Adapting *S. cerevisiae* Chemical Genomics for Identifying the Modes of Action of Natural Compounds

by

Kerry Andrusiak

A thesis submitted in conformity with the requirements
for the degree of Master of Science

Graduate Department of Molecular Genetics
University of Toronto

Your file Votre référence
ISBN: 978-0-494-85855-4

Our file Notre référence
ISBN: 978-0-494-85855-4

Adapting *S. cerevisiae* Chemical Genomics for Identifying the Modes of Action of Natural Compounds

Kerry Andrusiak

Master of Science

Graduate Department of Molecular Genetics
University of Toronto

2012

Abstract

Natural compounds have been largely excluded from characterization via high-throughput profiling strategies due to their limited abundance. Herein, I describe the modification of high-throughput yeast chemical genomic (CG) interaction profiling to permit identifying the modes of action of natural compounds. The previous assay proceeded by evaluating the genome-wide yeast deletion collection for drug-hypersensitivity in a volume of 0.7mL. Compound consumption was minimized with the adapted approach by reducing the assay volume 70% through simplifying the complexity of the yeast deletion pool screened. By recreating each yeast mutant in a drug-hypersensitive background, I created a novel resource that increases compound efficiency and further diminishes compound use. Evaluating a series of characterized compounds analyzed previously by the traditional CG approach validated the adaptations incorporated did not negatively affect the quality of data yielded. Ultimately, this modified strategy will be used to screen thousands of natural compounds contained within the RIKEN NPDepo library.

Acknowledgments

First and foremost, many thanks to my supervisors Dr. Charlie Boone and Dr. Guri Giaever for their support throughout my thesis project. I offer my sincerest gratitude to Charlie for investing in me throughout the past two years, providing me with a challenging environment that has undoubtedly shaped me into a better scientist. Your high caliber thinking and research excellence has truly inspired me.

I also thank my committee members Dr. Dan Durocher and Dr. Corey Nislow for their support of this project and their constructive comments at my committee meetings – this advice indisputably helped to develop this thesis into what I present today.

To my collaborators Dr. Jeff Piotrowski and Dr. Marissa LeBlanc at the RIKEN institute in Japan – many thanks for all your hard work. Much of this project would not have been possible without your experimental support and advice. I would also like to thank the head of the RIKEN chemical genomics research group, Dr. Minoru Yoshida, and Charlie for providing me the opportunity to travel to Japan to complete a portion of this project. This was a once in a lifetime experience that will be difficult to match.

To Dr. Chad Myers and his graduate student Raamesh Deshpande – your collaborative efforts on the computational aspects of this project have been invaluable. It has been awesome working with you and thank-you specifically to Raamesh for taking the time to answer my many (and sometimes-repetitive) questions on the ins-and-outs of the computational strategies employed in this thesis.

Throughout this thesis I have had the pleasure to work alongside good company in the Boone lab, from fellow graduate students, post docs, research associated and technical support. I would like to specifically thank past graduate students Owen and Cheuk Hei for their mentorship, and present members Michael, Bryan, Anastasia, Sondra and Pierre for their advice and support with various aspects of this project. I would like to give a special thanks to Elena for your support, your companionship, and your ability to lighten the mood on stressful days. Best of luck to everyone in their future research endeavors.

I would like to acknowledge funding I have received throughout this project from the National Sciences and Engineering Research Council of Canada as a lab grant and as student funding, as well as student funding from the Ontario Graduate Scholarship Program.

Lastly and perhaps most importantly, I would like to thank my family and friends for their unwavering support. To my mom – thanks for your encouragement no matter what path I chose. To Stephen – thanks for pretending you're interested in my experiments and driving me downtown on weekends just to check whether they worked. I am lucky to have such a loving and supporting person in my life; to you I dedicate this work.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

Δ	Gene deletion allele
ABC	ATP-binding cassette
AMB	Amphotericin B
ATP	Adenosine triphosphate
AUPR	Area under precision-recall
BP	Base pair
CG	Chemical genomic
DNA	Deoxyribonucleic acid
DMSO	Dimethyl sulfoxide
gDNA	Genomic DNA
G418	Geniticin
IC_{50}	Inhibitory concentration – 50%
KET	Ketoconazole
LEU	Leucine
MDR	Multi-drug resistance
MET	Methotrexate
MOA	Mode of action
NAT	Nourseothricin
NP	Natural product
NPDepo	Natural products depository
ORF	Open reading frame

OD	Optical density
PCR	Polymerase chain reaction
PDR	Pleiotropic drug resistance
PR	Precision-recall
SD	Synthetic dextrose
SGA	Synthetic genetic array
SGD	Saccharomyces genome database
URA	Uracil
WT	Wild-type
YPD	Yeast peptone dextrose
YPGal	Yeast peptone galactose

1.0 INTRODUCTION

1.1 Chemical genetics

Modern DNA sequencing technology has accelerated the completion of genome sequences from many organisms of interest, ranging from species of pathogenic yeast (Jones et al., 2004; Nierman et al., 2005) to the humans they infect (Venter et al., 2001; Lander et al., 2001). These projects have yielded thousands of predicted genes whose products require functional characterization to better understand the complex molecular biology of each organism. A desire to complete this task efficiently underscores the importance of developing and implementing large-scale functional genomic methodologies.

Comprehending the biological role of a gene's product normally requires a means to alter its function. Traditionally, this has been accomplished via classical genetics, whereby an activating or inactivating genetic mutation is used to alter the expression of a gene product of interest. While this strategy has been successfully exploited in genetically tractable model organisms such as yeast, flies and worms, its use can be problematic in more complex mammalian cells and biological systems less amenable to genetic manipulation (Stockwell, 2000). An increasingly appealing alternative is to directly alter the function of a gene's product using target-specific small molecules with an approach coined chemical genetics (Mayer, 2003; **Figure 1**). In contrast to permanently modifying gene function with a genetic mutation, chemical inhibitors act transiently, largely because their physiological effects are often reversed rapidly due to metabolism and cellular detoxification (Spring et al., 2005). This affords increased temporal and spatial control over the targeted molecule and enables resulting phenotypes to be analyzed independent of indirect compensatory effects that can result from permanent genetic mutations (Zheng & Chan, 2002). The conditional nature of chemical perturbations also permits studying genes that are essential or function during specific stages of development, avoiding complications which can be encountered when creating and studying conditional-essential alleles with genetic-based approaches (Shogren-Knaak et al., 2001). In addition to expanding genes accessible for analysis, severing the reliance on a genetic mutation to probe protein function makes the chemical genetics approach systems independent, especially when the chemical agents used possess cross-cell or species activity. Finally, unlike genetic mutations, chemicals can be used to perturb one function of a multifunctional enzyme to permit better understanding the

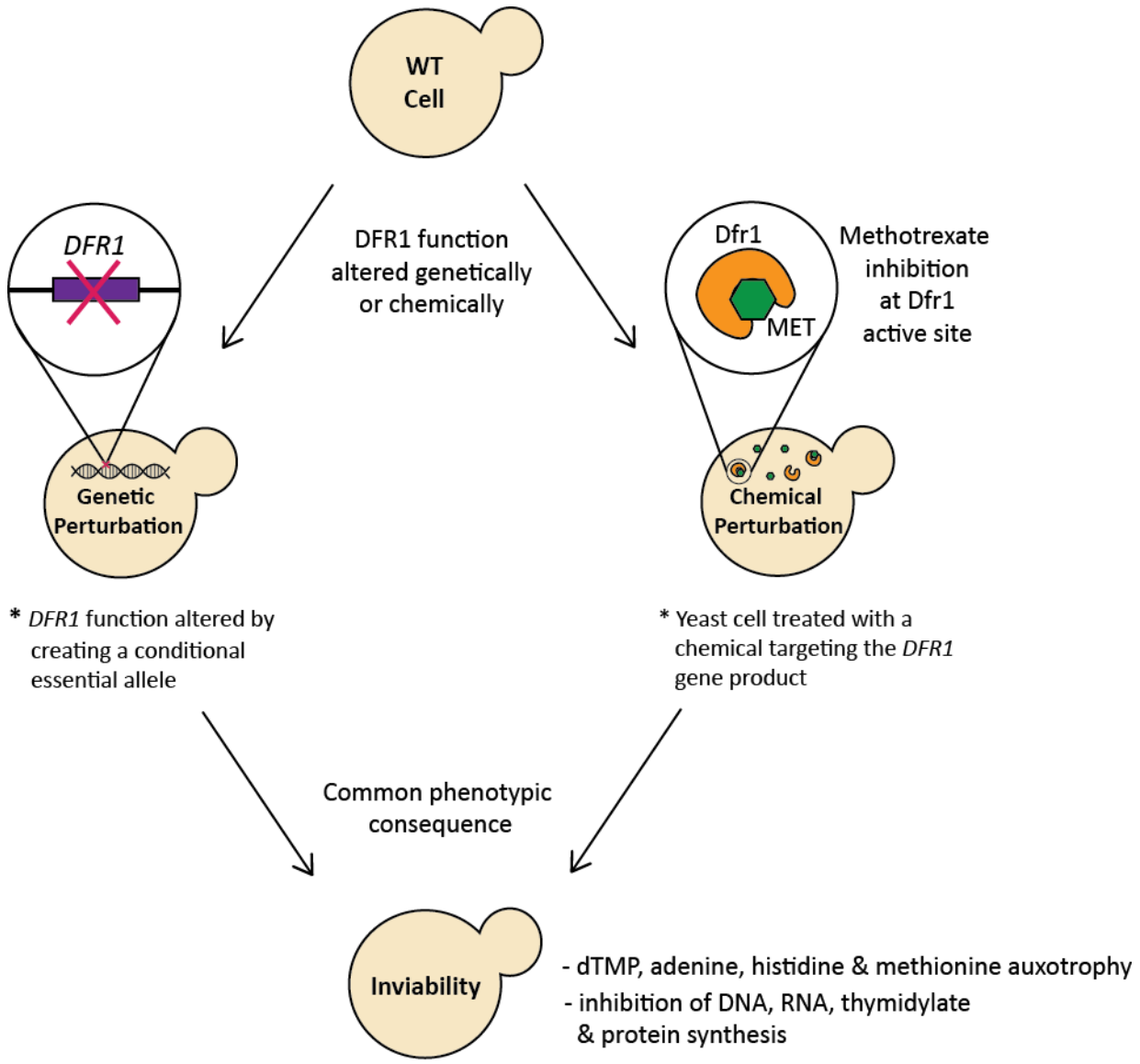

Figure 1. Chemical genetics: an example

Inhibition of a specific gene product using either a genetic or chemical means can have common phenotypic consequences to permit understanding gene function. For example, the essential yeast gene *DFR1* encodes dihydrofolate reductase, an enzyme integral to the biosynthesis of tetrahydrofolic acid. The function of *DFR1* can be altered genetically by, for example, creating a conditional essential allele (at left), or through the use of a chemical agent that targets the *DFR1* gene product (at right) (Huang et al., 1992). One such chemical agent is the drug methotrexate (MET), which is a high affinity Dfr1p antagonist through competitive inhibition at the enzyme active site. Inhibition of *DFR1* by either means causes cell inviability, which results from an auxotrophy for a series of nucleic and amino acids that require the Dfr1p product (tetrahydrofolic acid) as a cofactor in their biosynthesis. Alteration of *DFR1* function by either a genetic or chemical means therefore ultimately inhibits DNA, RNA thymidylate and protein synthesis.

multiple biological roles of different gene products (Kawasumi & Nghiem, 2007). This advantage is evidenced in the study of protein kinases; the phosphotransferase activity of a kinase can be perturbed and thus investigated independent of its protein scaffold activity using kinase inhibitors (Knight & Shokat, 2007).

1.2 Applying chemical genetics genome-wide to probe biological function and identify novel drug leads

The use of chemical ligands to probe gene product function on a genome-wide scale constitutes a chemical genomics (CG) analysis. Results obtained from this strategy can be used to gain important biological insights, or identify compounds that may provide potential pharmaceutical leads. Regarding the former, CG analyses can enable understanding the biological role of gene products in the context of a cellular environment to gain insights into diverse cellular processes. For instance, the well-known fungicide benomyl has been used extensively as a chemical-based microtubule-perturbing agent in cell-cycle research to tease apart the highly regulated microtubule network (Li & Murray, 1991; Sterns et al., 1990). In another important example, the neurotransmitter dopamine was discovered and its role in mediating chemical transmission within the nervous system was identified using the chemical agents reserpine and chlorpromazine as probes of neurobiology (LeDoux, 2002). Furthermore, there are also instances where CGs can exploit chemical ligands lacking direct protein targets to indirectly probe biological process by forcing the cell to respond to an environmental stressor. Such is the case for DNA damaging agents like methyl methanesulfonate (MMS), which have been used to characterize components of the cellular pathways responsible for responding to and repairing DNA damage (e.g. Jelinsky & Samson, 1999; Gasch et al., 2001; Workman et al., 2006; Rooney et al., 2009).

The potential to identify novel therapeutic agents is an equally important motivation for performing chemical genomic analyses. A recent study of FDA approved drugs revealed that only ~400 of the many thousands of predicted proteins encoded by the human genome are currently targeted by chemical intervention (Yildirim et al., 2007). However, given that the Online Mendelian Inheritance of Man database currently reports on more than 3,300 disease-related genes, it is presumed that a much larger number of human proteins will be of therapeutic

significance. Identifying modulators of these 'untapped' molecules that may serve as potential targets could facilitate the development of treatments for the many human diseases that currently lack drug therapies. In addition, there is also a need to expand upon the repertoire of available antibacterial and antifungal compounds to contest the continued emergence of multi-drug resistance microbial pathogens. A CGs approach has the potential to identify chemical agents serving these therapeutic purposes by rapidly screening large compound libraries to identify those that induce a desired physiological response through targeting a disease causing or pathogen specific molecule or pathway. Because this analysis is performed genome-wide, the interrelationships that often occur between a single compound and many gene products can be identified simultaneously and used to identify all potential biological targets. This information is beneficial to eliminate chemicals early on that possess undesirable secondary targets and thus may exhibit unfavorable side effects if used as a drug therapy. However, knowledge of additional drug targets could also be used to prioritize compounds whose combined targets may enhance a desirable effect based on known biological interactions. The development of multi-targeted therapies is an emerging field that shows great promise for treating multi-genic diseases such as cancer, and combating drug resistant pathogens (Zimmermann et al., 2007; Sams-Dodd, 2005; Onyewu & J Heitman, 2007).

1.3 Using cell-based compound profiling assays in modern drug discovery

Identifying compounds that modulate the function of validated targets using a target-oriented approach has been a major avenue employed by the pharmaceutical industry for developing novel drug therapies (Schreiber, 2000). Some of the characteristics relied on when selecting a validated target with which to focus these research efforts include the essentiality of a target, general knowledge of a targets biological role(s), whether the target has been implicated in a specific disease, and the potential 'druggability' of the target, as reviewed in Hopkins & Groom, 2002. However, one prevailing disadvantage of this focused approach is a bias towards only identifying modulators of well-characterized molecules and biological pathways. Additionally, these assays are typically performed *in vitro* and little knowledge is therefore known initially about whether potential lead compounds will exhibit bioactivity in a more complex cellular

environment. Given these downsides, a heavy reliance on the target-oriented strategy has been used as one component of a multifactorial explanation for an observed flat-line in the number of registered new chemical entities over the past few decades, despite a continued financial influx into the pharmaceutical industry (Higgins & Graham, 2009). As a consequence, there has been renewed interest in developing and implementing unbiased cell-based compound screening strategies, such as chemical genomics assays. These approaches are advantageous because compound leads identified are known to possess cell permeability, and were characterized in a cellular environment that favors native confirmation and association with other molecules and cofactors. Additionally, while cell-based assays require detailed follow-up experiments to validate drug-target predictions, they provide a means to rapidly screen large compound libraries to identify chemicals affecting the function of various gene products based on first-pass target predictions. One disadvantage of most cell-based strategies is direct screening in mammalian systems is currently inaccessible. However, well-studied model organisms have provided an ideal test bed for developing these technologies and identifying modulators of gene products that are human or pathogen conserved.

1.4 Budding yeast is a premier model organism in which to study bioactive compounds

The budding yeast *Saccharomyces cerevisiae* is positioned as an ideal model organism for analyzing bioactive compounds with a CGs approach based on several attributes. This unicellular organism possesses a compact genome that consists of ~6000 genes encoded in 12,000 kilobases of DNA sequence, organized among 16 chromosomes (Goffeau et al., 1996). Many *S. cerevisiae* genes are biologically conserved in humans; it has been shown that about half of yeast proteins share at least part of their primary amino-acid sequence with one known or predicted human protein (~2,700 at a BLAST E-value < 10^{-10}; Hughes, 2002). Importantly, among the conserved genes are several which have been implicated in human diseases (Botstein et al., 1997; Tu et al., 2006). For instance, this includes the yeast genes *MSH2* and *MLH1* whose human counterparts have been linked to inherited nonpolyposis colon cancers (Strand et al., 1993), and the yeast DNA helicase *SGS1*, which shares high sequence identity to the human gene connected to Werner's syndrome (Sinclair et al., 1997). Moreover, a number of drugs that target human

proteins also possess activity in yeast through interacting with the corresponding yeast orthologues; several well known examples are reviewed in Cardenas et al., 1999. Specifically, it was demonstrated recently that out of a subset of 1318 known human drug targets, *S. cerevisiae* possesses orthologues for around 30% (Gunnarsson et al., 2008). Budding yeast therefore provides a simple eukaryotic model to study potential human drug targets and identify lead compounds that modulate their function. *S. cerevisiae* also serves is an ideal system for studying antifungal compounds because it shares many genes with *Candida albicans* and *Aspergillus fumigatus* (Jones et al., 2004; Hu et al., 2007), two of the most common opportunistic pathogenic fungi. As such, several of the most commonly used antifungal compounds were originally characterized in *S. cerevisiae*, including Caspofungin, one of the most-recently developed (Douglas et al., 1994; Douglas et al., 1997). Given continued increases in fungal infections worldwide and the emergence of new pathogenic fungi (Richardson, 2005), *S. cerevisiae* provides a safe, non-pathogenic test bed for studying fungal-specific targets to identify novel antifungal compounds.

From a technical standpoint, *S. cerevisiae* possesses many desirable characteristics that facilitate CG analyses. As the first eukaryotic organism sequenced, *S. cerevisiae* has been the focus of extensive molecular biology research that has lead to the annotation of ~75% of all predicted yeast ORFs. This feature can aid in interpreting the results of CG studies, as it is likely the native biological function of many predicted drug targets will be known. Importantly, this information is also easily accessible due to the creation of comprehensive online resources, such as the Saccharomyces Genome Database (SGD) (Dwight et al., 2004). In addition to rapid and inexpensive culturing, *S. cerevisiae* can be easily manipulated in either a haploid or diploid state, permitting genes to be added, deleted or tagged through homologous recombination (Goffeau et al., 1996). This genetic manipulability has enabled the creation of many valuable *S. cerevisiae* genomic tools, including a genome-wide deletion collection constructed by an international consortium of laboratories. This mutant set consists of over 20,000 strains created in diploid heterozygous or diploid/haploid homozygous backgrounds and covers 96.6% of all predicted ORFs (Giaever et al., 2002). Each deletion mutant is tagged with two unique 20mer nucleotide sequences flanked by priming sequences common to all strains (**Figure 2**). These 'molecular barcodes' allow the fitness of all deletion mutants to be rapidly assessed in parallel in response to

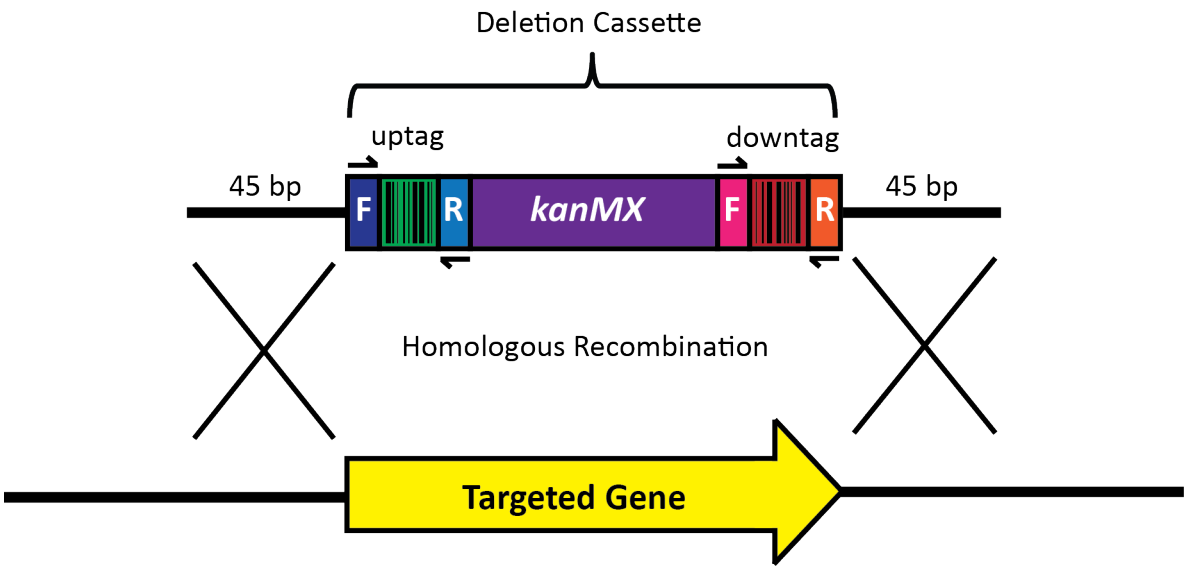

Figure 2. Strategy used for constructing mutants in the *S. cerevisiae* deletion collection.

