an unordered gene list, and calculating both under- and over-enrichment. One thousand simulations were performed, and a significance cutoff of 1 was employed to identify scores and p values for all input categories. Enrichment of our data were calculated within the functional categories specified in Costanzo et al. (2010) and in other pre-existing DNA damage screens (Figure 3A; Table S4; Alvaro et al., 2007; Aouida et al., 2004; Askree et al., 2004; Begley et al., 2002; Bennett et al., 2001; Chang et al., 2002, 2011; Gatbonton et al., 2006; Hartman and Tippery, 2004; Hillenmeyer et al., 2008; Huang et al., 2002; Levy and Siegal, 2008; Parsons et al., 2004; Stirling et al., 2011; Woolstencroft et al., 2006; Yuen et al., 2007).

Pearson Correlation of Hits with Phenotypic and Evolutionary Traits

The relationship between foci score (Fisher's score) and several gene / protein-level features was computed to characterize the properties of genes implicated in the DNA damage response pathway. For each quantitative feature described below, the Pearson correlation coefficient (PCC) between the foci score and the 3885 array genes was calculated (Figure S4B).

- Negative genetic interaction (GI) degree: negative interactions were used directly from published SGA data (Costanzo et al., 2010).
- Phenotypic capacitance: Used directly from (Levy and Siegal, 2008), and summarizes variance across a range of morphological phenotypes upon deletion of each non-essential gene.
- Single mutant fitness defect: Single mutant fitness for all non-essential deletion mutants was derived from mutant colony size data as described (Baryshnikova et al., 2010; Costanzo et al., 2010). The fitness defect ($1\text{-}f_i$) for a single mutant fitness (f_i) was used.
- Multi-functionality: A quantitative standard for gene multi-functionality was defined from annotations to "biological process" terms of the Gene Ontology. The total number of annotations across the set of functionally distinct GO terms was used as a multi-functionality index (Costanzo et al., 2010; Myers et al., 2006).
- Yeast conservation: the number of species that possess an ortholog of a given gene, when considering 23 divergent species of Ascomycota fungi (measure described with the term "persistence"), and the corresponding ortholog data were downloaded from https://portals.broadinstitute.org/regev/orthogroups/. The 23 species are an expanded set of the original 17 species described previously (Wapinski et al., 2007), with the additions of *S. octosporus, S. japonicus, L. elongosporus, C. parasilosis, C. tropicalis* and *C. guilliermondii.*
- Chemical-genetic degree: data measuring the sensitivity of all non-essential deletion mutants to a library of drugs, and a variety of environmental conditions were used (Hillenmeyer et al., 2008). The number of drug and environmental sensitivities for a specific deletion mutant in the homozygous dataset that met a minimum cutoff of p value < 0.05 were summed.
- Protein-protein interaction degree (PPI) is the number of physical interactions reported in BioGRID, version 2.0.58 (Stark et al., 2006) and consists of: Affinity Capture-MS, Affinity Capture-RNA, Affinity Capture-Western, Biochemical Activity, Co-crystal Structure, Co-fractionation, Co-localization, Co-purification, Far Western, FRET, PCA, Protein-peptide, Protein-RNA, Reconstituted Complex, and Two-hybrid.
- Expression variation: represents the average number of mRNA copies of each transcript per cell as assessed in (Holstege et al., 1998).
- Whole genome duplicate (WGD): the list of duplicate pairs is comprised of those identified as the result of a whole genome duplication event (Byrne and Wolfe, 2005). Additionally, any pair of genes fulfilling established similarity requirements (Gu et al., 2002) was also considered a duplicate pair resulting from a small scale duplication event. Specifically, a gene pair must have sufficient sequence similarity score (FASTA Blast, E = 10), and sufficient protein alignment length (> 80% of the longer protein). A pair must also have an amino acid level identity of at least 30% for proteins with aligned regions longer than 150 aa, and $0.01n + 4.8L - 0.32^{(1+\exp(-L/1000))}$ for shorter proteins, where L is the aligned length, and n = 6 (Gu et al., 2002; Rost, 1999). Pairs from the WGD event were combined with pairs determined through sequence alone.
- SGA Ratio: a measure of LOG(positive interactions / negative interactions) for each non-essential mutant (Costanzo et al., 2010).