Each yeast ORF was replaced start codon to stop codon with a deletion cassette consisting of the kanamycin resistance marker (*kanMX*) and two unique 20mer nucleotide barcodes (uptag and downtag) flanked by primer sites common to all deletion strains (indicated by half arrows). 45 bp of sequence, homologous to regions up- and down-stream of the targeted gene, were incorporated onto each deletion cassette to facilitate gene deletion via homologous recombination.

a variety of environmental conditions (e.g. Giaever et al., 2002; Giaever et al. 2004; Lee et al., 2005; Hillenmeyer et al., 2008; Ericson et al., 2008). This feature enabled the development of several yeast high-throughput CG assays, which aim to infer a compounds MOA or molecular target(s) by identifying the yeast strains that exhibit altered sensitivity to compounds of interest (reviewed in Hoon et al., 2008; Chan et al., 2009; Smith et al., 2010; Ho et al., 2011). The remainder of this thesis will focus on one such strategy, known as *S. cerevisiae* CG interaction profiling.

1.5 Characterizing bioactive compounds using *S. cerevisiae* chemical genomic interaction profiling

Chemical genomic interaction profiling seeks to understand a compounds MOA by screening the non-essential yeast deletion collection for drug-hypersensitivity to identify drug-gene interactions (Parsons et al., 2006). Within biological circuits there is genetic redundancy that functions to buffer organisms from both genetic and environmental perturbations – a feature emphasized by the observation that only ~20% of S. cerevisiae ORFs are essential for haploid viability in rich media (Giaever et al., 2002). Applying a bioactive compound to yeast cells harboring a gene deletion may overcome this genetic redundancy and establish drug-induced conditional essentially if the compound's target interacts with the deleted gene product. By profiling these chemical genetic interactions genome-wide, a compressive interaction index or profile can be established to provide important clues regarding the mechanism of drug action. As described in subsequent paragraphs, testable MOA hypotheses can be formulated from these genome-wide profiles by comparison to an interaction knowledgebase.

To facilitate the high-throughput characterization of large compound libraries, all ~5000 non-essential deletion mutants are pooled and CG interactions are evaluated in a highly parallel manner. This is achieved by first competitively growing the mutant pool in the presence of a drug at a sub-lethal concentration. A series of experimental steps are then employed to quantify the relative abundance of each strain's molecular barcode as a proxy for strain fitness or drug sensitivity (**Figure 3A**). This approach often identifies hundreds strains sensitive to each drug screened and this data is compiled to construct genome-wide CG interaction profiles. Although this pipeline cannot directly identify a drug's target based on its absence from the strain

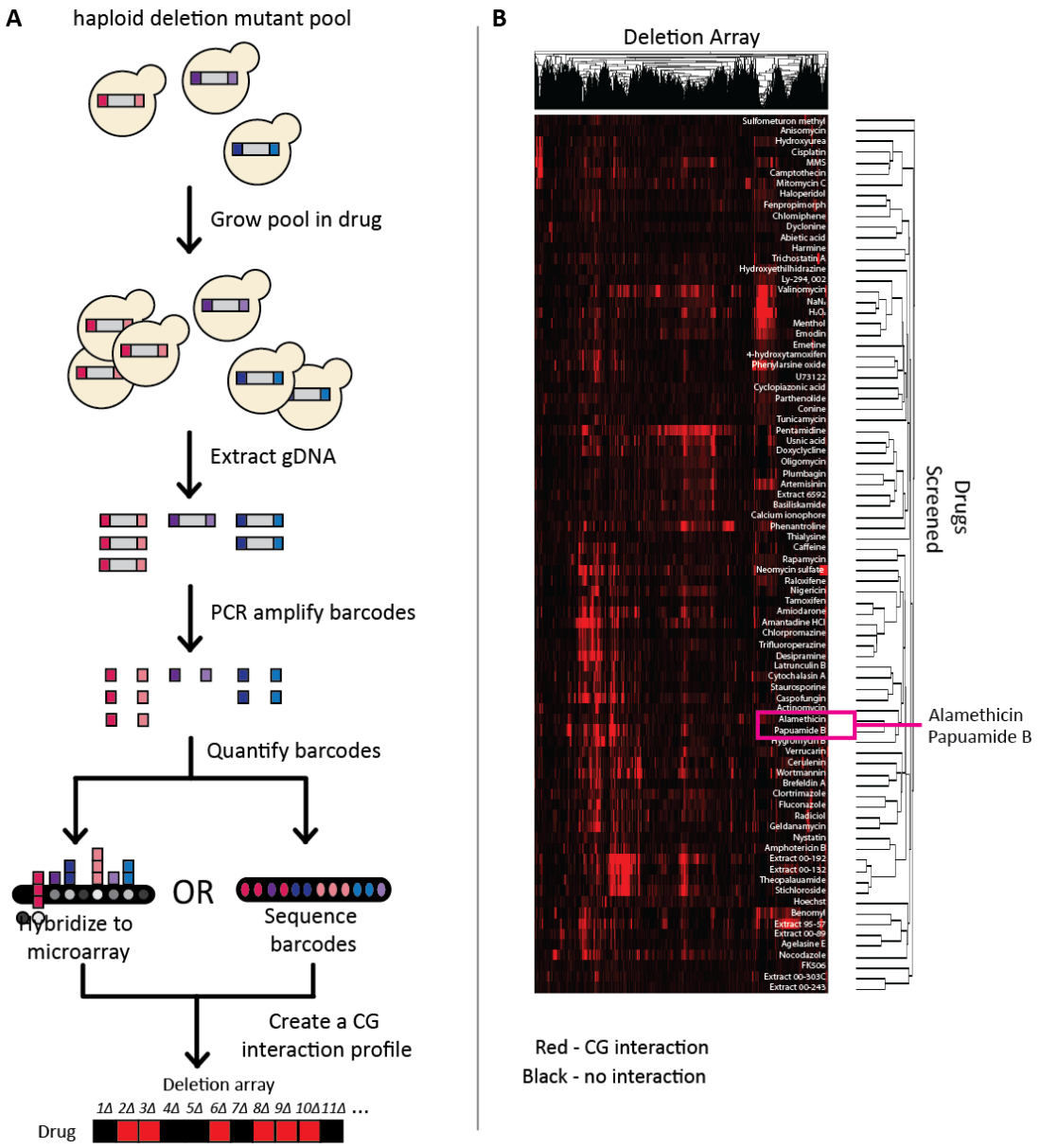

Figure 3. *Saccharomyces cerevisiae* CG interaction profiling.

A) CG profiling in *S. cerevisiae* proceeds by growing a pool of all yeast deletion mutants in the presence of a drug at a sublethal concentration (e.g. IC_{80}). Strains deleted for genes required to respond to the applied drug will exhibit increased sensitivity and become depleted from the pool. To quantify the relative abundance of all strains, genomic DNA is extracted, the strain-specific barcodes are PCR amplified by primers targeting the common priming sites, and barcode abundance is measured by hybridizing to a custom DNA microarray or more recently, through next generation sequencing. Data compiled from this experiment is used to construct CG interaction profiles that summarize the drug-sensitivity of all deletion strains examined. Figure adapted from Smith et al. (2010). **B)** Two-dimensional hierarchical clustering of individual CG interaction profiles will group drugs with common MOAs or drugs targeting related pathways. For instance, the grouping of papuamide B next to alamethecin among a clustergram of 82 total CG interaction profiles provided evidence that these drugs functioned through a common MOA: disruption of the cell membrane. Figure B created using CG data from Parsons et al. (2006).

collection, the drug-gene interactions can identify gene products that belong to the targeted pathway, buffer the targeted pathway and are important for cellular function when the targeted pathway is compromised, or products that are involved in general drug metabolism, transport or stress response (Ho et al., 2011). This information is particularly useful to understand the biological function of compounds that lack direct protein targets and function through forcing the cell to respond to a drug-induced environmental stress, such as increased cellular permeability or DNA damage. For example, while many existing chemotherapeutic agents lack a direct protein target and function generally through eliciting DNA damage, the CG interactions identified by screening these compounds are often highly informative and enriched for members of the DNA repair pathway (Birrell et al., 2001; Chang et al., 2002; Lee et al., 2005; Workman et al., 2006; Yu et al., 2008). CG interaction profiles can be further analyzed to infer a compounds MOA using a guilt-by-association rationale. Compounds that function through related MOAs and exhibit comparable biological activity often yield similar CG interaction profiles (Parsons et al., 2004; Lee et al., 2005; Brown et al., 2006; Parsons et al., 2006). Consequently, global clustering of CG interaction profiles will group drugs based on common MOAs or targets (**Figure 3B)**. This comparison is highly informative when the profiles of well-characterized drugs that function through known and diverse MOAs are included. For instance, a systematic comparison of CG interaction profiles produced by 82 compounds revealed the previously novel natural product papuamide B grouped closet to alamethichin, an antibiotic that binds to lipids and disrupts the cell membrane (Parsons et al., 2006). This similarity suggested papuamide B functioned through a related MOA, which was later corroborated using a series of requisite follow-up experiments to identify the phospholipid phosphotidylcholine as the target. While the guilt-by-association analysis of CG interaction profiles requires detailed biochemical follow-up experiments to validate MOA or target predictions, it is a rapid and relatively straightforward method to generate a testable hypothesis. The follow-up experiments required to explore these hypotheses depend on the predictions formulated, but could include creating and analyzing the sensitivity of specific deletion mutants to drugs known to function through a related MOA, or using function-specific assays to test whether the compound alters the predicted biological process. Regardless, the identification and analysis of CG interaction profiles provides a powerful means for gaining insight into the MOA of novel bioactive compounds.

1.6 Integrating chemical genomic and genetic interaction data to facilitate drug-target prediction

Integrating CG data with the growing genetic interaction knowledgebase provides an alternative analysis strategy for identifying the pathways and molecules targeted by bioactive compounds. As defined by the multiplicative model, a genetic interaction occurs when the fitness of a double mutant deviates from the expected multiplicative effect of the two single mutants (Mani et al., 2008). Specifically, a synthetic lethal interaction results when two nonessential genes that impinge upon a common essential function are deleted in a single strain (Tong et al., 2004). Chemical genetic interactions can mimic this phenomenon when a drug targets the product of one gene participating in a synthetic lethal interaction pair (**Figure 4A**). Consequently, the deletion strains that exhibit sensitivity to a given drug possess mutations in genes that should participate in synthetic lethal interactions with the gene(s) encoding the drugs target(s) (Parsons et al., 2004). The CG interaction profile of a bioactive compound should therefore resemble the synthetic lethal interaction profile produced by its targeted molecule or molecules belonging to the targeted cellular process (**Figure 4B)**. Parsons et al., (2004) first established the validity of this concept by showing the CG interaction profiles of several compounds were highly similar to the genetic interaction profiles of genes belonging to biological processes they were known to target. These similarities could be visualized based on profile grouping using a two-dimensional hierarchical clustering analysis. In a more recent example, the anti-fungal activity of the previously uncharacterized compound, named Erodoxin in the study, was linked to inhibition of the *ERO1* gene product based on a hypothesis formed initially using a comparison of interaction data across experiments (**Figure 4C**; Costanzo et al., 2010). While this comparison strategy also requires follow-up experiments to validate drug target predictions, it provides a powerful method to rapidly interpret CG interaction profiles that are often complex, owing to the hundreds of strains commonly sensitive to a given compound. One present limitation of this approach is not all yeast genes have been screened for digenic interactions and a full genome-wide comparison is not yet possible. However, a significant proportion of interactions that compressively cover all biological processes have been analyzed (Costanzo et al., 2010) and it is expected that data for remaining interactions will be compiled in the near future.

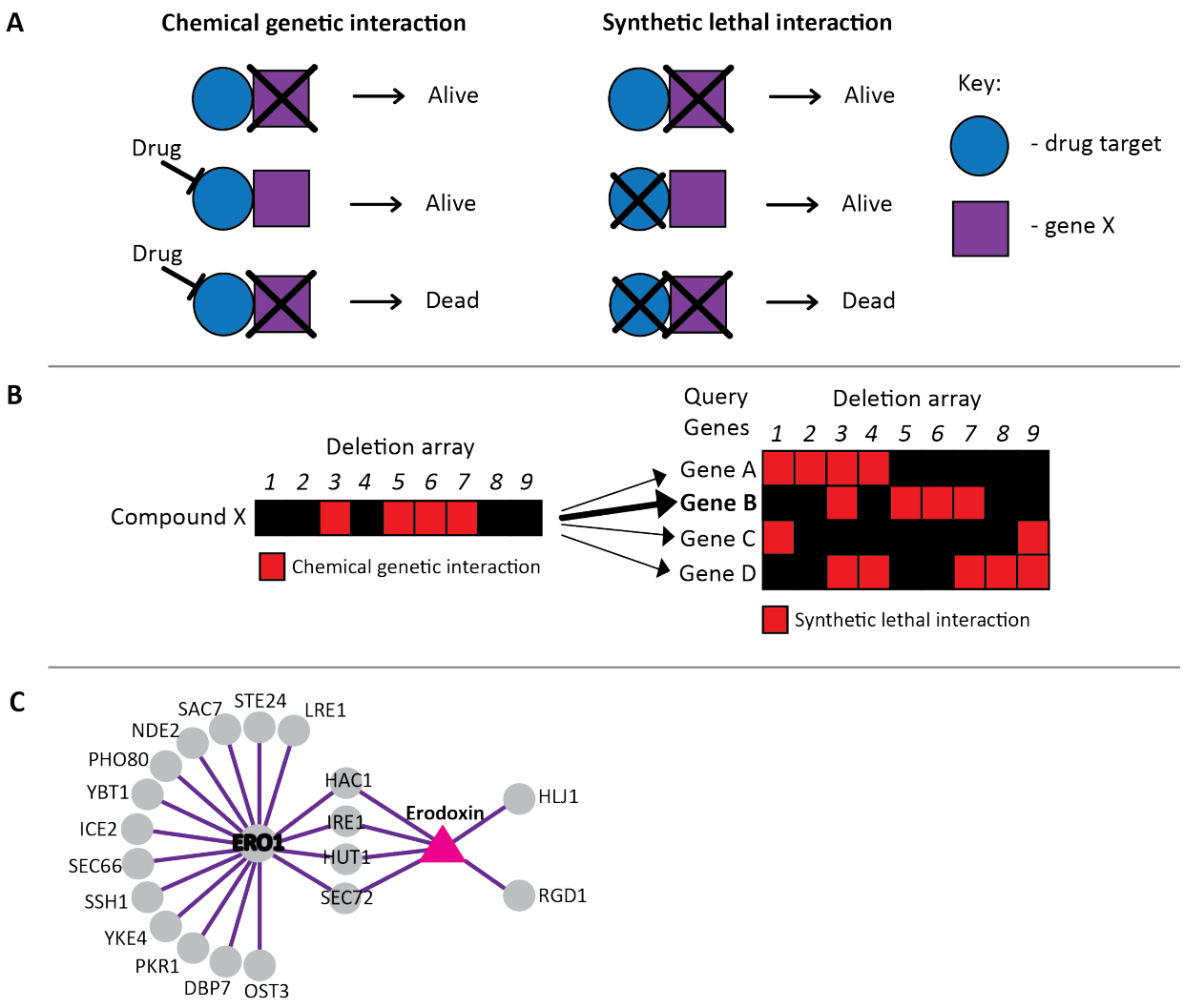

Figure 4. Gaining insight into a compounds biological target using a global comparison of CG and synthetic lethal interaction data.

A) Chemical genetic interactions mimic synthetic lethal interactions. In a chemical genetic interaction, a strain deleted for gene x is hypersensitive to a sublethal concentration of drug (at left). In the corresponding synthetic lethal interaction, a strain deleted for both the drug target and gene x is inviable (at right). Gene deletions that exhibit chemical genetic interactions should therefore exhibit synthetic lethality with the gene encoding the drug's target. **B)** The CG interaction profile of a compound should resemble the synthetic lethal interaction profile produced by the targeted molecule or molecules participating in the targeted pathway. For example, the CG interaction profile of compound X resembles the synthetic lethal interaction profile of Gene B, suggesting the product of Gene B is the putative target. Figures **A)** and **B)** were adapted from Parsons et al. (2004). **C)** The overlap of genes exhibiting synthetic lethality to *ERO1* with the strains exhibiting sensitivity to the previously uncharacterized compound Erodoxin was one component of evidence used to form the hypothesis that erodoxin targeted the *ERO1* gene product. Adapted from Costanzo et al. (2010).

1.7 The role of multidrug resistance in *S. cerevisiae* chemical genomic interaction profiling

Multidrug resistance is the generalized lack of sensitivity to a broad spectrum of functionally and structurally unrelated compounds. The overexpression of membrane transporters dedicated to ATP-dependent drug efflux is a major contributor to this phenotype (Chang, 2003). In yeast, there exists an extensive pleiotropic drug resistance (PDR) network, consisting of many ABC transporters and several transcription factors that regulate transporter expression (reviewed in Bauer et al., 1999; Jungwirth & Kuchler, 2006; Kuchler & Schuller 2007). This natural drug resistance machinery can be problematic for *S. cerevisiae* CG interaction profiling, as it de-sensitizes the cell to the compounds exogenously applied for characterization and increases the concentration required to induce a physiological response. The cell's PDR response can also complicate CG interaction profiles generated from this analysis, as many strains deleted for drug-resistance machinery can be identified as drug-sensitive even though they are not directly related to the MOA of the compound screened (Parsons et al., 2004; Hillenmeyer et al., 2008). The components of the yeast PDR network are not essential for cell viability and they can therefore be deleted to diminish drug-resistance and increase sensitivity to a variety of compounds (Kolaczkowski et al., 1998; Michalkova-papajova et al., 2000; Rogers et al., 2001). Recently, this strategy was employed to construct a drug-hypersensitive overexpression library in the fission yeast *Schizosaccharomyces pombe* (Arita et al., 2011). As a result of deleting two *S. pombe* drug resistance genes, each strain had increased chemical sensitivity and drugs could be analyzed at lower concentrations to conserve resources. While a similar drug-sensitized strain collection has yet to be completed in budding yeast, it is expected that deleting PDR genes would be an ideal avenue to achieve this goal and overcome the current complications associated with multi-drug resistance in CG interaction profiling.

1.8 Analyzing synthetic versus natural product compound libraries

Given the theoretical scope of organic small molecule chemical space ($1 \times 10^{60} – 1 \times 10^{200}$; Tan, 2005), there are an immense number of potential compounds that could be screened with *S. cerevisiae* CG interaction profiling. The selection of which chemicals are investigated has a large impact on the quality of biological insights gained from the data and potential drug-leads

identified. As such, extensive analysis has been performed to identify the biochemical and biophysical characteristics that relate to important drug-like features, such as solubility, cell permeability, bioavailability and toxicity (Lipinski, 2000; Lipinski et al., 2001; Kerns & Di, 2009; Mishra et al. 2009). An extension of this analysis is the comparison of compounds from synthetic or natural origins, which has initiated an on-going debate into which class of compounds has greater biological merits and should be the focus of pharmaceutical profiling efforts. The following paragraphs briefly summarize the main features of both compound classes and highlights why natural products are an ideal compound source to study with yeast CG analyses.

1.8.1 Synthetic compounds

Since the inception of combinatorial chemistry in the 1990s, synthetically produced compounds have been central to many drug-based functional genomic analyses, particularly within the pharmaceutical industry (Koehn & Carter 2005; Li & Vederas, 2009). This has resulted partly because these compounds are inherently compatible with high-throughput drug–characterization approaches. Specifically, synthetic compounds are pure, possess a known molecular structure and can be obtained relatively inexpensively, often as components of large, commercial synthetic compound libraries. Furthermore, the compounds contained within these synthetic-based libraries often adhere to the drug-like physical properties described by Lipinski's 'Rule of Five' to increase the potential of identifying ideal therapeutic agents. These characteristics include that compounds should possess low liphophilicity, have fewer than 5 hydrogen bond donors and fewer than 10 hydrogen bond acceptors, and possess a molecular mass of less then 500 Daltons (Lipinski et al., 2001). One prevailing disadvantage of focusing exclusively on analyzing synthetic compounds is the lack of structural diversity that can currently be achieved by synthesis techniques, despite advances in combinatorial chemistry (Dolle et al., 2005). This limited diversity has been used as another component of the multifactorial answer used to explain the flat-lining number of registered new chemical entities over the past decade, as mentioned in section 1.3 (Rouhi & Washington, 2003; Lam, 2007). Therefore in recent years, there has been some renewed interest in reverting research efforts back to the characterization of natural-based compound, as was prior to the synthetic compound era (Galm & Shen, 2007).

1.8.2 Natural products

Natural products (NP) are selected for precise biological functions in nature and are therefore a rich source of specific small molecule inhibitors. These compounds are typically derived from terrestrial micro-organisms (e.g. the penicillin antibiotics from *Penicillium* fungi), plants (e.g. the pain-killer morphine from *Papaver somniferum*) and more recently, marine organisms (e.g. the breast cancer drug Halaven from *Halichondria okadai*). Throughout history, NPs have been the basis of a significant proportion of therapeutic agents; ~80% of all drugs prior to 1990 were either natural products or natural product inspired analogs (Li & Vederas, 2009). In particular, they have been predominantly influential in the fields of cancer treatment, immunosuppression and antibacterial therapies (Newman & Cragg, 2007). In addition to inspiring this large percentage of marketed drugs, within the last three decades over 50% of all new registered chemical entities possessed a natural origin (Newman & Cragg, 2007). The biological value of NPs, as evidenced by these statistics, can be attributed to their intrinsic cell permeability, their inherently specific yet varied biological targets and their immense chemical diversity. Compared to synthetic compounds, NPs populate a larger proportion of chemical space, which overlaps with the structural diversity observed among existing successful therapeutic agents (Feher & Schmidt, 2003). In spite of these characteristics and their overwhelmingly favorable success, the use of natural-based compounds by many pharmaceutical companies has declined based on their incomparability with high-throughput screening approaches (Harvey, 2008). This is largely because NPs are often difficult and expensive to obtain and are therefore available in very limited quantities. Furthermore, when crude natural product extracts are first acquired, they are often composed of highly complex chemical mixtures that require fractionation prior to analysis (Roemer et al., 2011). This added step can be both time and resource consuming when performed on an industrial scale (Roemer et al., 2011). However, there have been advances in this field and diverse chromatographic separation procedures are available (Colegate & Molyneux, 2008). Ultimately, the potential benefits of analyzing NPs outweigh the costs and refocusing high-throughput screening efforts to NP characterization will likely yield myriads of rich biological information and many potential pharmaceutical leads.

1.9 Project rationale

The utility of yeast CGs for characterizing bioactive compounds and a desire to extend this

analysis to natural products has been the impetus for my thesis research. This project has the long-term goal of characterizing a relatively new and understudied collection of NP and NP-based compounds contained within the Natural Products Depository (NPDepo) held at the RIKEN institute in Japan. This valuable resource consists of over 40,000 structurally diverse, low molecular weight compounds isolated from a multitude of natural sources, including bacterial fermentations, secondary metabolites of fungi and compounds extracted from plants and marine organisms. One significant asset of this collection is that each compound is pure and possesses a known structure cataloged in the NPEdia database (Tomiki et al., 2006). This resource also provides detailed information on the compounds origin and biological and physical properties, which will ultimately facilitate interpreting the results obtained following their analysis. However like most natural products, the compounds contained within this library are limited in quantity and are therefore incompatible for screening with the current yeast CG interaction profiling strategy.

To establish yeast chemical genomics as a powerful approach for characterizing low-abundance NPs, this thesis had two main objects. First, I sought to modify the current CG profiling strategy to reduce the quantity of each compound required for analysis. This entailed selecting a diagnostic subset of yeast deletion mutants to screen and regenerating each mutant in a novel drug-hypersensitive background. Second, I endeavored to demonstrate the utility of the adapted assay by screening a series of previously characterized compounds and comparing the results to those obtained previously using a 'traditional' CG strategy. Validating that informative biological information can be obtained using the adapted approach will pave the road for analysis of the valuable RIKEN NPDepo library. These studies have involved collaborations with Jeff Piotrowski and Marissa LeBlanc, two post-doctoral fellows at the RIKEN institute in Japan and Raamesh Deshpande, a computational scientist from Chad Meyer's lab at the University of Minnesota.

2.0 RESULTS

2.1 Overview: A high-throughput chemical genomics assay for the analysis of natural products

To characterize valuable NP compound libraries, I developed a modified CG interaction profiling strategy to reduce the amount of each NP required for analysis (**Figure 5**). Specifically, the volume of each experiment was miniaturized to 0.2mL. This change required selecting a diagnostic subset of viable deletion mutants to screen for compound sensitivity in lieu of the traditionally used genome-wide collection to maintain adequate representation of each strain in the initial inoculum. This adaptation also permitted scaling experiments to a 96-well format to increase the throughout of compound analysis. The chosen deletion mutants were then regenerated in a drug-hypersensitive background deficient for natural drug-resistance machinery. By diminishing the concentration of compound required to inhibit cell growth, this feature will further reduce the quantity of each NP screened. Furthermore, to quantify strain abundance following each pooled growth experiment, a recently developed 96-plex next-generation sequencing strategy was adopted (Smith et al., 2010b), replacing the previously used microarray strategy. Once completed, this modified assay was used to screen a series of previously characterized compounds, demonstrating the described adaptations do not affect the quality of CG data yielded from screening bioactive compounds with this approach.

2.2 Selection of a diagnostic subset of *S. cerevisiae* deletion mutants to miniaturize pooled growth analysis

With the existing high-throughput *S. cerevisiae* chemical genomic assay, a pool comprised of the entire collection of viable deletion mutants is screened for compound sensitivity in a culture volume of 0.7mL, in 48-well format (Pierce et al., 2007). Given the initial culture is inoculated with the pool at an OD_{600} of 0.06, this strategy enables each strain to be represented at least 300 times in the initial inoculum to reduce noise in generated CG data that can result from sampling errors (Pierce et al., 2007). For my modified assay, I sought to miniaturize this approach by decreasing the volume in which the pooled growth analysis is performed to 0.2mL. Assuming the same concentration of drug is maintained, this 70% reduction in volume affords a 3.5 fold

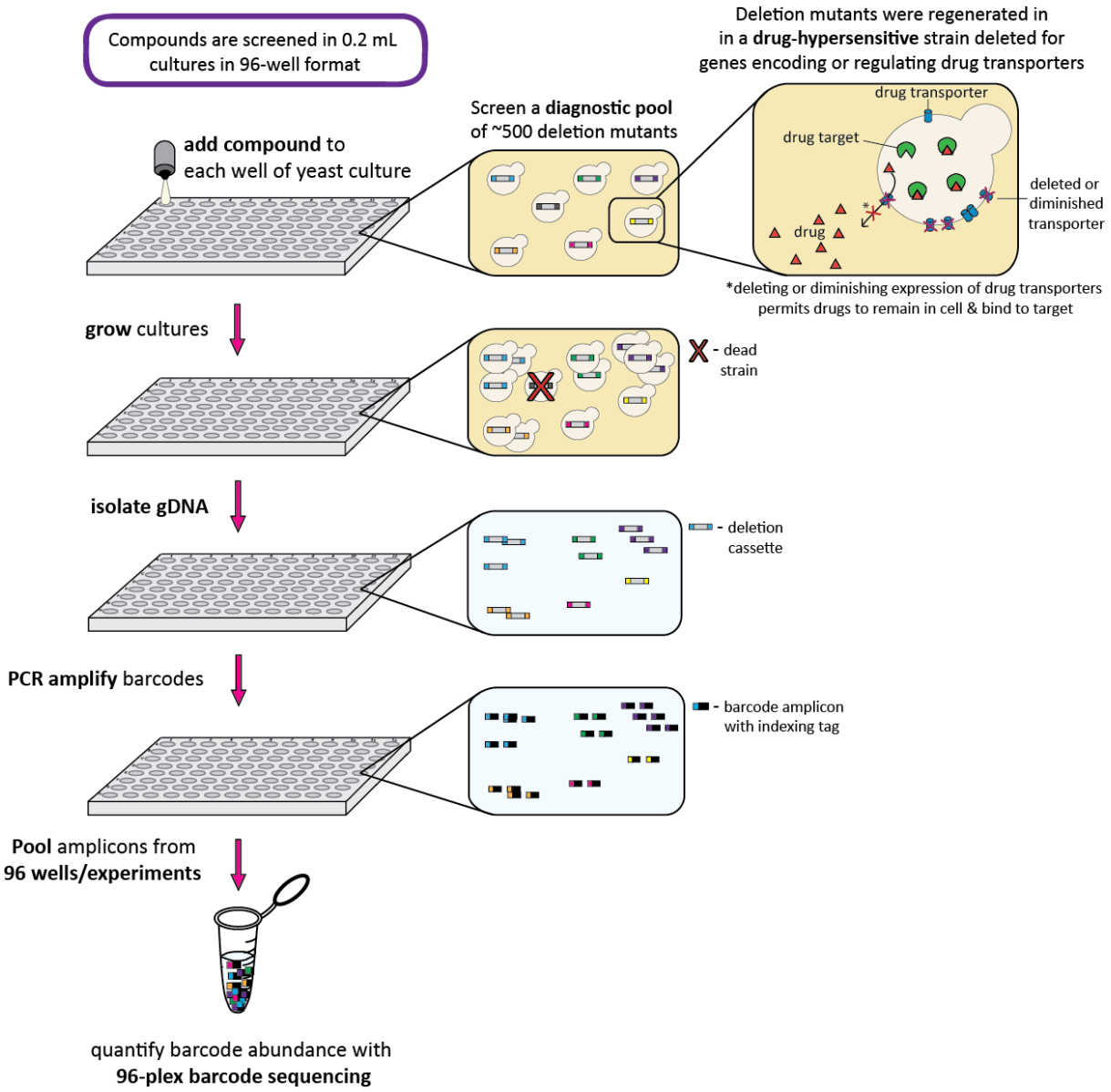

Figure 5. A modified chemical genomic profiling assay for characterizing low abundance natural products.

To reduce the quantity of each compound required for analysis and permit screening in a 96-well format, pooled growth experiments are performed in a reduced volume of 0.2mL. A diagnostic subset of ~500 deletion mutants was selected to constitute the mutant pool analyzed to facilitate this change. Each of the diagnostic mutants was also regenerated in a drug-hypersensitive background, deleted for PDR drug-transport related genes to further reduce the quantity of NP required. After adding a NP to each well inoculated with the novel pool of deletion mutants and growing the cultures for 5 generations, the relative abundance of each strain is quantified by first isolating genomic DNA and then PCR amplifying the unique barcodes, as previously described (Pierce et al., 2007). However, modified primers are used to add experiment-specific indexing tags to the amplicons from each well to permit pooling the PCR products from 96 experiments. Barcode abundances from each experiment are then quantified using a previously described 96-plex next generation sequencing approach (Smith et al., 2010b).

reduction in the quantity of each NP required for analysis. Additionally, screening at a volume of 0.2mL is compatible with analysis in a 96-well format to double the scale of compounds screened at once relative to analysis in the current 48-well format. However, to maintain an adequate coverage of each strain in the initial inoculum, the complexity of the mutant pool evaluated had to be reduced. I reasoned that given the following equation, reducing the complexity of the mutant pool screened 10-fold to around 500 strains would ensure a more then ideal coverage of each strain could be maintained.

$$cells\ per\ strain = starting\ OD \times culture\ volume \times \left(\frac{cells\ per\ mL}{number\ of\ strains}\right)$$
$$= 960$$

When: *number of strains = 500*
starting OD = 0.06
culture volume = 0.2mL
cells per mL = 4 x 10^7 (measured for the haploid strain used)

My main objective when selecting this subset of deletion mutants was to pick strains that participated in strong, diagnostic interactions. This would enable the CG interaction profiles generated by assaying the streamlined collection to cluster with other interaction data as well as if the entire collection of 5000 mutants were analyzed. To make predictions on which genes would serve this purpose, I utilized a computational-based approach, developed in collaboration with Raamesh Deshpande, a student from Chad Myers lab at the University of Minnesota, to analyze our lab's compendium of genetic interaction data. Importantly, the genetic interactions contained within this data set provide a comprehensive coverage of most cellular processes (Costanzo et al., 2010). I also employed a complementary, manual-based approach to expand upon the computationally selected genes and add functional redundancy to the diagnostic gene list. These two approaches are described in detail in the subsequent paragraphs.

Computational Gene Selection. To select the most informative array genes present in our genetic interaction dataset, an approach coined the simple greedy algorithm was developed to assess the ability of each array gene to individually re-group interacting query genes correctly relative to a published GO co-annotation gold standard (Myers et al., 2006; logic depicted in Figure **6A**). This comparison was quantified using a precision recall analysis to generate an AUPR curve score for each array gene (see Materials and Methods). After assigning an AUPR

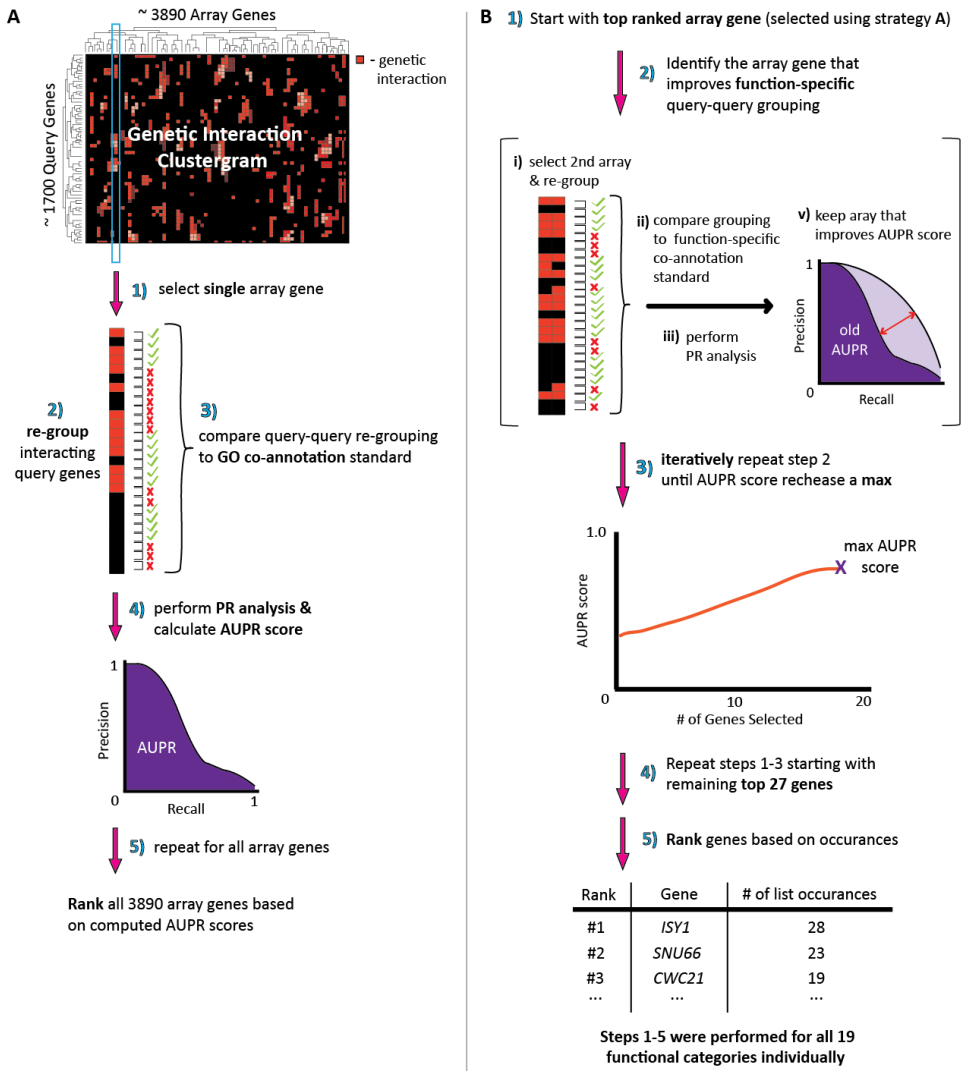

Figure 6. Computational strategy for selecting a diagnostic subset of yeast deletion strains.

A) A simple greedy algorithm was developed to evaluate the ability of each array gene to informatively re-group interacting query genes. Conceptually, this strategy proceeded by: **1)** selecting an individual array gene, **2)** re-grouping query genes based on genetic interaction scores, **3)** comparing the query-query re-grouping to a GO co-annotation standard, **4)** quantifying the comparison using a PR analysis to calculate an AUPR curve score and, **5)** repeating steps 1-4 for the remaining 3879 array genes and finally, ranking all genes based on their AUPR curve score. **B)** The exhaustive greedy algorithm was developed to address shortcomings of the simple greedy algorithm. Conceptually, this strategy proceeded by: **1)** selecting the top ranked array gene for each of the 19 different functional categories by repeating A) using a function-specific GO co-annotation standard, **2)** identifying the second array genes that could improve query-query re-grouping specific to each of the 19 different functional processes by assessing all pairwise combinations with the top array and retaining the gene which could improve the AUPR curve score (i-v), **3)** iteratively repeating step 2 until the AUPR curve scores reach a maximum, **4)** repeating steps 1-3 using the next 27 top array genes identified by the simple greedy algorithm and, **5)** ranking the genes present in the 28 lists based on the number of instances which they were identified. *Steps 1-5 of strategy B) were performed individually for the 19 functional categories.

curve score value to all ~3890 array genes, a ranked gene list was formulated to identify the top performing array genes. Inspection of this list revealed the top ranked array genes were representative of a limited number of biological processes and were therefore not diagnostic of the entire cellular landscape. I reasoned that one factor which likely contributed to this result was the approach focused on assessing array genes in isolation, neglecting to evaluate how combinations of array genes would perform. In other words, two genes that may yield the top AUPR scores when evaluated independently may not enhance query-query grouping when combined, and screening both would therefore be redundant. Additionally, the limited biological processes represented among the top genes in the initial list suggests a possible bias in how informative genes are from different processes, highlighting the value of selecting genes from each process separately to ensure comprehensive representation. These issues were addressed with the development of an alternative computational algorithm loosely based upon a matching pursuit algorithm (Mallat & Zhang, 1993), coined the exhaustive greedy approach. This strategy proceeded by first subdividing the GO co-annotation gold standard into 19 different previously defined functional categories (Costanzo et al., 2010) and repeating the simple greedy algorithm to generate 19 function-specific ranked lists. Using the top ranked array gene selected for each functional category, the exhaustive greedy algorithm then identified which array genes should be added to improve the function-specific re-grouping of interacting query genes, as measured by an increase in the calculated AUPR curve score (**Figure 6B**). The algorithm then iteratively repeated this analysis and expanded the gene lists for each functional category until the AUPR curve score could not be further improved with an additional array gene. To account for possible variations resulting from the use of different starting array genes, the algorithm was initiated using the top 28 ranked genes from each functional category that were identified by the simple greedy approach. The exhaustive greedy algorithm therefore yielded 28 different gene lists for each of the 19 functional categories. The genes selected for each of the 19 functional categories were then ranked based on the number of instances they occurred in the 28 lists. After filtering out uncharacterized or un-annotated genes, I generated a list of 112 diagnostic genes by selected the top occurring genes for each category (**Appendix A**).

Manual Gene Selection. Analogous to the computational-based approach described above, I manually selected a second set of deletion strains driving query-query clustering across different biological processes. This enabled me to increase the functional redundancy of the

computationally selected gene list, as the algorithm could only select a handful of genes from each biological process due to the AUPR curve metric rapidly saturating. To manually select a diagnostic set of genes, I visually inspected the genetic interaction clustergram and selected the array genes that appeared to contribute to the formation of defined clusters (**Figure 7**). A complete analysis of the ~1700 query gene by ~3890 array gene genetic interaction clustergram yielded a list of 442 genes (**Appendix A**). This list included 56% (63/112) of the 112 genes selected computationally. Therefore in total, I chose a diagnostic subset of 491 unique array genes for the adapted CG assay, which ultimately permitting me to reduce the pooled growth assay volume the desired 70%.

2.3 Validating the chosen 491 diagnostic yeast deletion mutants by analysis of interaction data

To assess the diagnostic capacity of the chosen 491 deletion mutants, I wanted to first evaluate whether genetic interaction profiles restricted to the 491 array genes would cluster informatively, relative to clustering with complete profiles. To achieve this aim, three unique two-dimensional clustergrams of the genetic interaction data produced by our lab were generated (see Materials and Methods). The first clustergram was created using genetic interactions involving the entire set of ~3890 array genes, the second was created by restricting genetic interactions to the chosen 491 array genes and the third was created by restricting genetic interactions to 491 randomly chosen array genes. By comparing the query-query grouping in each clustergram to the GO co-annotation gold standard (Myers et al., 2006) and using a PR analysis to evaluate the comparison, the quality of query-query grouping in each clustergram was assessed. When this analysis was performed using a global GO co-annotation standard covering all biological processes, the clustering of genetic interaction profiles restricted the chosen 491 genes was comparable to the clustering observed using complete profile (**Figure 8A**). Furthermore, both the complete and diagnostic array genes clustered interacting query genes more informatively than an equally sized number of random genes, as indicated by an obvious reduction in the PR line yielded by the random gene set (**Figure 8A**). To extend this analysis, subdivided GO co-annotation gold standards were used to assess function-specific clustering for the 19 previously defined functional categories (Costanzo et al., 2010). Additionally, the differences between PR

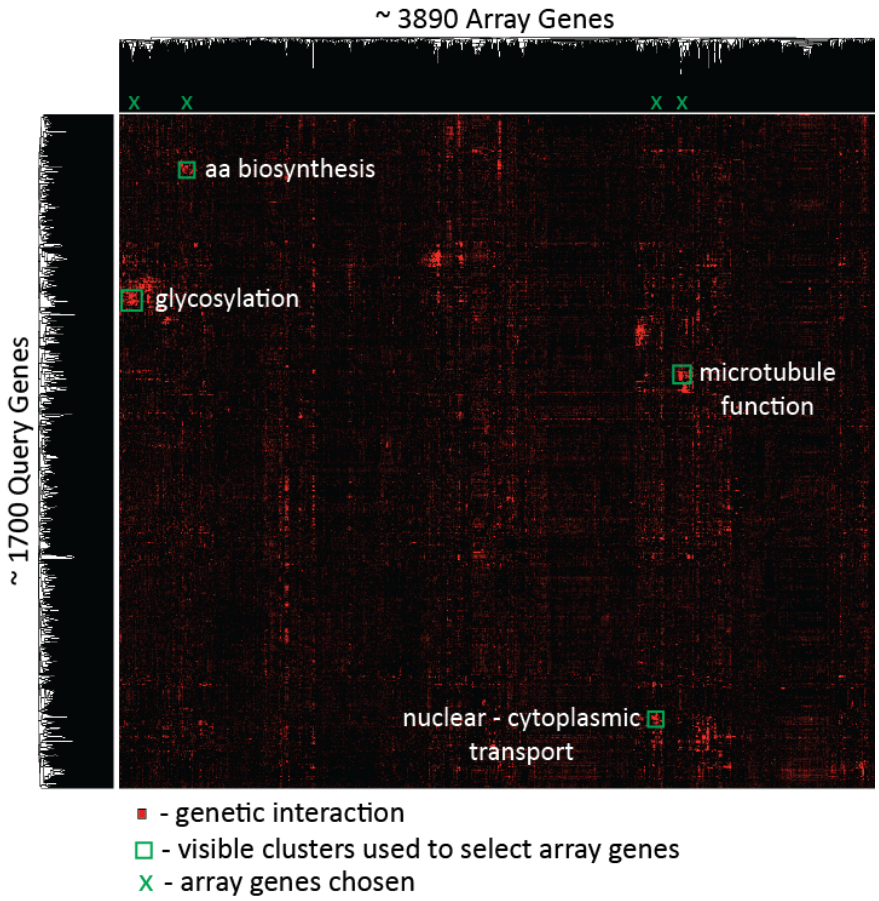

Figure 7. Manual strategy used for selecting the diagnostic subset of yeast deletion strains.

Similar to the computational approach, the aim of the manual strategy was to select array genes that participated in diagnostic interactions driving query-query grouping within a two-dimensional clustergram of genetic interaction data. To achieve this manually, a genetic interaction clustergram based on interactions between ~1700 query genes and ~3890 array genes was visually inspected to identify defined clusters. Highlighted in green are four examples of defined clusters identified, with the biological processes corresponding to the query genes represented in the cluster listed. The array genes participating in the interactions defined by each of the visible clusters identified were compiled. Ultimately, inspection of all clusters yielded a final list of 442 genes.