Association between Vid22 DNA-binding Sites and DNA Elements

To identify potential biological functions for Vid22 at specific loci, the association of Vid22 with some known genomic features was analyzed. Given the query sets of all possible Vid22 ChIP binding sites and calling card binding sites, four reference sets (G-quadruplex DNA, γH2A sites, loci with elevated basal levels of RNA-DNA hybrids, and loci with elevated levels of RNA-DNA hybrids in an *rnh1*Δ *rnh202*Δ double mutant background) were assessed to identify the number of overlapping regions between the reference and query sets. Regions of G-quadruplex (G4) DNA and γH2A sites in the yeast genome were assessed based on previously reported data (Capra et al., 2010). Direct overlap of ChIP-seq and calling card binding regions with these genomic structures was assessed using an expanded form of the regions (500 bp up and downstream of the G4 or γH2A site), since these genomic features are very short (average length is 60.9bp ± 36.8bp and 57.9bp ± 2.9bp, respectively). Loci with elevated levels of RNA-DNA hybrid formation were assessed at the ORF level rather than precise overlapping sequence information, as sequence information was not available for these features (Chan et al., 2014). Fold enrichment was calculated using the following formula:

$$\text{Fold Enrichment} = (s/S)/(p/P)$$

in which *s* represents the number of successes in the given sample (e.g., number of G4 sites that overlap with Vid22 ChIP binding sites), *S* represents the total sample size (e.g., the total number of Vid22 ChIP binding sites), *p* represents the number of successes in the population (e.g., the total number of G4 DNA regions in the genome), and *P* represents the total population size (e.g., the total number of possible G4 regions, based on the cumulative size of the genome). In the case of both G4 regions and γH2A sites, the average size of the feature including the expanded 500bp window was taken into account to identify the total number of possible sites in the genome (i.e., p = Total length of the genome / Average length of expanded feature). In the case of loci with elevated levels of RNA-DNA hybrids, the total number of yeast ORFs was used as the total population size (i.e., p = 6117).

DATA AND SOFTWARE AVAILABILITY

Data Resources

Genetic Interaction Analysis of VID22

Genetic interactions and correlations with *VID22* were identified using the SGA score (Baryshnikova et al., 2010; Costanzo et al., 2010).

Data are available at http://andrewslab.ccbr.utoronto.ca/supplement/styles2015/.

Images

Raw image data will be made available on request.

SOFTWARE AVAILABILITY

The segmentation pipeline used to identify fluorescently tagged cells (listed below) is compatible with CellProfiler version 1.0.5811. The SVM-based classifier used to identify cells with Rad52-GFP foci and the training set of cells, either positive or negative for Rad52-GFP foci, that was generated using CellProfiler Analyst to train the classifier are available at https://github.com/lzamparo/styles2016. In addition, this link contains the MATLAB model file of the classifier and the code used to calculate the B-score for each screen.

CellProfiler Pipeline

Module #1: LoadImages

- Text-Exact match: Type the text that one type of image has in common (for TEXT options), or their position in each group (for ORDER option): *.flex*
- What do you want to call these images within CellProfiler? *Orig*
- Type the text that one type of image has in common (for TEXT options), or their position in each group (for ORDER option). Type "Do not use" to ignore: *Do not use*
- What do you want to call these images within CellProfiler? (Type "Do not use" to ignore) *Do not use*
- Type the text that one type of image has in common (for TEXT options), or their position in each group (for ORDER option): *Do not use*
- What do you want to call these images within CellProfiler? *Do not use*
- Type the text that one type of image has in common (for TEXT options), or their position in each group (for ORDER option): *Do not use*
- What do you want to call these images within CellProfiler? *Do not use*
- If using ORDER, how many images are there in each group (i.e., each field of view)? *3*
- What type of files are you loading? *tif,tiff,flex movies*
- Analyze all subfolders within the selected folder? *Yes*
- Enter the path name to the folder where the images to be loaded are located. Type period (.) for default image folder.