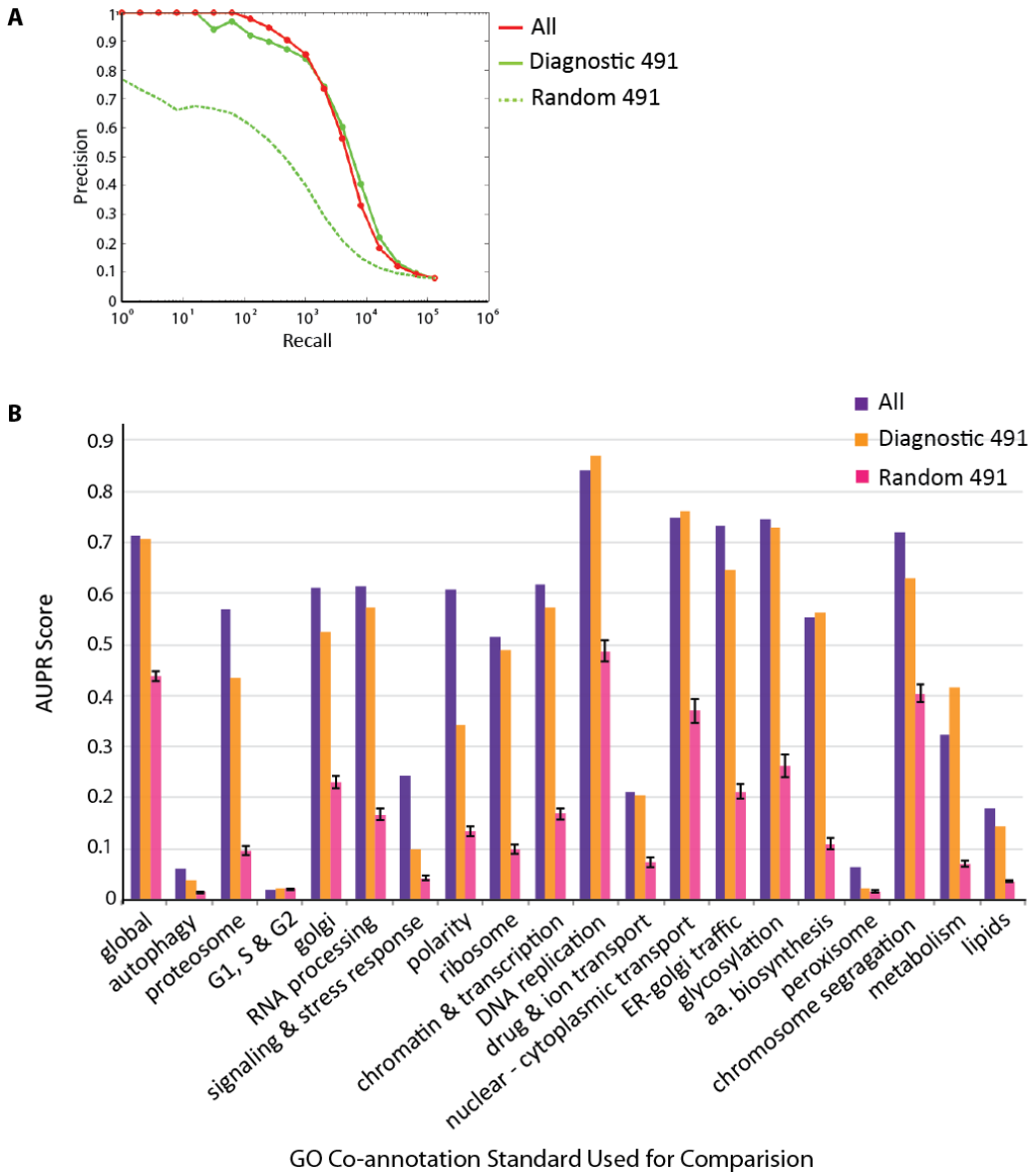

Figure 8. An analysis of genetic interaction clustering using the chosen 491 diagnostic deletion mutants.

A) A global PR analysis was performed on the chosen 491 diagnostic deletion mutants to evaluate their ability to correctly re-cluster genetic interaction data generated by our lab. Clustergrams were generated using genetic interactions involving either the complete array of ~3890 deletion mutants, the diagnostic subset of 491 deletion mutants or a random list of 491 deletion mutants. The resulting query-query grouping was compared to a GO co-annotation standard covering all biological processes to generate a global precision-recall plot. **B)** An AUPR score was calculated to quantify the differences between PR plots generated using the three gene lists. The precision-recall analysis and AUPR curve score calculation was repeated for 19 specific biological processes individually by comparing query-query grouping to 19-function specific GO co-annotation standards. *Results plotted for the random gene set represent an average of results calculated for 100 random gene sets.

plots generated for each comparison were quantified by calculating AUPR curve scores. Similar to the results obtained from the global analysis, the query-query grouping present in the clustergrams generated using interaction profiles restricted to the diagnostic genes were analogous to the grouping present in the complete profile clustergrams for most of the functional categories tested (**Figure 8B**). Interestingly, the diagnostic subset of genes appeared to perform slightly better than the complete set of genes at informatively grouping query-genes classified among the DNA replication, nuclear to cytoplasmic transport and metabolism functional categories. Conversely, both groups appeared to perform as poorly as the random subset of genes at grouping genes classified among the autophagy, G1/S & G2 (cell cycle) and peroxisome functional groups. This result suggests that genes from these functional categories may be under sampled in our genetic interaction data, proposing that the diagnostic genes chosen for these functional categories may not be the most informative given the reliance on genetic interaction data for selecting the diagnostic gene list. This potential downfall can be addressed in the future as additional interaction data is obtained, providing a more comprehensive view of the currently under-sampled biological processes to better assess the genes chosen and potentially select a more informative gene set for these categories. Overall, this analysis demonstrated genetic interaction profiles restricted to the diagnostic subset of 491 deletion mutants could cluster as informatively as genome-wide interaction profiles, endorsing the strategy used for gene selection.

To further analyze the 491 deletion strains selected, I wanted to evaluate whether CG interaction profiles restricted to the 491 genes could be used to informatively predict the pathways or molecules targeted by compounds screened with the assay. Briefly, these predictions can be made by computing the cosine similarity correlation coefficient between a given drug's CG interaction profile and the genetic interaction profiles produced by all ~1700 query genes present in the genetic interaction dataset (see Materials and Method). The cosine similarity metric employed resembles the commonly used Pearson correlation coefficient for comparing datasets, but was selected for use in this study based on its lack of reliance on centering data around an experimental mean (see Materials and Method). This feature ensures the sign of interactions evaluated is maintained and increases the speed with which large datasets can be processed. For my analysis, I used the CG interaction data from Parsons et al. (2006) to make target predictions for three scenarios: using profiles involving interactions with all ~3890 array genes, with profiles

restricted to interactions involving the diagnostic subset of array genes and with profiles restricted to an equally sized number of random array genes. Of note, only 449 of the 491 chosen strains could be incorporated into this analysis because Parsons et al. (2006) did not provide CG data for the remaining 42 strains. Ideally, the accuracy of target predictions could be analyzed through comparison with a gold standard that details the known targets of all compounds screened. However, such a resource does not exist and I therefore instead compared the predictions made using the restricted CG interaction profiles to the predictions made using the complete CG interaction profiles. This comparison was quantified by calculating cosine similarity correlation coefficients for each of the 82 drug profiles contained within the dataset. As shown in **Figure 9A**, the majority of the points fell to the right of the diagonal ($p = 1.1260 \times 10^{-8}$). This result indicated target predictions made using CG profiles restricted to the diagnostic genes were more correlated and thus similar to predictions made using the complete profiles, relative to predictions made using CG profiles restricted to a random set of genes. Averaging the correlation coefficients calculated for each of the 82 drugs reiterates this conclusion; the average correlation calculated for the diagnostic gene set was significantly greater then the average correlation calculated for the random gene set ($p = 1.6211 \times 10^{-9}$; **Figure 9B**). This outcome was particularly important, as it validates the idea that information produced by the analysis of restricted CG interaction profiles is comparable to the information gained from the analysis of CG interaction profiles generated using the genome-wide deletion collection. Consequently, this suggests that data obtained by screening the diagnostic subset of 491 deletion mutants should yield CG interaction profiles from which informative predictions can be made regarding the pathways or molecules targeted by each compound. The aforementioned analyses performed on the diagnostic subset of 491 deletion strains were done in collaboration with Raamesh Deshpande.

2.4 An assessment of two PDR mutants reveals the need to create a novel drug-hypersensitive strain

The non-essential yeast PDR drug transporters serve as ideal targets for deletion to create a drug-hypersensitive strain. While the overexpression of PDR-related genes is connected to multi-drug resistance, their deletion has been shown to sensitize yeast to a wide-variety of compounds

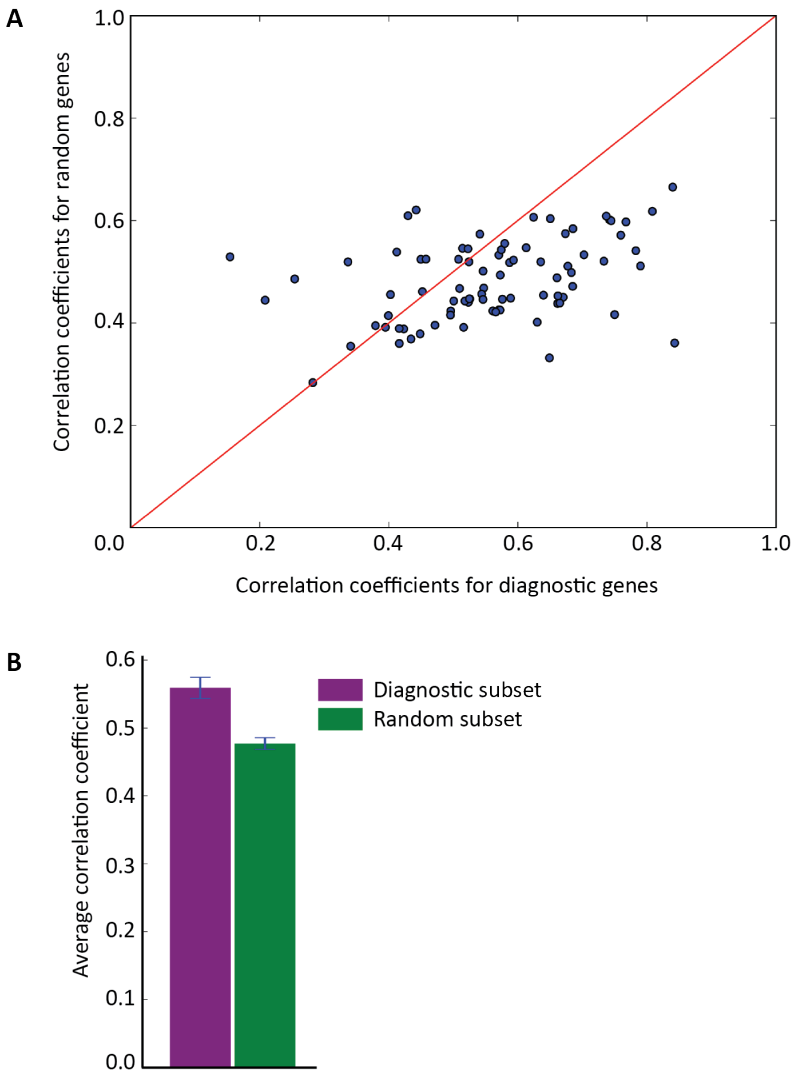

Figure 9. An analysis of compound-target predictions obtained using CG interaction profiles restricted to the diagnostic subset of deletion mutants.

A) To further examine the diagnostic subset of deletion mutants selected, CG interaction profiles from the Parsons et al. (2006) dataset were restricted to the diagnostic strains and target predictions were made for all 82 drugs using a correlation-based analysis. This analysis was repeating by instead restricting the CG interaction profiles to an equally sized number of random deletion strains. All predictions were then compared to predictions made using complete CG interaction profiles and the resulting cosine similarity correlation coefficients calculated are plotted. Results computed by restricting CG profiles to the diagnostic subset of strains and the random subset of strains are plotted on opposing axes for comparison. This analysis reflects the results for only 449 of the 491 diagnostic genes because data was not available for the remaining 42 genes. Results plotted for the random gene set represent an average of results calculated for 10 equally sized random gene sets. **B)** The cosine similarity correlation coefficient values calculated for all 82 drugs were averaged for both the diagnostic and random gene sets for further comparison of the plots in A).

(Kolaczkowski et al., 1998; Michalkova-papajova et al., 2000; Rogers et al., 2001). I therefore opted to evaluate two strains deleted for PDR transport-related genes as potential backgrounds for my modified CG assay. The first strain, described in Rogers et al. (2001), was deleted for seven genes encoding PDR transporters (*YOR1*, *SNQ2*, *PDR5*, *PDR11*, *YCF1*, *PDR15*, *PDR10*) and two genes encoding transcription factors (*PDR1* and *PDR3*) that control expression of various drug transport genes. Previously, this 9Δ strain was found to exhibit a 2 to 200-fold increase in sensitivity to a variety of compounds (Rogers et al., 2001). I created the second strain evaluated by deleting only *PDR1* and *PDR3* to determine whether the absence of these homologous TFs alone was sufficient to sensitize the cell. This avenue was selected for exploration based on the known importance of *PDR1* and *PDR3* for regulation of the PDR response (reviewed in Moye-Rowley, 2003), and a previous observation that deletion of these two TFs increases drug-sensitivity (Rogers et al., 2001). The *pdr1Δ pdr3Δ* double mutant was generated by sequentially deleting *PDR1* and *PDR3* using PCR-directed mutagenesis (see Materials and Methods).

To evaluate the chemical sensitivity of the two PDR mutant strains, I employed a high-throughput chemical halo assay, in collaboration with Jeff Piotrowski at the RIKEN institute. This procedure was completed by pin transferring 440 NPs to plates seeded with yeast culture and evaluating whether the compounds were toxic based on the formation of a visible area of growth inhibition (see Materials and Methods). The chemical halo assay was chosen to investigate the drug sensitivity of these strains because it permits rapidly assessing many compounds and yields a binary outcome for easy data interpretation. While both mutants were found to have an increase in sensitivity relative to the wild-type control, the chemical-sensitivity of the 9Δ mutant was much greater then that of the *pdr1Δ pdr3Δ* mutant (**Figure 10A**). Although this result suggested the 9Δ mutant would be ideal for use as the drug-hypersensitive background, further analysis revealed the 9Δ strain possessed other undesirable phenotypes. By microscopically visualizing cells stained with Con A, I determined the 9Δ mutant possessed a budding defect, manifesting as the formation of multicellular clumps due to daughter cells remaining attached to mother cells at the bud site septum following cell division (**Figure 10B**; see Materials and Methods). This defect was not present in the *pdr1Δ pdr3Δ* mutant. Additionally, the 9Δ mutant was unable to sporulate, indicating that while this strain was highly drug sensitive it was clearly sick and unsuitable for use as a background in which to create a

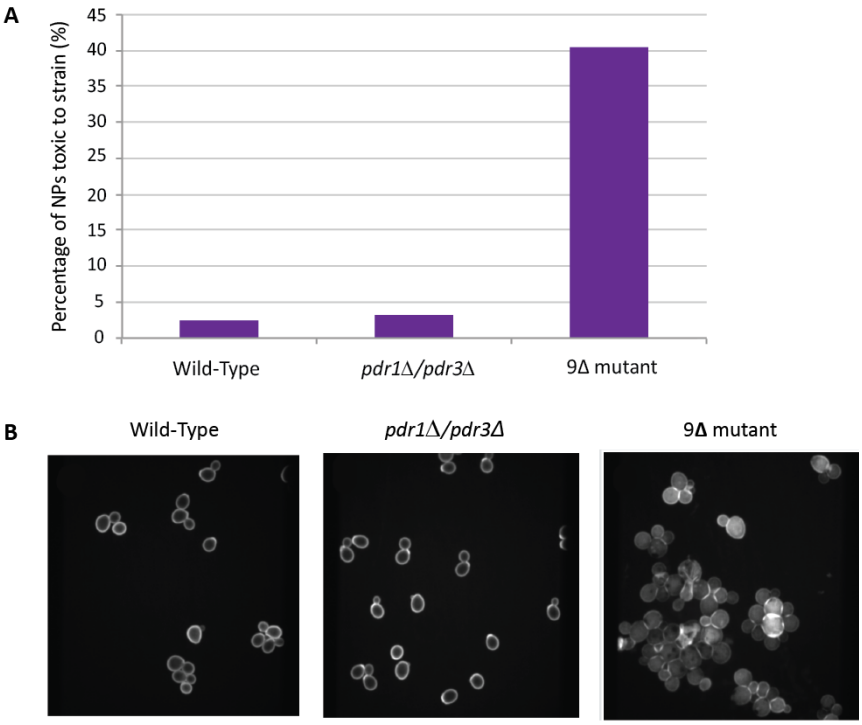

Figure 10. Phenotypic analysis of two *S. cerevisiae* PDR mutant strains.

A) The chemical-sensitivity of two PDR mutant strains was assessed by robotically pinning 440 NPs onto YEPD plates seeded with yeast culture. After incubating plates overnight, a compound was labeled toxic if it created a visible area of growth inhibition. Values plotted indicate the percentage of 440 compounds that were identified as toxic to the strain tested. BY4741 was used as a wild-type control. **B)** Images of the gross appearance of the PDR mutant strains were generated by staining log-phase cells with 100 ug/mL of Con A and imaging with fluorescence microscopy at 63X magnification. BY4741 was again used as a wild-type control.

novel deletion collection. Because the *pdr1Δ pdr3Δ* mutant did not possess these phenotypes, I chose to explore whether I could further sensitive the healthy *pdr1Δ pdr3Δ* mutant with the addition of a third drug-sensitizing mutation, instead of creating an entirely novel hypersensitive strain.

2.5 An analysis of chemical genomic and genetic interaction data identifies potential drug-sensitizing mutations

Three criteria were considered when selecting a third gene to delete in the *pdr1Δ pdr3Δ* mutant to increase drug sensitivity. First, deletion of the gene should confer sensitivity to many compounds and thus the gene should participate in many CG interactions. Second, deletion of the gene should not sensitize the cell to other genetic perturbations and therefore, the gene should exhibit few genetic interactions. It was important to ensure the chemical-sensitivity of the deleted gene was the result of a true MDR phenomenon and not simply due to the deleted gene genetically interacting with many different drug targets. The latter scenario would significantly complicate downstream CG interaction profile analysis. Finally, deletion of the gene should not result in a cell-fitness defect, as the future deletion mutants created in a sick background would perform poorly in a pooled growth CG assay. To identify genes possessing these features, CG interaction data produced by screening 2243 compounds was first assessed (data from St Onge et al. in preparation). All ~4000 genes present within the dataset were ranked based on the number of drugs they conferred sensitivity to when deleted, enabling me to identify the top 500 drug-sensitizing genes. I then utilized our lab's compendium of genetic interaction data to identify the number of genetic interactions each of the 500 genes participated in (Costanzo et al., 2010). The genes ranked among the top 500 based on their number of measured digenetic interactions were removed from the list of 500 drug-sensitizing genes. Single mutant fitness values were then obtained for the remaining 312 genes from data previously generated by our lab (Costanzo et al., 2010). As depicted in **Figure 11**, several genes were identified that fit the three criteria described above; many of the top drug-sensitizing genes also ranked low with regards to digenic interactions and caused no fitness defect when deleted. Since creating all 312 triple mutants was unrealistic and beyond the scope of my thesis, I selected a small subset of genes from two regions of the scatterplot for further analysis (F**igure 11**). The first sub-group was selected from

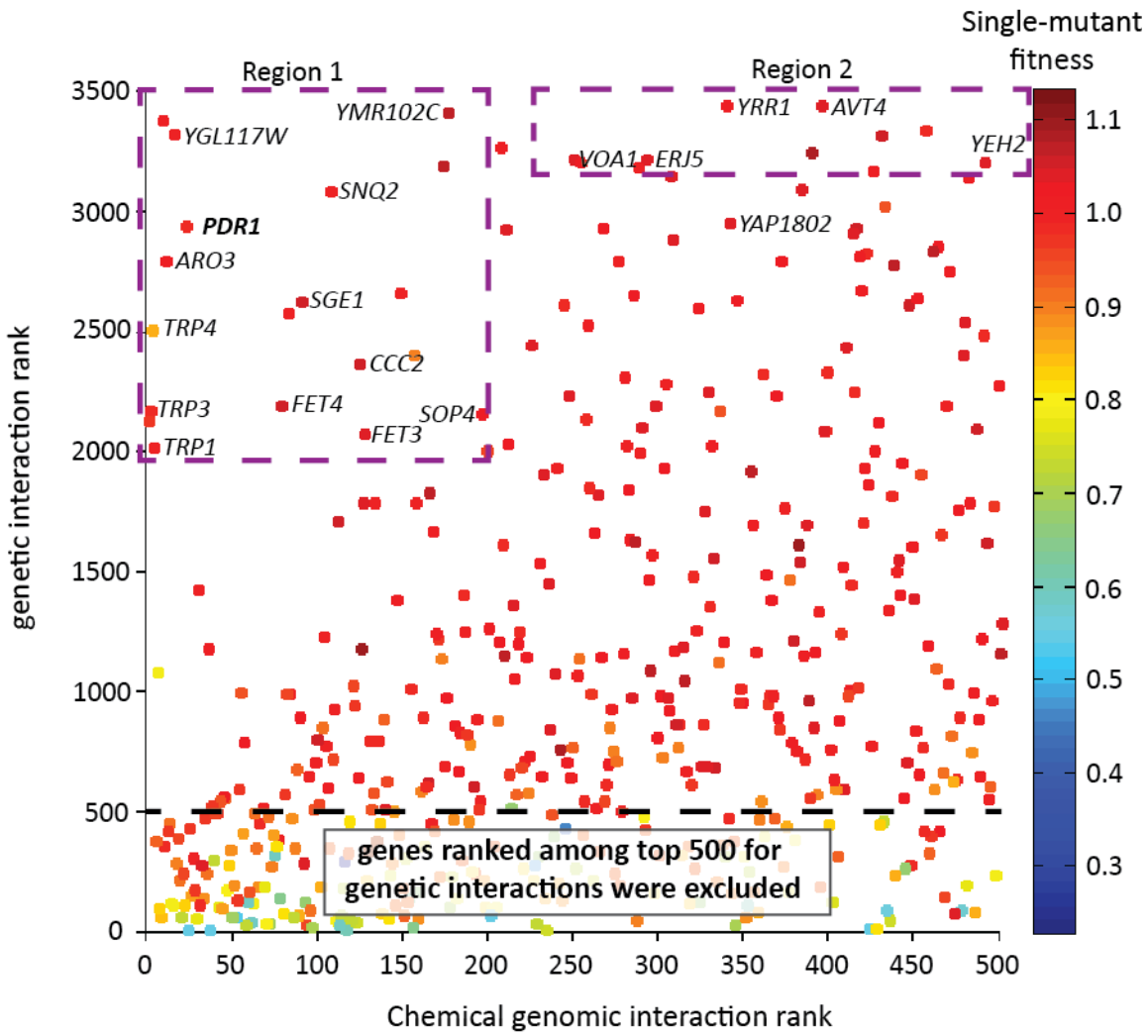

Figure 11. Identifying potential drug-sensitizing mutants through an analysis of CG and genetic interaction data.