Module #2: GroupMovieFrames

- What did you call the movie you want to extract from? *Orig*
- How many frames should be extracted each cycle? *2*
- Are the frames grouped by cycle interleaved (ABCABC...) or separated (AA..BB..CC..)? *Interleaved*
- What do you want to call frame 1 in each cycle (or "Do not use" to ignore)? *GFP*
- What do you want to call frame 2 in each cycle (or "Do not use" to ignore)? *RFP*
- What do you want to call frame 3 in each cycle (or "Do not use" to ignore)? *Do not use*
- What do you want to call frame 4 in each cycle (or "Do not use" to ignore)? *Do not use*
- What do you want to call frame 5 in each cycle (or "Do not use" to ignore)? *Do not use*
- What do you want to call frame 6 in each cycle (or "Do not use" to ignore)? *Do not use*

CellPress

Module #3: RescaleIntensity

- What did you call the image to be rescaled? *RFP*
- What do you want to call the rescaled image? *RescaledRFP*
- Rescaling method. (S) Stretch the image (0 to 1). (E) Enter the minimum and maximum values in the boxes below. (G) rescale so all pixels are equal to or Greater than one. (M) Match the maximum of one image to the maximum of another. (C) Convert to 8 bit. (T) Divide by loaded text value. See the help for details. *Stretch 0 to 1*
- (Method E only): Enter the intensity from the original image that should be set to the lowest value in the rescaled image, or type AA to calculate the lowest intensity automatically from all of the images to be analyzed and AE to calculate the lowest intensity from each image independently. *AA*
- (Method E only): Enter the intensity from the original image that should be set to the highest value in the rescaled image, or type AA to calculate the highest intensity automatically from all of the images to be analyzed and AE to calculate the highest intensity from each image independently. *AA*
- (Method E only): What value should pixels at the low end of the original intensity range be mapped to (range [0,1])? *0*
- (Method E only): What value should pixels at the high end of the original intensity range be mapped to (range [0,1])? *1* (Method E only): What value should pixels *below* the low end of the original intensity range be mapped to (range [0,1])? *0*
- (Method E only): What value should pixels *above* the high end of the original intensity range be mapped to (range [0,1])? *1* (Method M only): What did you call image whose maximum you want rescaled image to match? *Orig*
- (Method T only): What did you call the loaded text in the LoadText module?

Module #4: RescaleIntensity

- What did you call the image to be rescaled? *GFP*
- What do you want to call the rescaled image? *RescaledGFP*
- Rescaling method. (S) Stretch the image (0 to 1). (E) Enter the minimum and maximum values in the boxes below. (G) rescale so all pixels are equal to or Greater than one. (M) Match the maximum of one image to the maximum of another. (C) Convert to 8 bit. (T) Divide by loaded text value. See the help for details. *Stretch 0 to 1*
- (Method E only): Enter the intensity from the original image that should be set to the lowest value in the rescaled image, or type AA to calculate the lowest intensity automatically from all of the images to be analyzed and AE to calculate the lowest intensity from each image independently. *AA*
- (Method E only): Enter the intensity from the original image that should be set to the highest value in the rescaled image, or type AA to calculate the highest intensity automatically from all of the images to be analyzed and AE to calculate the highest intensity from each image independently. *AA*
- (Method E only): What value should pixels at the low end of the original intensity range be mapped to (range [0,1])? *0*
- (Method E only): What value should pixels at the high end of the original intensity range be mapped to (range [0,1])? *1*
- (Method E only): What value should pixels *below* the low end of the original intensity range be mapped to (range [0,1])? *0*
- (Method E only): What value should pixels *above* the high end of the original intensity range be mapped to (range [0,1])? *1*
- (Method M only): What did you call image whose maximum you want rescaled image to match? *Orig*
- (Method T only): What did you call the loaded text in the LoadText module?

Module #5: IdentifyPrimAutomatic

- ☐ What did you call the images you want to process? *RFP*
- ☐ What do you want to call the objects identified by this module? *Nuclei*
- ☐ Typical diameter of objects, in pixel units (Min,Max): *6,40*
- ☐ Discard objects outside the diameter range? *Yes*
- ☐ Try to merge too small objects with nearby larger objects? *No*
- ☐ Discard objects touching the border of the image? *Yes*
- ☐ Select an automatic thresholding method or enter an absolute threshold in the range [0,1]. To choose a binary image, select "Other" and type its name. Choosing 'All' will use the Otsu Global method to calculate a single threshold for the entire image group. The other methods calculate a threshold for each image individually. "Set interactively" will allow you to manually adjust the threshold during the first cycle to determine what will work well. *Otsu Global*
- ☐ Threshold correction factor *2*
- ☐ Lower and upper bounds on threshold, in the range [0,1] *0.0013,1*
- ☐ For MoG thresholding, what is the approximate fraction of image covered by objects? *0.01*
- ☐ Method to distinguish clumped objects (see help for details): *Intensity*
- ☐ Method to draw dividing lines between clumped objects (see help for details): *Intensity*
- ☐ Size of smoothing filter, in pixel units (if you are distinguishing between clumped objects). Enter 0 for low resolution images with small objects ($\sim < 5$ pixel diameter) to prevent any smoothing. *Automatic*
- ☐ Suppress local maxima within this distance, (a positive integer, in pixel units) (if you are distinguishing between clumped objects) *Automatic*
- ☐ Speed up by using lower-resolution image to find local maxima? (if you are distinguishing between clumped objects) *Yes*

- ☐ Enter the following information, separated by commas, if you would like to use the Laplacian of Gaussian method for identifying objects instead of using the above settings: Size of neighborhood (height, width), Sigma, Minimum Area, Size for Wiener Filter (height, width),Threshold *Do not use*
- ☐ What do you want to call the outlines of the identified objects (optional)? *NucleiOutline*
- ☐ Do you want to fill holes in identified objects? *Yes*
- ☐ Do you want to run in test mode where methods for distinguishing clumped objects are compared? *No*

Module #6: MeasureObjectAreaShape

- ☐ What did you call the objects that you want to measure? *Nuclei*
- ☐ Would you like to calculate the Zernike features for each object? *Yes*

Module #7: MeasureObjectIntensity

- ☐ What did you call the greyscale images you want to measure? *GFP*
- ☐ What did you call the objects that you want to measure? *Nuclei*

Module #8-15: MeasureTexture

- What did you call the greyscale images you want to measure? *GFP*
- What did you call the objects that you want to measure? *Nuclei*
- What is the scale of texture? *1-8*

Module #16: MeasureObjectIntensity

- ☐ What did you call the greyscale images you want to measure? *RFP*
- ☐ What did you call the objects that you want to measure? *Nuclei*

Module #17-24: MeasureTexture

- ☐ What did you call the greyscale images you want to measure? *RFP*
- ☐ What did you call the objects that you want to measure? *Nuclei*
- ☐ What is the scale of texture? *1-8*

Module #25: ExpandOrShrink

- ☐ What did you call the objects that you want to expand or shrink? *Nuclei*
- ☐ What do you want to call the expanded or shrunken objects? *ExpandNuclei*
- ☐ Were the objects identified using an Identify Primary or Identify Secondary module (note: shrinking results are not perfect with Secondary objects)? *Primary*
- ☐ Do you want to expand or shrink the objects? *Expand*
- ☐ Enter the number of pixels by which to expand or shrink the objects, or "Inf" to either shrink to a point or expand until almost touching, or 0 (the number zero) to simply add partial dividing lines between objects that are touching (experimental feature). *2*
- ☐ What do you want to call the outlines of the identified objects (optional)? *ExpandedNucleiOutline*

Module #26: MeasureObjectAreaShape

- ☐ What did you call the objects that you want to measure? *ExpandNuclei*
- ☐ Would you like to calculate the Zernike features for each object? *Yes*

Module #27: MeasureObjectIntensity

- ☐ What did you call the greyscale images you want to measure? *GFP*
- ☐ What did you call the objects that you want to measure? *ExpandNuclei*

Module #28-35: MeasureTexture

- ☐ What did you call the greyscale images you want to measure? *GFP*
- ☐ What did you call the objects that you want to measure? *ExpandNuclei*
- ☐ What is the scale of texture? *1-8*

Module #36: MeasureObjectIntensity

- What did you call the greyscale images you want to measure? *RFP*
- What did you call the objects that you want to measure? *ExpandNuclei*

Module #37-44: MeasureTexture

- What did you call the greyscale images you want to measure? *RFP*
- What did you call the objects that you want to measure? *ExpandNuclei*
- What is the scale of texture? *1-8*