Using CG data for 2243 compounds (from St Onge *et al.* in preparation), all array genes were ranked based on the number of drugs they conferred sensitivity to when deleted. Each spot plotted represents one of the top 500 drug-sensitizing genes. The color scale depicts the single-mutant fitness score, relative to WT, of each gene-deletion mutant. The genetic interaction ranks corresponding to these 500 genes were determined by ranking all array genes based on the number of digenic interactions each participates in using the genetic interaction data from Costanzo et al., (2010). Genes ranked among the top 500 for genetic interactions were excluded from further analysis. The two regions highlighted were used for gene selection. Genes found within region 1 participate in the most CG interactions and few genetic interactions, while the genes found within region 2 participate in fewer CG interactions but have fewer than 5 genetic interactions. The spots labeled represent the 18 genes selected for further analysis. *PDR1* was also identified from this analysis and is labeled. Genes not chosen from regions 1 or 2 were excluded because they were either uncharacterized genes or their deletion has been shown to result in undesirable phenotypes, according to the SGD database. For example, *YHP1* was located in region 1, but *yhp1Δ* mutants have abnormal cell morphology and possess a budding defect.

the upper-left region of the plot, consisting of genes ranked among the top 200 for drug sensitivity and the lowest 2000 for digenic interactions. This represented genes that exhibited sensitivity to 86 – 1066 compounds when deleted and participated in 11 – 19 digenic interactions. The second sub-group was selected from the top margin of the plot, consisting of genes participating in fewer than 5 digenic interactions and broadly ranked the among the top 500 drug-sensitizing genes. A total of 18 genes were selected from these regions and are listed in **Table 1,** along with their corresponding CG and genetic interaction data values. Independent from this analysis I selected an additional PDR transcription factor gene (*PDR8*) to evaluate. While it was not identified by the computational analysis, *PDR8* plays a key role in regulating the expression of PDR transporters (Hikkel et al., 2003) and its deletion was found to sensitive cells to a variety of drugs by our collaborators at the RIKEN institute (unpublished).

2.6 The creation and analysis of 19 *xxxΔ pdr1Δ pdr3Δ* mutants highlights four potential drug-hypersensitive background strains

Each of the 19 selected genes were deleted individually in the *pdr1Δ pdr3Δ* mutant using PCR-directed mutagenesis with the *K.lactis LEU2* marker (see Materials and Methods). To test the chemical sensitivity of the 19 newly created triple mutant strains, I employed a dose-response assay with eight different bioactive compounds. This work was done in collaboration with Marissa LeBlanc, a former post-doc at the RIKEN institute. The 8 drugs chosen for analysis were all well characterized and functioned through diverse MOAs. The dose-response assay performed involved culturing the deletion strains in a wide range of drug concentrations and performing a four-parameter logistic regression on the resulting optical density measurements to calculate the concentration of drug required to inhibit cell growth by 50% (IC_{50}; see Materials and Methods). By dividing the IC_{50} values obtained for a wild-type control by the IC_{50} values obtained for all 19 triple mutants, I calculated fold changes in chemical sensitivity for all triple mutants across the 8 drugs tested (**Table 2**).

From this analysis, I sought to identify the triple mutants that exhibited the greatest fold increase in drug sensitivity to the most compounds. I achieved this aim by first highlighting the strains that exhibited one of the top three fold-increases in sensitivity to each compound, and then

Table 1. Genes selected for deletion in the *pdr1Δ pdr3Δ* mutant strain to increase chemical-sensitivity.

The listed interaction values and chemical/genetic ranks were obtained from the analysis described in **Figure 11.**

ORF (Gene Name)	# of CI's*	Chemical Rank	# of GI's*	Genetic Rank	Gene Function/Description**
YDR035W (*ARO3*)	521	11	6	2788	DAHP synthase: catalyzes the first step in aromatic amino acid biosynthesis
YNL101W (*AVT4*)	61	397	0	3439	Vacuolar transporter: exports large, neutral amino acids from the vacuole
YDR270W (*CCC2*)	19	125	9	2367	Cu^{+2}-transporting P-type ATPase: required for Cu^{+2} export from cytosol
YFR041C (*ERJ5*)	69	294	3	3214	Type 1 membrane protein: required to preserve folding capacity of the ER
YMR058W (*FET3*)	114	128	12	2070	Ferro-O_2-oxoreductase: required for high-affinity iron uptake
YMR319C (*FET4*)	152	79	11	2192	Low-affinity iron transporter of the plasma membrane
YPR198W (*SGE1*)	136	91	8	2624	Plasma membrane transporter of the major facilitator superfamily (MFS)
YDR011W (*SNQ2*)	124	108	4	3083	Plasma membrane transporter of the ATP-binding cassette (ABC) family
YJL192C (*SOP4*)	87	197	11	2158	ER-membrane protein: deletion slows the export of Pma1p from the ER
YDR007W (*TRP1*)	875	4	12	2014	Phosphoribosylanthranilate isomerase: catalyzes third step in Trp biosynthesis
YKL211C (*TRP3*)	1066	2	11	2169	Bifunctional enzyme: involved in tryptophan biosynthesis
YDR354W (*TRP4*)	990	3	8	2502	Anthranilate phosphoriosyl transferase: catalyzes phorphoribosylation of anthranilate
YGR106C (*VOA1*)	75	251	3	3220	ER protein: functions with other factors in the assembly of the V_0 sector of the vacuolar ATPase
YGR241C (*YAP1802*)	64	343	5	2952	Protein involved in clathrin cage assembly
YLR020C (*YEH2*)	55	483	4	3137	Steryl ester hydrolase: catalyzes steryl ester hydrolysis at plasma membrane
YGL117W	417	16	2	3322	Unknown function
YMR102C	93	177	1	3412	Putative multidrug ABC transporter
YOR162C (*YRR1*)	65	341	0	3441	Zn_2-Cys_6 zinc-finger TF: activates genes involved in multi-drug resistance
YLR266C (*PDR8*)	33	> 500	10	2577	Transcription factor: regulates genes involved in pleiotropic drug resistance

* CI = chemical genomic interactions; GI = genetic interactions

** Gene descriptions/functions obtained from the SGD website: http://www.yeastgenome.org/

Table 2. The fold change in triple mutants sensitivity to the eight compounds tested.

Values were calculated by dividing the IC_{50} of the BY4741 wild-type strain by the IC_{50} for the triple mutant. Asterisks beside values indicate the fold change calculated was one of the top three identified out of the 19 triple mutants screened. The four triple mutants selected for further analysis, which exhibited the largest fold increase in sensitivity to the most compounds tested, are highlighted in grey. All gene deletions listed were present in a *pdr1Δ pdr3Δ* mutant background.

Gene Deletion	Drug Screened							
	AMB	ANIS	CYX	DAU	KETO	MICA	TERB	NYS
pdr8Δ	*1.8	1.7	2.5	*1.9	54.6	0.5	5.0	1.6
sge1Δ	0.9	1.4	2.5	1.4	36.6	1.2	4.3	1.6
snq2Δ	1.1	1.0	*10	1.4	25.4	1	0.2	*2.6
ymr102cΔ	1.5	1.6	3.3	*3.4	40.5	1.0	4.1	1.6
yrr1Δ	1.4	0.7	2	*2.0	6.9	0.6	*13.9	*2.1
aro3Δ	1.6	1.3	2.5	1.4	42.3	1.3	1.3	1.8
avt4Δ	1.2	1.0	2	1.2	21.2	1.3	1.3	1.6
trp1Δ	1.6	0.8	2	1.2	49.4	1.2	7.6	1.9
trp3Δ	1.6	1.2	2.2	1.1	*87.3	1.2	*8.7	1.6
trp4Δ	1.6	1.7	2	1.3	*63.6	1.3	8.3	1.6
ccc2Δ	1.6	1.0	3.3	0.9	22.3	1.2	1.1	1.8
fet3Δ	1.6	1.3	2	1.1	24.0	1.3	5.8	1.8
fet4Δ	*1.8	*1.9	*6.7	1.8	51.4	*1.5	2.9	1.8
erj5Δ	1.6	1.5	5	1.2	45.9	1.2	0.6	1.6
sop4Δ	1.1	1.0	2.2	1.1	*80.9	1.3	3.2	1.9
voa1Δ	1.6	1.5	2	1.3	23.4	1.2	2.6	1.7
yap1082Δ	1.6	1.2	2.2	1.1	27.8	1.2	1.4	1.7
yeh2Δ	1.5	*3.7	4	0.5	41.2	1.2	*11.5	1.5
ygl117wΔ	1.5	*5.2	*6.7	1.6	25.6	1.2	1.1	*2.1

retaining the strains that ranked among the top for the greatest number of compounds. This analysis distinguished four triple mutants to explore further. The top sensitive strain was deleted for *FET4*, which encodes a low-affinity iron transporter. This result was interesting, as Fet4p has not been linked previously to drug transport or MDR, yet the strain harboring a *FET4* deletion ranked among the top three drug-sensitive strains for four out of the eight compounds tested. The triple mutant deleted for *YRR1* exhibited the second largest increase in drug-sensitivity, followed by the *PDR8* and *SNQ2* triple mutants. All three of these genes have been previously linked to MDR; *YRR1* and *PDR8* encode transcription factors that regulate the expression of drug transporters (Le Crom et al., 2002; Hikkel et al., 2003), while *SNQ2* encodes a known PDR transporter (Balzi et al., 1994). Interestingly, the expression of *SNQ2* was initially thought to be controlled exclusively by the Pdr1p and Pdr3p TFs (Decottignies et al., 1995). However, the observed increase in drug-sensitivity associated with deletion of *SNQ2* in a *pdr1Δ pdr3Δ* mutant corroborates a more recent finding that suggests the expression of *SNQ2* is controlled by other PDR transcription factors, such as Yrr1p (Le Crom et al., 2002).

2.7 Deletion of *SNQ2* in the *pdr1Δ pdr3Δ* mutant yields a highly drug sensitive strain

To further evaluate the chemical sensitivity of the top four drug-sensitive strains identified from the dose-response analysis, I performed a high-throughput chemical-halo assay using a collection of 440 natural products in collaboration with Jeff Piotrowski (see Materials and Methods). While all four strains exhibited an increase in drug-sensitivity relative to the *pdr1Δ pdr3Δ* mutant strain, the triple mutant deleted for *SNQ2* exhibited sensitivity approaching that of the 9Δ drug-hypersensitive strain (**Figure 12A**). To evaluate whether this chemical-sensitivity was due to the deletion of *SNQ2* alone or resulted from the unique combination of the *pdr1Δ, pdr3Δ* and *snq2Δ* deletions, I used PCR mutagenesis to create single and double deletion strains harboring all possible combinations of *pdr1Δ, pdr3Δ,* and *snq2Δ*. These additional five strains were then evaluated using the same chemical-halo assay. As depicted in **Figure 12B**, I found the chemical-hypersensitivity was unique to the *pdr1Δ pdr3Δ snq2Δ* triple mutant, as all single and double mutants were markedly insensitive. Further analysis of the *pdr1Δ pdr3Δ snq2Δ* triple mutant

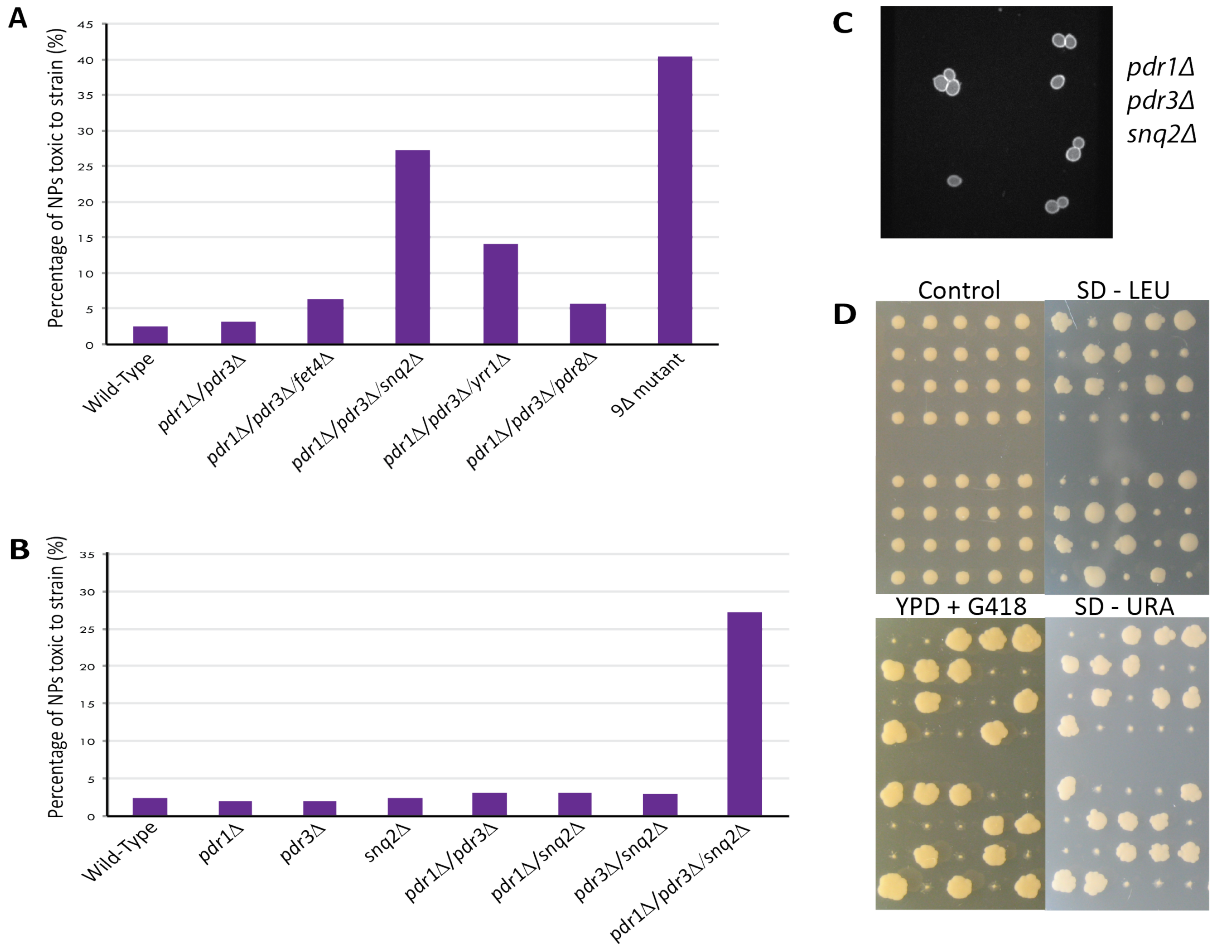

Figure 12. Phenotypic analysis of potential drug-hypersensitive background strains.

A) The chemical-sensitivity of the top four drug-hypersensitive strains was assessed by robotically pinning 440 NPs onto YPD plates seeded with yeast culture. After incubating plates overnight, a compound was labeled toxic if it created a visible area of growth inhibition. Values plotted indicate the percentage of 440 compounds that were identified as toxic to the strain tested. BY4741 was used as a wild-type control. **B)** To evaluate the contribution of the *pdr1Δ, pdr3Δ* and *snq2Δ* gene deletions to the drug-hypersensitivity of the triple mutant strain, all possible single and double mutant combinations were created and tested using the same approach. **C)** The gross appearance of the *pdr1Δ pdr3Δ snq2Δ* triple mutant strain was visualized by staining log-phase cells with 100 µg/mL of Con A and imaging with fluorescence microscopy at 63X magnification. **B)** To evaluate sporulation, cells were cultured in supplemented SPO media for 7 days and resulting tetrads were dissected (control plate). Spores were replica-plated onto YPD + G418, SD –URA and SD – LEU media to ensure each marker segregated 1:1 as expected.

revealed the strain did not possess the clumpy phenotype present in the 9Δ drug-hypersensitive strain and was capable of sporulating correctly, as evidenced by tetrad analysis (**Figure 12C-D**; see Materials and Methods). This triple mutant was therefore chosen as the drug-hypersensitive background for my CG profiling assay.

2.8 Regenerating the diagnostic subset of yeast deletion mutants in the *pdr1Δ pdr3Δ snq2Δ* background using a modified SGA protocol

The cumbersome task of re-creating the selected 491 deletion mutants in the *pdr1Δ pdr3Δ snq2Δ* background was circumvented using a modified synthetic genetic array (SGA) pipeline. Use of the SGA strategy was possible because the *pdr1Δ pdr3Δ snq2Δ* mutant was created in the SGA query strain background (y7092), which harbors the important *can1Δ::STEpr-SP_his and lypΔ* reporters (Tong et al., 2001). The drug-hypersensitive query mutant was crossed to an ordered array of the 491 deletion mutants and quadruple mutant progeny were selected for, in lieu of traditional double mutant selection, using a series of replica plating steps via robotic pinning (**Figure 13**; see Materials and Methods). All SGA output colonies were then streaked for single colonies to isolate a single clonal population. This process ultimately resulted in the creation of 491 *MAT***a** haploid strains carrying all four marked deletions. I then used a series of quality control measures to validate the phenotype of these strains and assess integrity of the modified SGA pipeline used for their creation. First, I used a flow-cytometry analysis to evaluate the ploidy of the SGA output strains to verify each was haploid (see Materials and Methods). I opted to validate the ploidy of the entire set of 491 output strains because this characteristic was essential to prevent screening diploid cells, which would ultimately create false negatives in future CG data generated using this deletion collection. Similar to the results depicted in **Figure 14A**, all 491 mutants yielded profiles characteristic of either 1n or 2n DNA content, as evidenced by comparison to the haploid and diploid controls. All 491 SGA output strains were therefore classified as haploid. To further interrogate the modified SGA pipeline, I then evaluated a subset of strains present on each array plate to ensure the *MAT***a** mating type was correctly selected. This was performed using a pheromone based mating type test, whereby a strain is identified as

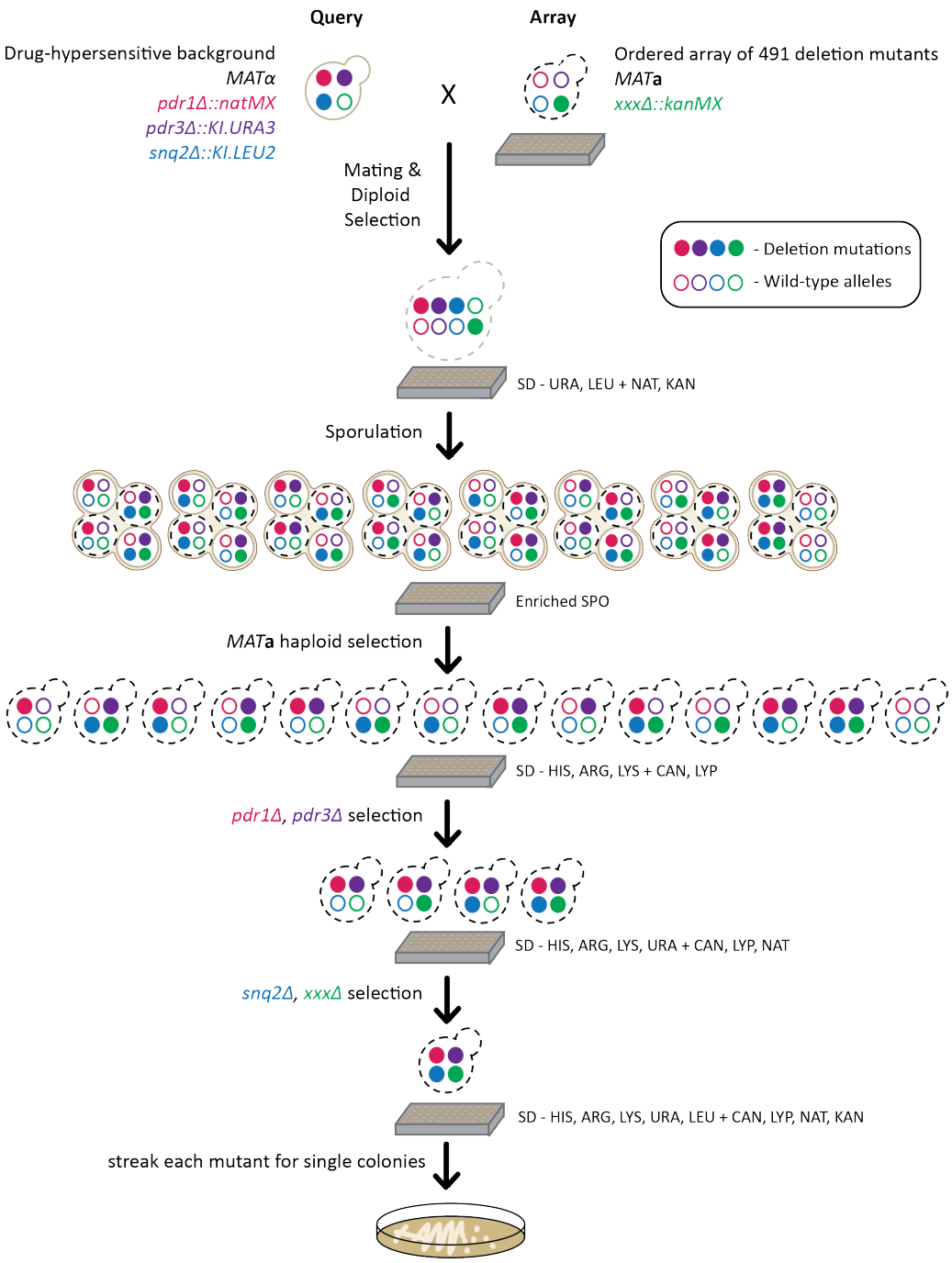

Figure 13. A modified SGA protocol for regenerating the 491 deletion mutants in the drug-hypersensitive background.