Module #45: IdentifySecondary

- ☐ What did you call the primary objects you want to create secondary objects around? *Nuclei*
- ☐ What do you want to call the objects identified by this module? *Cells*
- ☐ Select the method to identify the secondary objects (Distance - B uses background; Distance - N does not): *Propagation*
- ☐ What did you call the images to be used to find the edges of the secondary objects? For DISTANCE - N, this will not affect object identification, only the final display. *RescaledRFP*
- ☐ Select an automatic thresholding method or enter an absolute threshold in the range [0,1]. To choose a binary image, select "Other" and type its name. Choosing 'All' will use the Otsu Global method to calculate a single threshold for the entire image group. The other methods calculate a threshold for each image individually. Set interactively will allow you to manually adjust the threshold during the first cycle to determine what will work well. *Otsu Global*
- ☐ Threshold correction factor *0.8*
- ☐ Lower and upper bounds on threshold, in the range [0,1] *0.04,1*
- ☐ For MoG thresholding, what is the approximate fraction of image covered by objects? *0.01*
- ☐ For DISTANCE, enter number of pixels by which to expand the primary objects [Positive integer] *10*
- ☐ For PROPAGATION, enter the regularization factor (0 to infinity). Larger = distance,0 = intensity *0.05*
- ☐ What do you want to call the outlines of the identified objects (optional)? *CellOutline*
- ☐ Do you want to run in test mode where each method for identifying secondary objects is compared? *No*

Module #46: MeasureObjectAreaShape

- ☐ What did you call the objects that you want to measure? *Cells*
- ☐ Would you like to calculate the Zernike features for each object? *Yes*

Module #47: MeasureObjectIntensity

- ☐ What did you call the greyscale images you want to measure? *GFP*
- ☐ What did you call the objects that you want to measure? *Cells*

Module #48-55: MeasureTexture

- ☐ What did you call the greyscale images you want to measure? *GFP*
- ☐ What did you call the objects that you want to measure? *Cells*
- ☐ What is the scale of texture? *1-8*

Module #56: MeasureObjectIntensity

- ☐ What did you call the greyscale images you want to measure? *RFP*
- ☐ What did you call the objects that you want to measure? *Cells*

Module #57-64: MeasureTexture

- ☐ What did you call the greyscale images you want to measure? *RFP*
- ☐ What did you call the objects that you want to measure? *Cells*
- ☐ What is the scale of texture? *1-8*

Module #65: OverlayOutlines

- ☐ On which image would you like to display the outlines? *RescaledRFP*
- ☐ What did you call the outlines that you would like to display? *CellOutline*
- ☐ Would you like to set the intensity (brightness) of the outlines to be the same as the brightest point in the image, or the maximum possible value for this image format? *Max of image*
- ☐ What do you want to call the image with the outlines displayed? *CellRFP*
- ☐ For color images, what do you want the color of the outlines to be? *Red*

Module #66: OverlayOutlines

- ☐ On which image would you like to display the outlines? *RescaledGFP*
- ☐ What did you call the outlines that you would like to display? *ExpandedNucleiOutline*
- ☐ Would you like to set the intensity (brightness) of the outlines to be the same as the brightest point in the image, or the maximum possible value for this image format? *Max of image*
- ☐ What do you want to call the image with the outlines displayed? *ExpNucleiGFP*
- ☐ For color images, what do you want the color of the outlines to be? *Green*

Module #67: ExportToDatabase

- ☐ What type of database do you want to use? *MySQL*
- ☐ For MySQL only, what is the name of the database to use? *FociDB*
- ☐ What prefix should be used to name the tables in the database (should be unique per experiment, or leave "Do not use" to have generic Per_Image and Per_Object tables)? *Do not use*

- ☐ What prefix should be used to name the SQL files? *SQL_*
- ☐ Enter directory where the SQL files are to be saved. Type period (.) to use the default output folder.
- ☐ Do you want to create a CellProfiler Analyst properties file? *Yes*

Module #68: CreateBatchFiles

- ☐ What is the path to the folder where the batch control file (Batch_data.mat) will be saved? Leave a period (.) to use the default output folder.
- ☐ If pathnames are specified differently between the local and cluster machines, enter that part of the pathname from the local machine's perspective, omitting trailing slashes. Otherwise, leave a period (.) */Volumes/MetaXpress/*
- ☐ If pathnames are specified differently between the local and cluster machines, enter that part of the pathname from the cluster machines' perspective, omitting trailing slashes. Otherwise, leave a period (.) */home/MetaXpress/*

Note: This module must be the last one in the analysis pipeline.