A *MAT*α query strain carrying the *pdr1Δ pdr3Δ snq2Δ* deletions, as well as the *can1 Δ::STEpr-SP_his* and *lypΔ* reporters, was crossed to an ordered array of the 491 *MAT*a *xxxΔ::kanMX* deletion mutants. The resulting zygotes were selected and transferred to enriched SPO to induce sporulation. First the *MAT***a** haploid meiotic progeny carrying the *can1Δ::STEpr-SP_his and lypΔ* reporters were selected, followed by concurrent selection of the *pdr3Δ::KI.URA3* + *pdr1Δ::natMX* deletions and the *snq2Δ::KI.LEU2* + *xxxΔ::kanMX* deletions. Final output strains were streaked for single colonies. This approach was modified from Tong et al. (2001).

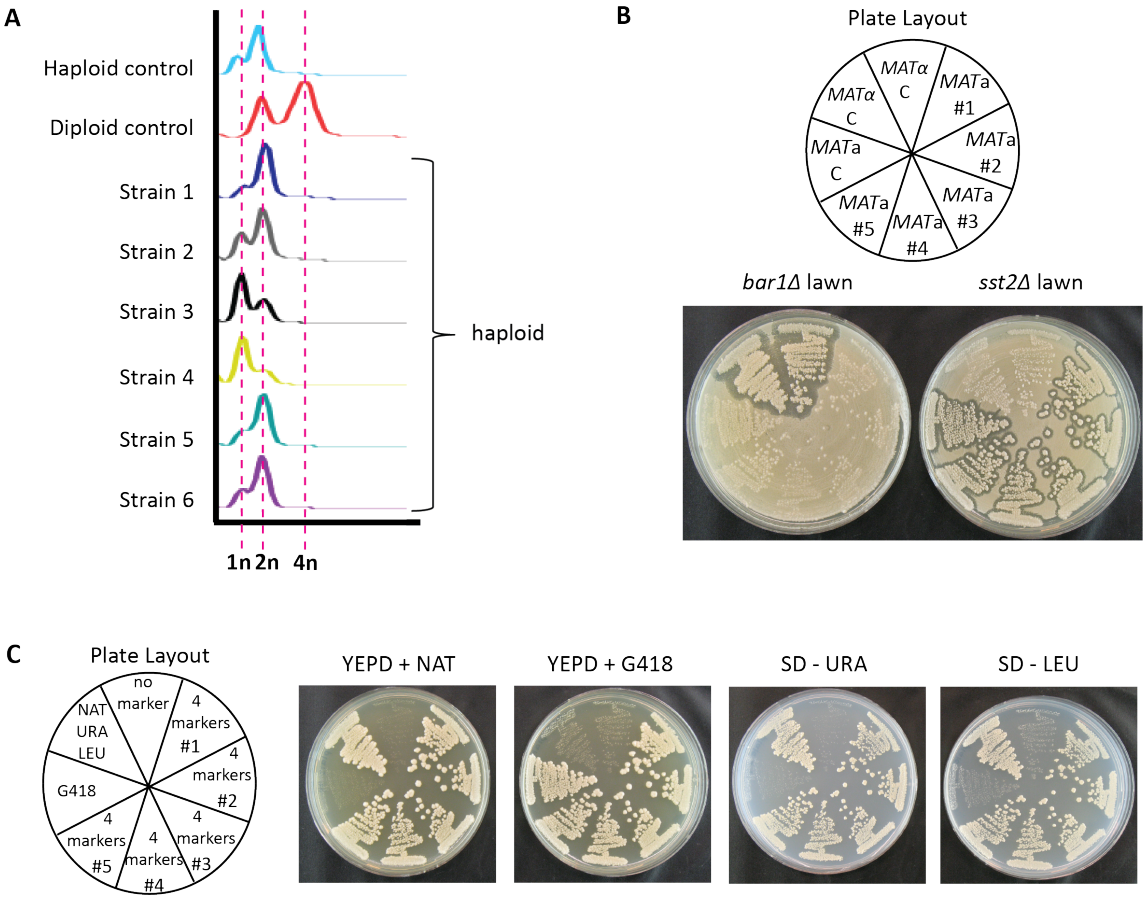

Figure 14. Quality control of SGA output strains.

Quality control experiments were used to confirm the phenotype of SGA output strains and validate the modified SGA pipeline. **A)** Each strain's ploidy was assessed by fixing and staining treated cells grown to mid-log phase and running samples through a flow cytometer. Similar to the 6 strains depicted, all 491 SGA output strains were haploid, containing only 1n or 2n DNA content. BY4741 was used as the haploid control while BY4743 was used as the diploid control. **B)** Pheromone halo assays were used to verify the mating type of 5 randomly selected strains from each array plate. After streaking out strains and growing overnight on YEPD media, cells were replica plated onto lawns of *bar1Δ* or *sst2Δ* cells. The presence of a halo on the *bar1Δ* lawns indicates *MATα* cells and the presence of a halo on the *sst2Δ* lawns indicates *MAT***a** cells. Similar to depicted, all SGA output strains tested were *MAT***a**. The *pdr1Δ pdr3Δ snq2Δ* strain and the SGA query starting strain were used as *MATα* positive controls, while the *his3Δ::kanMX* strain from the deletion collection was used as a *MAT***a** positive control. **C)** The mutants streaked out for B) were also replica plated onto the following media: YEPD + NAT to test for the *pdr1Δ::natMX* deletion, YEPD + G418 to test for the *xxxΔ::kanMX* deletion, SD – URA to test for the *pdr3Δ::KI.URA3* deletion, and SD – LEU to test for the *snq2Δ::KI.LEU2* deletion. All SGA strains tested harbored all four deletion markers, as evidenced by growth on the appropriate media. The *pdr1Δ pdr3Δ snq2Δ* strain was used as a positive control for the *NAT*, *KI.URA3* and *KI.LEU2* markers, the *his3Δ::kanMX* strain from the deletion collection was used as a positive control for the *kanMX* marker and the SGA query starting strain was used as negative control for all markers.

*MAT***a** or *MAT*α based on their production of **a** or α type mating pheromone which induces G1 cell cycle arrest in *sst2Δ* or *bar1Δ* mutants, respectively (see Materials and Methods). Like the results shown in **Figure 14B**, all SGA output strains tested generated halos of growth inhibition on an *sst2Δ* lawn, verifying they were of the *MAT***a** mating type. Finally, I chose to ensure all four marked deletions (*pdr1Δ::natMX, pdr3Δ::KIURA3, snq2Δ::KILEU2* and *xxxΔ::kanMX*) were selected correctly by replica plating a subset of strains from each array plate onto the appropriate selection media (see Materials and Methods). As expected, all strains evaluated grew on the four types of selection media, signifying they carried the four deletion markers (**Figure 14C**). These results therefore corroborated the phenotypic identity of the output strains to validate the integrity of the modified SGA pipeline used for their creation.

2.9 The increased chemical sensitivity of two SGA output strains verifies function of the drug-hypersensitive background

To confirm the utility of the SGA pipeline to regenerate deletion mutants in a functioning drug-hypersensitive background, I selected two of the 491 4Δ SGA output strains and evaluated their known chemical sensitivity. The first strain selected was deleted for *RAD52*, a member of the RAD52 epistasis group involved in the repair of double-stranded DNA breaks (Symington, 2002). Deletion of this gene therefore sensitizes yeast to various DNA damaging agents, such as the glycopeptide antibiotic phleomycin, which causes double stranded DNA breaks through inducing radical mediated DNA damage (Enserink *et al.*, 2009). The membrane integrity gene *VPS8* was deleted in the second strain chosen, increasing the cell's sensitivity to drugs like nystatin, that function through compromising cell membrane integrity (Giaever *et al.* 2002). I reasoned that if the *pdr1Δ, pdr3Δ,* and *snq2Δ* deletions were functional, then the SGA output strains harboring *rad52Δ* and *vps8Δ* deletions should possess increased sensitivity to phleomycin and nystain, respectively, when compared to strains deleted for these genes in a wild-type background. To evaluate this prediction, a liquid-growth assay measuring turbidity over time was performed to compare the chemical sensitivity of the two SGA output strains to the sensitivity of *rad52Δ* and *vps8Δ* mutants from the original yeast deletion collection (see Materials and Methods). As expected, this analysis revealed deletion of *rad52Δ* and *vps8Δ* in the drug hypersensitive background increased sensitivity to the tested compounds **Figure 15**. In addition

to validating the function of the drug-hypersensitive background, the observed increase in sensitivity further substantiates that I can evaluate a lower concentration of each NP with the adapted assay as a result of regenerating each deletion allele in the *pdr1Δ pdr3Δ snq2Δ* background.

2.10 Culturing yeast in galactose-based media further increases drug sensitivity, relative to the use of glucose-based media

Preliminary experiments performed by my collaborator Jeff Piotrowski suggested culturing in rich galactose-based media could be used to further sensitive yeast to various compounds (unpublished). This analysis was inspired by a previous study that explored the concept of altering culture media composition to increase the drug-permeability of yeast (Pannuzio et al., 2004). I opted to further explore whether replacing dextrose with galactose would sensitize yeast by performing a dose-response analysis using six control compounds (see Materials and Methods). For all six compounds tested, the measured IC_{50} value was lower when the analysis was performed in YPGal media (**Table 3**). Specifically, when culturing in YPGal media, cells experienced at least a two-fold increase in sensitivity to cyclohexamide, ketoconazole, daunorubacin, and ansiomycin, while the fold-change in sensitivity to amphotericin B and rapamycin was below 1.5 (**Table 3**). Although these changes are not drastic, culturing in galactose-based media will afford a further reduction in the concentration of compound required for analysis, in addition to the sensitivity change resulting from using the drug-hypersensitive genetic background. One simple explanation that may explain this observation is yeast grow slower in galactose-based media. This change could afford each compound an increased amount of time to interact with their target or exaggerate the resulting physiological response. Another hypothesis for this result is based on a previous study, which demonstrated that substituting glucose with galactose increased the expression and function of various amino acid permeases found in the plasma membrane though regulation by the carbon catabolite repression pathway (Peter et al., 2006). If these permeases can also import drugs, or if the signaling cascade responsible for the up-regulation of these permeases up-regulates expression of other proteins capable of drug-import, then this concept may provide an explanation for the observed increase in drug-sensitivity resulting from culturing in galactose-based media.

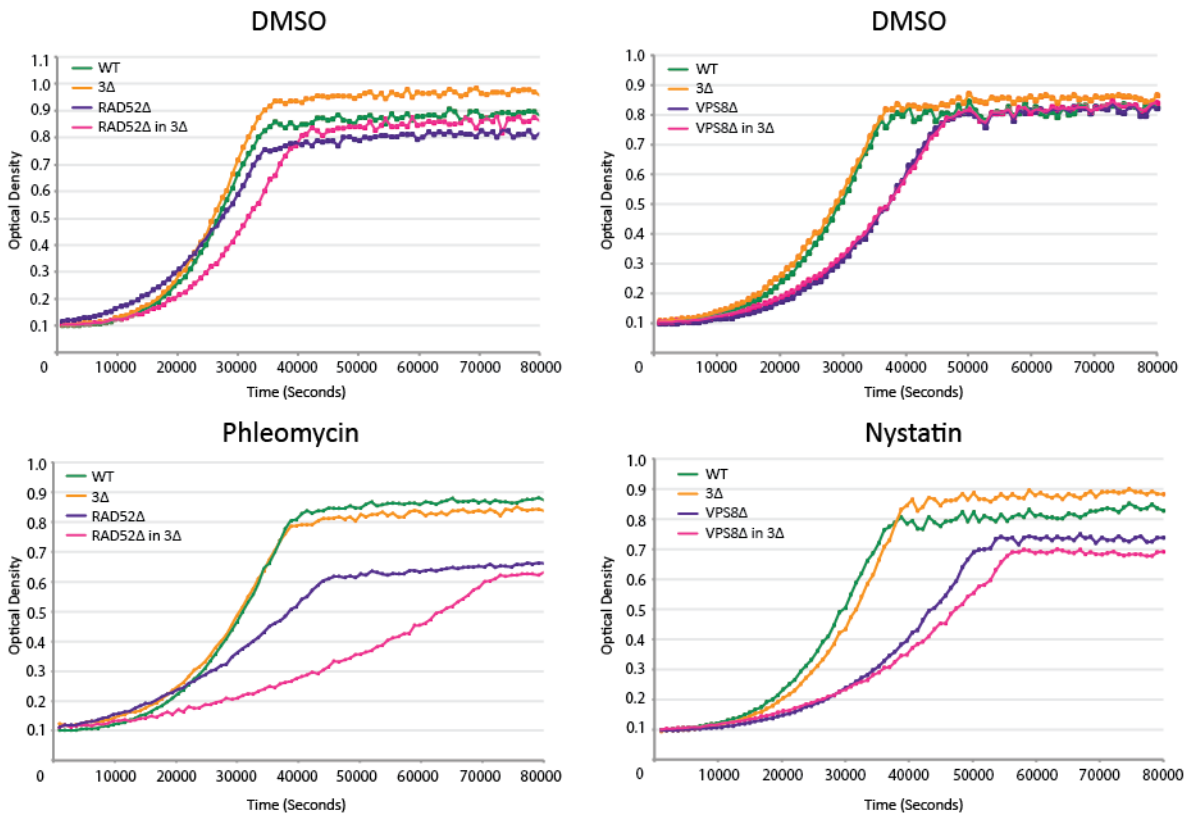

Figure 15. The chemical sensitivity of two SGA output strains.

The chemical sensitivity of strains deleted for *RAD52* and *VPS8* in both a wild-type background (*rad52Δ* & *vps8Δ*) and the hypersensitive *pdr1Δ pdr3Δ snq2Δ* background (*rad52Δ* in *3Δ* and *vps8Δ* in *3Δ)* were assessed with a liquid-growth analysis. BY4741 was used as wild-type control (WT) for the wild-type background mutants and the *pdr1Δ pdr3Δ snq2Δ* triple mutant (3Δ) was used as a control for the hypersensitive mutants. All strains were grown overnight in YEPD and after subculturing the next morning to ~OD_{600} = 0.0625, the *rad52Δ* cultures were spiked with 0.4 ng/µL of phleomycin and 0.05 µg/mL of nystatin was added to the *vps8Δ* cultures. Turbidity was monitored over ~22 hours. Culturing in 2% DMSO was used a no-drug control for both experiments.

Table 3. A comparison of IC_{50} values measured following growth in glucose vs. galactose-based media

IC_{50} values represent the concentrations of each compound required to inhibit 50% of cell growth when culturing yeast in YP media supplemented with 2% dextrose (YPD) or 2% galactose (YPGal). Values were determined by performing a four-parameter logistic regression of the results obtained using a standard dose-response assay. Fold-change values were calculated by dividing the YPD IC_{50} by the YPGal IC_{50}.

Drug Tested	Calculated IC_{50} Value		Fold IC_{50} Change
	YPD	YPGal	
Amphotericin B	0.2123	0.1987	1.07
Cyclohexamide	0.1055	0.0557	1.89
Ketoconazole	> 2.0	0.9331	> 2.14
Daunorubicin	> 10.0	5.2988	> 1.89
Rapamycin	0.0093	0.0082	1.13
Anisomycin	19.674	8.7249	2.25

2.11 Chemical genomic interaction profiles generated using the adapted assay are comparable to those produced by the traditional assay

Prior to screening the valuable NPDepo compounds with the modified assay, it was important to investigate whether the adaptations implemented would affect the quality of CG data produced. This question was addressed by selecting 5 control compounds, screened previously using the traditional CG assay, for analysis. The compounds chosen were the microtubule-disruptant benomyl, the DNA damaging agents camptothecin and mitomycin C, the estrogen receptor antagonist tamoxifen and the fungal cell-wall disturbing agent micafungin. To simulate the exact methodology that would be used for screening the NPDepo library, the 5 compounds were evaluated among a series of other compounds contained within four 96-well plates. Including the adaptations of a novel deletion collection, the use of YPGal media, and analysis in a 96-well format, these compounds were assayed as described previously using next-generation sequencing for barcode quantification (Smith et al., 2011; see Materials and Methods). Briefly, this entailed inoculating each well of a 96-well plate with the novel deletion pool and spiking with either compound or a 2% DMSO solvent. After culturing for 5 generations, gDNA was isolated from each well and the strain-specific barcodes were PCR amplified using experiment-specific indexing primers. The resulting barcode amplicons were then quantified using 96-plex next generation sequencing, and CG interaction profiles for each compound were constructed by computing drug-gene interaction z-scores for each strain.

One test to evaluate the quality of data yielded from the adapted strategy is to compare the CG interaction profiles obtained for the 5 compounds to CG interaction profiles obtained previously using the traditional assay. In collaboration with Raamesh Deshpande, this was achieved initially by combining the CG interaction profiles for benomyl and camptothecin with the 82 CG interaction profiles from the Parsons et al. (2006) dataset and using two-dimensional clustering to identify which profiles were similar. I reasoned that if the CG data generated by the new assay was comparable to the CG data yield produced by the traditional approach, then the benomyl and camptothecin profiles should group with the profiles of related drugs within the clustergrams. This outcome was observed for both compounds. The benomyl CG profile grouped beside the Parsons et al. (2006) benomyl profile and the profile produced by the other microtubule disrupting agent nocodazole (**Figure 16**). Similarly, the camptothecin CG profile grouped within

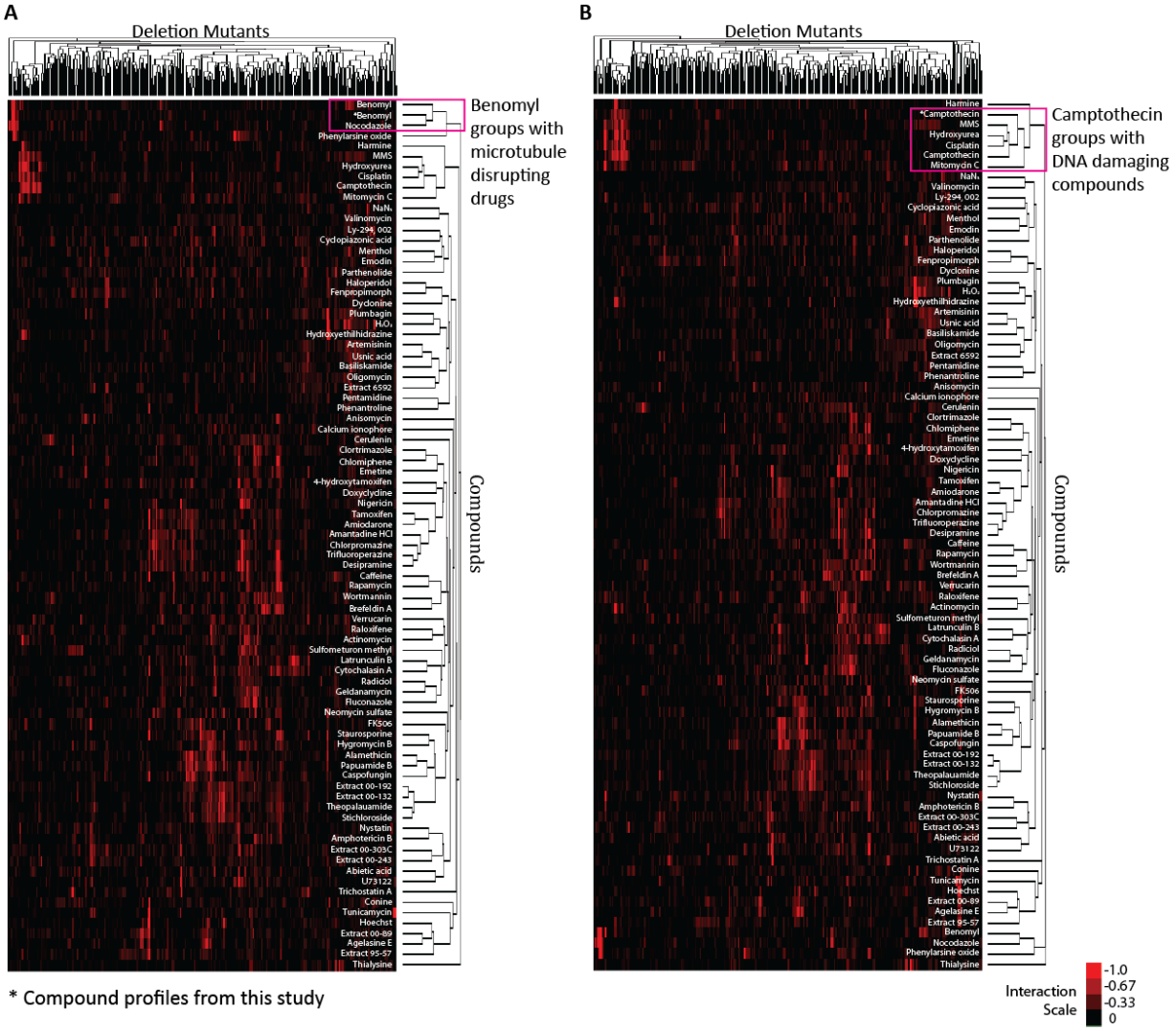

Figure 16. CG profiles generated by the adapted assay group informatively with previously generated CG interaction data using two-dimensional hierarchical clustering.

CG interaction profiles generated using the adapted assay were compared to the CG interaction data from Parsons et al. (2006) using two-dimensional hierarchical clustering. Given that only 491 genes were screened with the adapted assay, this analysis compared the CG interaction profiles for benomyl and camptothecin to the CG interaction profiles for the 82 Parsons et al. (2006) drugs restricted to same gene set. **A)** The CG interaction profile generated by benomyl grouped near the Parsons et al. (2006) benomyl profile, as well as near the other microtubule disrupting agent nocodazole **B)** The CG interaction profile generated by camptothecin grouped appropriately near a series of DNA damaging agents, including MMS, hydroxyurea, cisplatin, mitomycin C and the Parsons et al. (2006) camptothecin profile.

the profiles produced by a series of other DNA damaging agents, including the Parsons et al. (2006) camptothecin profile (**Figure 16**). To extend this analysis to the remaining 3 compounds and analyze the benomyl and camptothecin profiles further, I identified the top correlated CG interaction profiles from the Parsons et al. (2006) dataset by computing cosine similarity coefficients to quantify profile similarities. This analysis was again performed in collaboration with Raamesh Deshpande. For 4 out of the 5 control compounds assessed, the top correlated profiles from the Parsons et al. (2006) dataset were produced by the same tested compound, while most of the profiles that were second best correlated were produced by compounds functioning through a related MOA (**Table 4**). The exception to this observation was the result obtained for mitomycin C, whose profile correlated most with oligomycin (coefficient = 0.1789) and menthol (coefficient = 0.1678). Neither of these correlated compounds function through inducing DNA damage; oligomycin inhibits ATP synthase, while menthol is a terpene alcohol used as a flavorant. The lack of correlation between my mitomycin C CG profile and the Parsons et al. (2006) mitomycin C CG profile may have resulted from noise observed in both profiles, and not due to profile differences specific to the adapted assay. This hypothesis is based on a further analysis of the mitomycin C profile described below (results **section 2.12**). Another interesting result obtained from this correlation-based analysis was the high correlation found between my tamoxifen CG profile and the Parsons et al. (2006) desipramine profile (coefficient = 0.4245, with $p < 10^{-4}$). At first glance, this similarity was unexpected, as tamoxifen is an estrogen receptor antagonist and desipramine is a tricyclic antidepressant that inhibits norepinephrine reuptake. However, secondary to their primary MOAs, both tamoxifen and desipramine have been shown previously to perturb calcium homeostasis (Parsons et al., 2006; Qi et al., 2002). Because yeast lacks both estrogen receptors and a nervous system, the CG interaction profiles generated by these two compounds are likely representative of their secondary MOAs. Consequently, the observed similarity between these two profiles highlights the ability to use *S. cerevisiae* CG profiling to characterize bioactive compounds and provide MOA insights in cases where yeast lacks their primary mammalian targets. Comparing the CG interaction profiles generated by the adapted assay to a second large CG dataset (Hillenmeyer et al., 2008) yielded a similar result to the aforementioned analysis. The profiles for all compounds except tamoxifen were significantly correlated to the profiles produced by drugs functioning through related MOAs (**Table 4**). The top profiles found to correlate to tamoxifen were the N-WASP inhibitor wiskostatin and the smooth muscle relaxant alverine citrate. While

Table 4. Top correlated CG interaction profiles from the Parsons et al. (2006) and Hillenmeyer et al. (2008) datasets

Top correlated profiles from the Parsons et al. (2006) and the homozygous diploid component of the Hillenmeyer et al. (2008) datasets. Correlation coefficients were calculating using cosine similarity, with p-values estimated by comparing the actual correlation to a null distribution generated using a bootstrapping method (see Materials and Methods). The known MOAs of these compounds are listed in **Appendix B**.

Compound from this study	Profile Correlation with Other CG Data Sets			
	Parsons et al. 2006		Hillenmeyer et al. 2008	
	Compound	Correlation	Compound	Correlation
benomyl	benomyl	0.4560 ($p < 10^{-4}$)	benomyl	0.6111 ($p < 10^{-4}$)
	nocodazole	0.4567 ($p < 10^{-4}$)	mebendazole	0.2988 ($p = 0.0003$)
camptothecin	camptothecin	0.4191 ($p < 10^{-4}$)	camptothecin	0.4300 ($p < 10^{-4}$)
	hydroxyurea	0.3400 ($p < 10^{-4}$)	MMS	0.4169 ($p < 10^{-4}$)
tamoxifen	tamoxifen	0.4298 ($p < 10^{-4}$)	wiskostatin	0.2705 ($p < 10^{-4}$)
	desipramine	0.4245 ($p < 10^{-4}$)	alverine citrate	0.2502 ($p = 0.0005$)
micafungin	caspofungin	0.2273 ($p < 10^{-4}$)	nystatin	0.3150 ($p < 10^{-4}$)
	extract 95-97	0.2039 ($p < 10^{-4}$)	cantharidin	0.2801 ($p = 0.0007$)
mitomycin C	oligomycin	0.1789 ($p = 0.006$)	hydroxyurea	0.2428 ($p = 2427$)
	menthol	0.1678 ($p = 0.003$)	methotrexate	0.1854 ($p = 0017$)

wiskostatin has not been previously linked to calcium homeostasis, alverine citrate has been shown to relax smooth muscle through inhibiting calcium channels and increasing calcium influx (Hayase et al., 2007). Therefore, the observed correlation between the CG profiles for tamoxifen and alverine citrate likely resulted from both compounds altering yeast calcium homeostasis. As mentioned previously, this example further emphasizes the potential of using yeast CGs to identify the secondary MOAs of bioactive compounds and characterize drugs that possess mammalian specific targets.

Overall, the expected CG profile similarities identified using this correlation-based analyses suggests the profiles produced by the adapted assay are comparable to those produced using the traditional approach. Importantly, this observation demonstrates that comparing newly generated CG interaction profiles to our compendium CG interaction data should group related compounds to provide informative MOA insights.

2.12 The chemical genomic interaction profiles produced by screening a diagnostic subset of deletion mutants can be used to predict the targeted biological processes of bioactive compounds

The CG interaction profiles generated by screening the 5 control compounds with the adapted assay were further analyzed through comparing with genetic interaction data to predict the biological processes targeted by each compound. This work was performed in collaboration with Raamesh Deshpande. The ideal outcome of this comparison strategy is to predict the specific targets of compounds screened by identifying the genes that yield the top correlated genetic interaction profiles. However, because most drugs target essential genes, this outcome usually requires comparing CG profiles to the interaction profiles produced using essential query genes. Given that not all essential genes have been screened for genetic interactions to date, a more specific drug-target prediction is not yet possible and the comparison analysis described is currently restricted to broadly predicting the targeted biological processes of compounds screened. These biological process predictions can be formulated by: using cosine similarity to quantify the correlation between a given drug's CG interaction profile and the genetic interaction profiles produced by all ~1700 query genes present in the genetic interaction dataset, and then

identifying which cellular pathways or biological processes are enriched among the top correlated genes. I reasoned that if the correlated query genes identified using CG profiles from the adapted assay were enriched for genes functioning within the targeted pathway, then the CG data yielded by this approach is highly informative. Furthermore, if the predictions overlap with those generated using CG data from the traditional strategy, then the concept that data from these two approaches is comparable will be reinforced. From this analysis, three outcomes were observed and are described using the results for camptothecin, benomyl and mitomycin C.

Using a significance cutoff of $p = 0.01$, the DNA damaging agent camptothecin yielded 68 correlated query genes that could be classified into 6 functional categories, but were primarily composed of genes involved in DNA replication and repair (**Figure 17A**). Using GO ontology annotations downloaded from the gene ontology website (www.geneontology.org/) and a previously developed Java-based algorithm, the top three GO functional enrichments identified for these 68 correlated genes were DNA replication ($p = 7.55 \times 10^{-9}$), sister chromatid segregation ($p = 1.35 \times 10^{-7}$) and DNA metabolic processes ($p = 4.50 \times 10^{-7}$). Camptothecin functions through binding and stabilizing the DNA topoisomerase I enzyme/DNA complex, preventing DNA re ligation following replication and causing DNA double strand breaks (Liu et al., 2000). The enrichment observed among the 68 genes identified was therefore informative of camptothecin's known MOA. Furthermore, camptothecin has been shown to also affect the cell through inducing sister chromatic exchange (Zhao et al. 1992), supporting the observed enrichment for genes involved in sister chromatic segregation. Analysis of the camptothecin CG profile therefore suggests data yielded by the adapted assay can be used to predict a compounds targeted biological process. When predictions were instead made using the camptothecin CG interaction profile generated by the traditional assay, the 198 correlated genes identified were also highly enriched for informative biological processes. The top three GO enrichments identified were DNA metabolic processes ($p = 5.53 \times 10^{-27}$), DNA replication ($p = 7.13 \times 10^{-22}$) and DNA repair ($p = 2.48 \times 10^{-19}$). Additionally, these 198 correlated genes overlapped significantly ($p < 10^{-6}$) with the correlated genes identified using data from the adapted assay (**Figure 17B**). This result therefore demonstrates that screening a diagnostic subset of deletion mutants can generate CG data capable of capturing predictions formulated by screening the entire collection of ~4000 strains. Of note, a similar result was obtained when target predictions were made using the CG profiles produced by tamoxifen and micafungin.

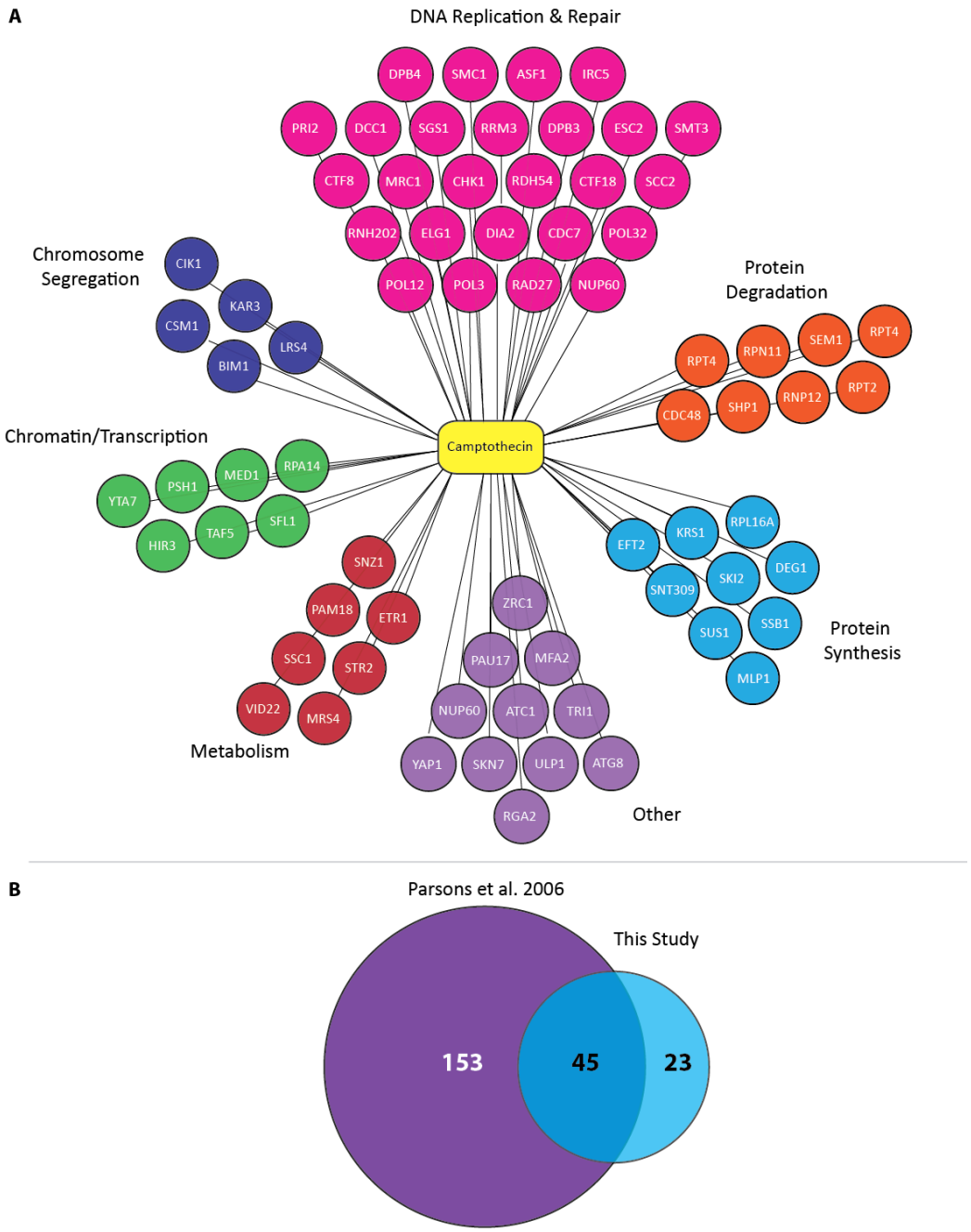

Figure 17. Analysis of the targeted processes predicted for camptothecin.

A) The 68 query genes correlated to the camptothecin CG profile ($p < 0.01$) were representative of 6 functional categories (classified according to a previously generated 19 category standard; Costanzo et al., (2010)). These 68 genes were functionally enriched for genes participating in DNA replication ($p = 7.55 \times 10^{-9}$), sister chromatid segregation ($p = 1.35 \times 10^{-7}$) and DNA metabolic processes ($p = 4.50 \times 10^{-7}$). **B)** These 68 correlated genes overlapped significantly ($p < 10^{-6}$) with the 198 correlated genes identified using the camptothecin CG data from Parsons et al. (2006).

An alternative outcome was observed when the biological processes targeted by benomyl were predicted. Again using a significance cutoff of p = 0.01, the benomyl CG interaction profile yielded 39 correlated query genes that fell into 4 functional categories, but were primarily represented by chromosome segregation, spindle and microtubule related genes (**Figure 18A**). The top three GO functional enrichments identified for these genes were mitotic cell cycle ($p = 7.35 \times 10^{-8}$), cell cycle ($p = 3.35 \times 10^{-6}$) and tubulin folding ($p = 4.97 \times 10^{-6}$). Inhibition of tubulin polymerization by benomyl can have broad cellular consequences, but may specifically block mitosis and prevent cell cycle progression. The GO term enrichments identified for the 39 correlated query genes are therefore informative of benomyl's known MOA. Furthermore, the top two correlated genes for benomyl were *CIN1* (correlation coefficient = 0.365; $p < 10^{-6}$) and *TUB3* (correlation coefficient = 0.343; $p < 10^{-6}$), which encode a tubulin folding factor and the α-tubulin protein, respectively. Consequently, the benomyl CG profile generated by the adapted assay is also predictive of a specific target, largely because benomyl does not target a specific essential protein. However, when target predictions were instead made using the benomyl CG interaction profile generated by the traditional assay, an alternative result was observed. While the two sets of correlated query genes overlapped significantly (**Figure 18B**), the correlated query genes identified using the Parsons et al. (2006) data were not significantly enriched for informative biological processes. Specifically, the top three GO enrichments identified using the Parsons et al. (2006) data were negative regulation of metabolic processes ($p = 2.59 \times 10^{-4}$), the negative regulation of biological processes ($p = 1.68 \times 10^{-3}$) and protein folding ($p = 1.48 \times 10^{-2}$). There were over 300 query genes correlated to the Parsons et al. (2006) benomyl CG profile, relative to the 39 query genes correlated to my benomyl CG profile. This result suggests the observed differences in enrichment may be due to increased noise in the Parsons et al. (2006) data, possibly owing to screening a larger library of deletion mutants. Therefore in contrast to the camptothecin example described above, analysis of the benomyl CG profile demonstrates the adapted may, in some instances, yield CG data that can be used to more informatively predict the processes targeted by compounds screened, relative to GC data produced by the traditional assay.

A final outcome was observed when the biological processes targeted by mitomycin C were predicted. Using a significance cutoff of p = 0.01, the mitomycin C CG interaction profile yielded 68 correlated query genes that fell into 6 functional categories, but were primarily represented by genes participating in DNA replication and repair (**Figure 19A**). The top three

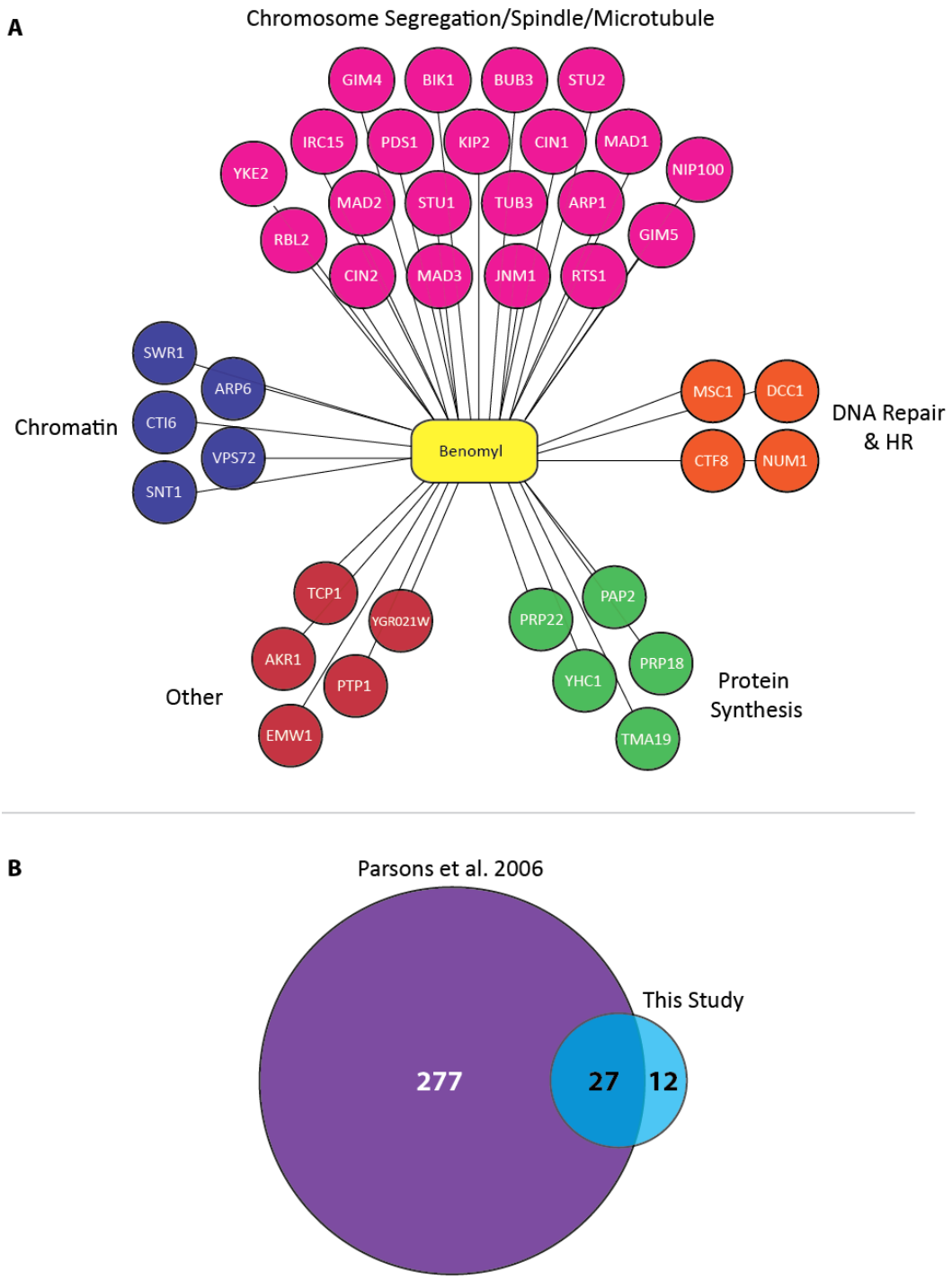

Figure 18. Analysis of the targeted processes predicted for benomyl.

A) The 39 query genes correlated to the benomyl CG profile were representative of 4 functional categories, classified according to a previously generated 19 category standard (Costanzo et al., (2010). These 39 genes were functionally enriched for genes participating in the mitotic cell cycle ($p = 7.35 \times 10^{-8}$), the cell cycle ($p = 3.35 \times 10^{-6}$) and tubulin folding ($p = 4.97 \times 10^{-6}$).**)** These 39 correlated query genes overlapped significantly ($p < 10^{-6}$) with the 304 correlated genes identified using the benomyl CG data from Parsons et al. (2006).

GO functional enrichments identified for these correlated genes were DNA repair ($p = 3.06 \times 10^{7}$), response to endogenous stimulus ($p = 1.32 \times 10^{-6}$) and response to DNA damage stimulus ($p = 2.93 \times 10^{-6}$). Furthermore, the top correlated gene identified was the non-essential gene *RAD52*, which functions to repair DNA damage that can result from DNA damaging drugs. Giving that mitomycin C is a potent DNA cross-linker that causes DNA damage, the biological processes predicted to be targeted by this drug are informative of mitomycin C's known MOA. When target predictions were instead made using the mitomycin C CG interaction profile produced by the traditional assay, the resulting predictions were also highly enriched for informative biological processes (**Figure 19B**). Specifically, the top three GO enrichments identified were DNA replication ($p = 8.54 \times 10^{-17}$), DNA metabolic process ($p = 1.35 \times 10^{-16}$) and DNA repair ($p = 1.92 \times 10^{-14}$). However, compared to the other drug profiles analyzed, there was a drastic reduction in the overlap of the correlated query genes identified using data generated by the two assays. This result suggests the mitomycin C profiles generated by both the adapted and traditional assays were noisy, and supports the lack of correlation observed between these two profiles described in section **2.11**. Interestingly, this noise did not appear to affect the target predictions that could be formulated, as evidenced by similar and informative GO functional enrichments identified using the set of correlated genes identified by analysis of either profile.

Inclusively, results from these profile analyses suggest CG interaction profiles produced by the adapted assay can be used to informatively predict the biological processes targeted by compounds screened. Importantly, these predictions are at least as good as predictions made using results generated by the traditional assay, supporting the adaptations implemented do not affect the quality of CG produced.

2.13 Confirming the top drug-sensitive strains for two compounds reinforces potential of screening with drug-hypersensitive strains

A final experiment performed that reinforced the purpose of the drug-hypersensitive background was confirming the top benomyl and micafungin-sensitive strains identified using the adapted CG assay. The top two strains found to exhibit a chemical genetic interaction with benomyl were deleted for *cin1Δ* ($z = -7.899$) and *tub3Δ* ($z = -5.38$), while the top two strains found to exhibit a

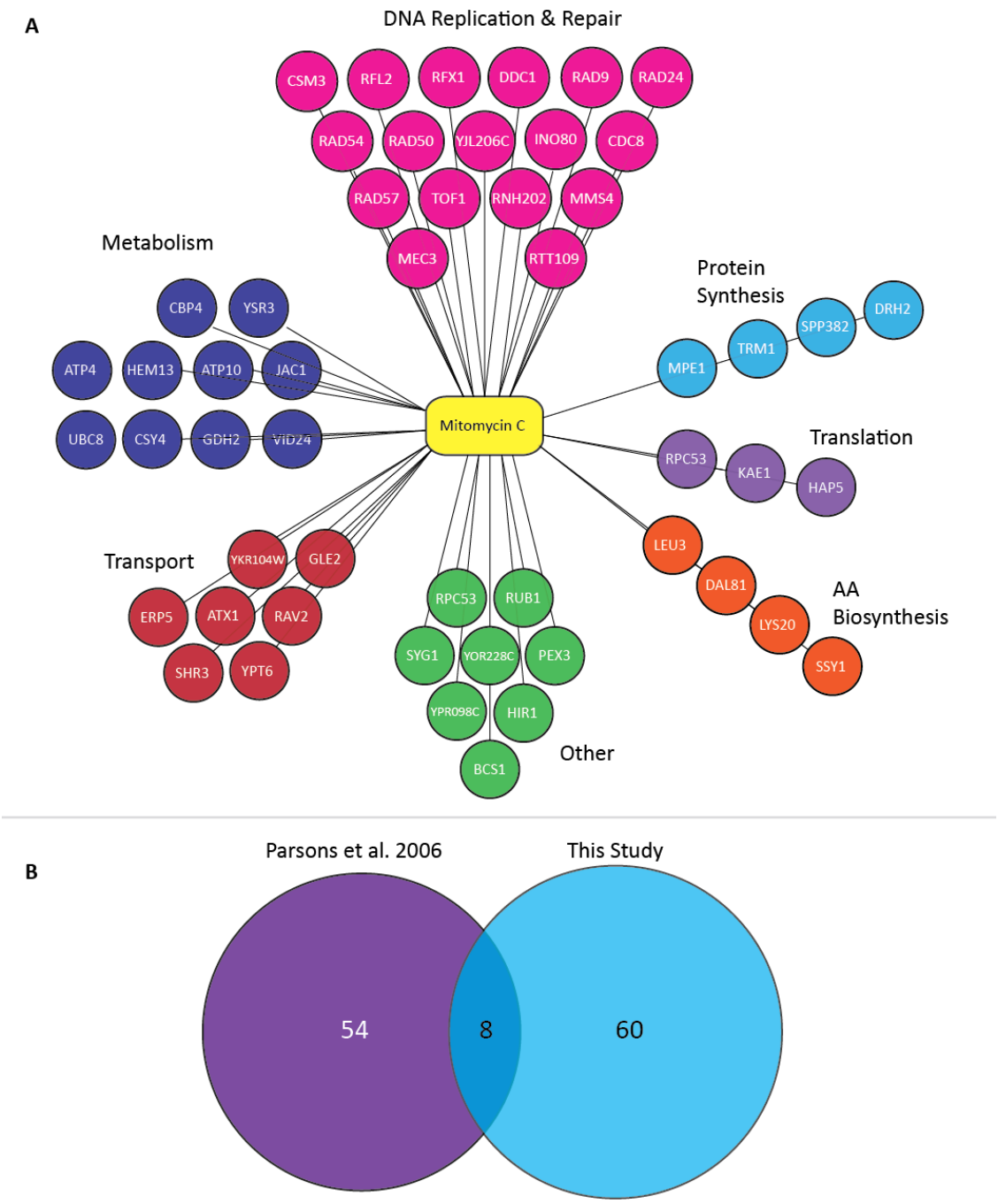

Figure 19. Analysis of the targeted processes predicted for mitomycin C.

A) The 68 predicted targets of mitomycin C were representative of 6 functional categories, classified according to a previously generated 19 category standard (Costanzo et al., (2010). These 68 genes were functionally enriched for genes participating in DNA repair ($p = 3.06 \times 10^{-7}$), response to endogenous stimulus ($p = 1.32 \times 10^{-6}$) and response to DNA damage stimulus ($p = 2.93 \times 10^{-6}$). **B)** These 68 correlated query genes overlapped significantly ($p = 0.003$) with the 62 correlated genes identified using the mitomycin C CG data from Parsons et al. (2006).

chemical genetic interaction with micafungin were deleted for *bck1Δ* (z = -5.355) and *rom2Δ* (z = -3.456) (See **Appendix C** for a list of the top 50 strains). To confirm the drug-hypersensitivity of these strains, a liquid growth assay was performed using benomyl and micafungin at the concentrations screened by the CG assay. Relative to a wild-type control, these four strains exhibited the expected drug-hypersensitivity (**Figure 20**). Testing the *bck1Δ* and *tub3Δ* mutants for sensitivity to benomyl and micafungin, respectively, demonstrated this sensitivity was specific to the expected deletion mutants. To further validate the use of a drug-hypersensitive background, strains harboring the four gene deletions in a wild-type background were also evaluated. At a benomyl concentration of 34.4 μM, the strains harboring *cin1Δ* and *tub3Δ* deletions in the *pdr1Δ pdr3Δ snq2Δ* drug-hypersensitive background demonstrated sensitivity, while strains harboring *cin1Δ* and *tub3Δ* deletions in a wild-type background were insensitive (**Figure 20**). The insensitivity of these wild-type background mutants persisted at a higher benomyl concentration of 50 μM. This result suggests the *pdr1Δ pdr3Δ snq2Δ* drug-hypersensitive background affords at least a 1.4 fold-reduction in the concentrations of benomyl required to inhibit cell growth. Similarly, only the strains harboring *bck1Δ* and *rom2Δ* deletions in the *pdr1Δ pdr3Δ snq2Δ* drug-hypersensitive background demonstrated sensitivity to 25nM of micafungin, while the strains harboring *bck1Δ* and *rom2Δ* deletions in a wild-type background were insensitive (**Figure 20**). When an elevated concentration of micafungin was tested, the *rom2Δ* mutant in the wild-type background exhibited sensitivity comparable to the *rom2Δ* drug-hypersensitive mutant. While the *bck1Δ* mutant in the wild-type background was not sensitized at this concentration, this result demonstrates the *pdr1Δ pdr3Δ snq2Δ* drug-hypersensitive background affords at least a 1.5 fold-reduction in the concentrations of micafungin required to inhibit cell growth. Overall, these results reinforce the power of using drug-hypersensitive strains to assay the RIKEN NPDepo library given the desire to conserve this NP resource.

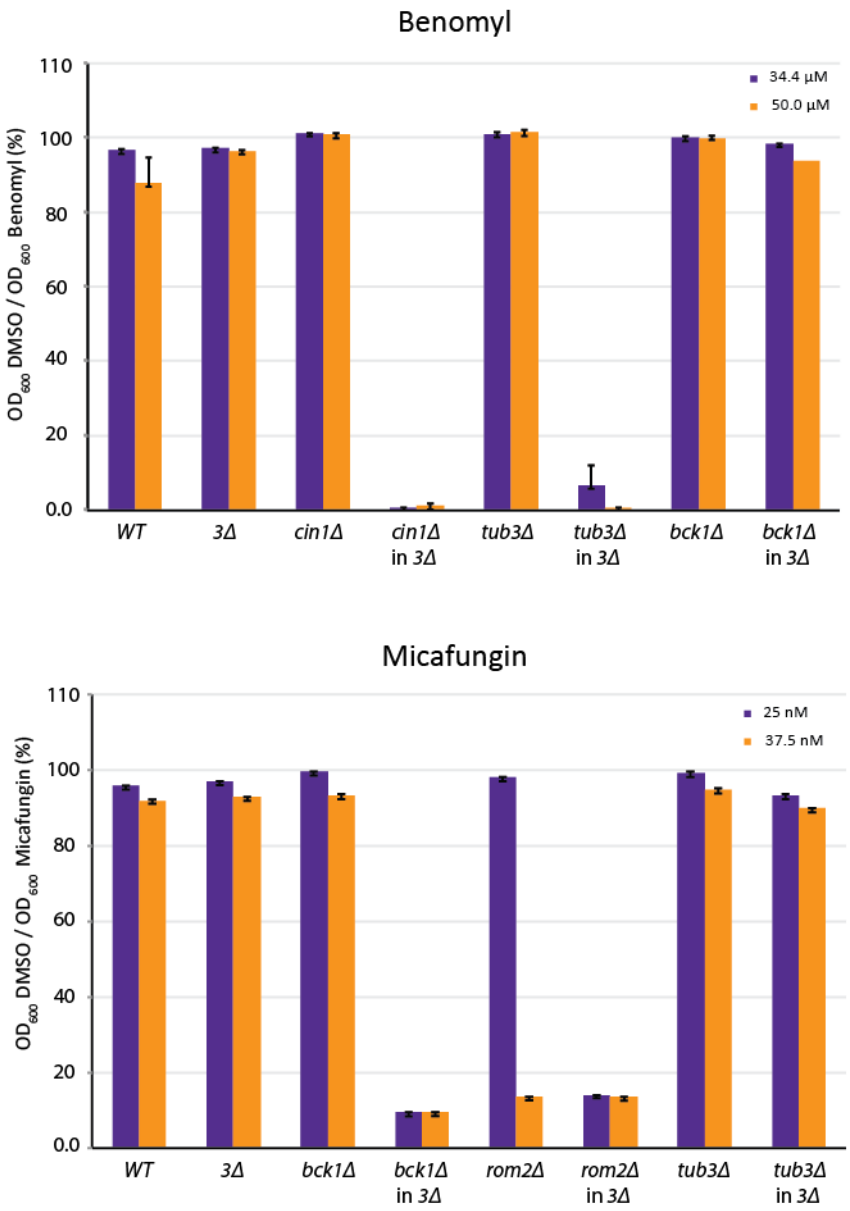

Figure 20. Confirming the chemical sensitivity of deletion mutants highlights the importance of screening compounds using the drug-hypersensitive deletion collection

The chemical-sensitivity of the top drug-sensitive deletion mutants identified from the adapted assay were confirmed by growing strains in the presence of the tested drug for 24 hours and recording the resulting optical density at 600nm. Strains tested harbored deletions either in a wild-type background or in the drug-hypersensitive *pdr1Δ pdr3Δ snq2Δ* background. Values plotted are percentages calculated by dividing the OD_{600} measured after growth in 2% DMSO by the OD_{600} measured after growth the marked concentration of compound and multiplying by 100. Y7092 was used as the WT control and the 3Δ *pdr1Δ pdr3Δ snq2Δ* mutant was used as the drug-hypersensitive control**.** (n = 3)

3.0 DISCUSSION

High-throughput CG analysis has been positioned as an ideal strategy for characterizing bioactive compounds, owing to the relative ease with which compounds can assayed to gain insights into their MOA. As such, there has been extensive work performed to develop and implement CG strategies for assaying compounds, particularly focused around model organisms like budding yeast. One outstanding limitation of many high-throughput compound analysis strategies is their methodology precludes screening natural products, which are generally available in very limited quantities and are thus incredibly valuable. This class of compounds has an outstanding resume as a candidate for the focus of future CG analysis. For instance, NPs have been the basis of a significant number of highly influential pharmaceuticals throughput history. Consequently, there is a desire to adapt existing high-throughput CG analysis strategies to make NP characterization accessible.

With this thesis, I described the development of a modified yeast CG interaction-profiling assay that will permit screening NP compounds. In particular, this strategy was developed with the ultimate goal of charactering a relatively novel collection of +40,000 NPs contained within the NPDepo at the RIKEN Institute in Japan. My main objective when developing the adapted methodology was to reduce the quantity of each NP that would be required for analysis. One strategy used to achieve this aim was to reduce the volume in which each experiment was performed 70%, affording a 3.5 fold reduction in the quantity of compound screened. Because the existing assay was optimized to analyze all yeast deletion mutants for drug hypersensitivity in a volume of 0.7mL, reducing the volume to 0.2mL required also reducing the complexity of the mutant pool screened. This was accomplished by selecting a subset of 491 deletion mutants that were diagnostic of genome-wide genetic interactions and could thus serve as a proxy for other deletion strains not assayed. To further reduce the quantity of each NP required for analysis, I recreated the 491 deletion mutants in a drug-hypersensitive background that was deficient for natural drug resistance machinery. By increasing the sensitivity of each strain, the concentration of compound required to inhibit cell growth was reduced, thereby affording a further reduction in the quantity of compound consumed by the assay. Once I had completed work to adopt these key modifications, I focused my efforts on implementing the assay to ensure

CG data yielded from this approach was both informative and comparable to data obtained using the traditional strategy. Through screening a series of control compounds for which CG data was available, I was able to demonstrate that in addition to exhibiting similarity to existing CG data, the CG profiles yielded by the adapted strategy could be used to provide insights into the biological processes targeted by each compound. Importantly, these concluding experiments endorsed the modified CG strategy I developed, priming the assay for application in NP characterization within the near future.

Selecting a subset of diagnostic strains to comprise the mutant pool screened raised an interesting question: could the non-essential yeast genome be simplified to a series of genes, capable of re-capturing the information contained within our knowledge base of genome-wide genetic interactions? The analysis of the 491 yeast deletion mutants chosen supports an affirmative answer to this question. First, we determined that genetic interactions restricted to the 491 deletion mutants were able to re-group interacting query genes as informatively as genome-wide interactions, within a genetic interaction clustergram. While this result was based on one specific metric developed to assess the 'correctness' of query-query grouping, interactions restricted to the diagnostic subset were also able to visually recapture the basic network topology present in a genome-wide yeast genetic interaction network-diagram (results not shown). Further corroboration of the diagnostic capability of the chosen 491 deletion mutants was obtained when we assessed the ability to use CG profiles restricted to interactions with the diagnostic genes to informatively predict the biological targets of various compounds. As hoped, target predictions formulated from CG profiles restricted to the diagnostic genes were comparable to the target-predictions made using the complete CG profiles. This result was particularly important given the ultimate goal of a CG analysis is to use the CG interaction profiles created to predict the MOA and targets of bioactive compounds screened. Overall, these results suggest that given an understanding of genome-wide genetic interactions within a particular organism, a diagnostic subset of genes can be chosen for use as a proxy to screen the remainder of the genome.

This outcome has interesting implications for developing CG strategies in other organisms. While our genome-wide genetic interaction knowledge base is currently restricted to budding yeast, there has been an extensive effort in recent years to map and understand interactions in other organisms, including *C. elegans* (Lehner et al. 2006; Byrne et al., 2007), *E. coli* (Babu et al., 2011), and mammalian cells (Lin et al., 2010). Once interactions are better understood in

these systems, subsets of mutants could be selected for analysis to either enhance existing CG strategies, transition available CG assays across organisms, or develop entirely new approaches. Achieving this aim would have exciting consequences for designing mammalian CG screens, as one overwhelming limitation of mammalian-based analyses is the immense complexity of mammalian genomes. Given advances in techniques for creating targeted or RNAi-based mammalian-gene knockouts and progress in completing large-scale deletion collections (Cullen & Arndt, 2005; Bu et al., 2010; Skarnes et al., 2011), the ability to select a diagnostic subset of mammalian genes could permit transitioning CG analysis to mammalian cells in the near future. These strategies would be important for following up on promising drug leads identified from a yeast-based CG approach, or analyzing compounds with the goal of identifying agents targeting mammalian-specific biological processes.

The drug-hypersensitive strain collection I created for this thesis is the first available large-scale deletion collection of its kind available. Currently, work is being undertaken to extend this resource genome-wide, which will ultimately provide a powerful reagent set for chemical-biology studies in *S. cerevisiae*. In addition to permitting the analysis of valuable natural compounds, use of these strains can save money by generally conserving compound resources. Additionally, the drug-sensitized background may permit analyzing compounds that don't typically exhibit antifungal activity. For instance, this feature could serve to better characterize human drugs that may not exhibit activity in wild-type yeast based on possible differences in the characteristics that determine compound bioactivity in these organisms. Even though yeast may lack the homologous mammalian target of a drug screened, this analysis can be used to understand a drug's secondary MOA, as described in **Section 2.12** for tamoxifen, desipramine and alverine citrate, and demonstrated previously (Ericson et al., 2008). These results could be especially beneficial for compounds that possess desirable therapeutic benefits but exhibit unspecific toxicity to understand the mechanisms through which the toxicity manifests. In doing so, researchers could focus on modifying existing chemicals they have invested significant time and money into instead of starting at square one in the drug-discovery pipeline.

One limitation of using drug-hypersensitive strains to characterize compounds is the results may not be directly transferable to strains with intact drug-resistance machinery. While a compound may be identified that elicits a desired physiological response or targets a known disease-causing gene, it is possible the compound may exhibit little-to-no activity in a wild-type strain, especially

if it exhibited low level activity in the drug-hypersensitive mutant. However, I believe this limitation can be addressed by one of two solutions.

Compounds identified that posses a desired target could be modified to increase bioavailability in strains with normal drug-resistance machinery. Over the last decade, extensive effort has been dedicated to characterizing the physical properties that contribute to bioavailability and drug-like features, such as the size of a compound size, lipophilicity and the number of hydrogen bonds it possesses (Lipinski et al., 2001). This information could be used to modify the chemical structure of compound leads identified from a CG analysis to enhance their potential as pharmaceutical agents. Generally, pharmaceutical profiling has focused on analyzing chemicals that have been synthetized to adhere to drug-like properties, which intrinsically restricts the range of chemicals that can be assessed. I believe that focusing on first screening all classes of chemicals and then modifying those that induce a desired physiological response is a better use of profiling efforts and would provide access to a larger diversity of chemical structures.

The second strategy to address the abovementioned limitation is the potential to use desirable compounds identified in concert with a second agent that inhibits MDR transport. This solution has shown particular promise for repurposing existing cancer therapies to treat cancers exhibiting resistance to a broad spectrum of chemotherapeutic agents (Wang et al., 2010). One phenotype common to many cancers either at diagnosis or following treatment is the overexpression of MDR transport genes like *ABCB1*, which encodes the p-glycoprotein (p-gp) ABC transporter most commonly linked to cancer drug-resistance (reviewed in Borowski et al., 2005). To overcome this phenotype and potentiate the activity of the currently used chemotherapeutic agents, there has been extensive effort to identify inhibitors of p-gp and other MDR transporters (e.g. Sikic et al., 1997; Chen et al., 2004; Coburger et al., 2009; Martelli et al., 2009). This concept has also been explored as a potential solution to combat the continued emergence of drug-resistance bacteria that exhibit increased expression of membrane drug-efflux pumps (Lomovskaya & Watkins, 2001; Mullin et al., 2004). Identifying and implementing a second agent that inhibits MDR transporters could therefore facilitate translating drug-leads or potential biological probes identified by the CG assay to cells with intact drug-resistance machinery.

Another important and related factor to consider when employing the drug-hypersensitive yeast collection for compound profiling is the potential that deleting the integral ABC membrane

proteins may affect cell membrane composition. Changes in yeast cell membrane composition have been shown previously to alter the cell's sensitivity to antifungal agents that target cell membrane components (Mukhopadhyay et al., 2002), suggesting NPs screened that possess similar targets may exhibit differential sensitivity in these strains. An additional change in the reaction of these strains to the NPs screened may result from screening in galactose-based media, since the biochemical composition of the cell likely changes as a consequence. To further explore these potential cellular variations introduced by using drug-hypersensitive strains and galactose-based media, it would be ideal to screen additional control compounds beyond the 5 described within this thesis, including several which are known to target the cell membrane. These screens should be performed using both mutants from the original deletion collection and the described drug hypersensitive deletion collection, in glucose and galactose based media. A detailed comparison of the resulting CG interaction profiles seeking to identify potential differences between the profiles yielded from each experimental condition should shed light on whether the described cellular changes affect the sensitivity of the strains tested and thus the output of the adapted assay. However, given the results obtained for the 5 compounds I described, I do not expect to observe significant differences.

By analyzing the CG profiles obtained for 5 previously characterized control compounds, I demonstrated the modifications adopted did not affect the quality of generated CG data. Specifically, I demonstrated that the CG profiles produced for the 5 drugs were comparable to CG profiles created for the same compounds using the traditional strategy. I also found that the CG profiles for the 5 compounds screened could be used to informatively predict their targeted biological processes. While these results support implementing the described assay for NP characterization, there are two outstanding issues encountered that must be first addressed. Among all of the experimental conditions tested (including the DMSO controls), I found that ~25% of the strains consistently had low raw sequencing counts (< 10 reads) or were completely absent from the data. This result suggests there are likely errors in either the barcodes or common priming sites marking those strains. The possibility of these errors is not a novel idea; re-characterization of the deletion collection by Smith et. al (2009) revealed errors in the barcode and common priming site sequences of many yeast deletion strains. More recently, our lab performed a bar-seq analysis of all haploid deletion mutants contained in the deletion collection and a comparison of the results demonstrated a significant overlap in the strains that were

missing or exhibited low counts in the barcode sequencing data (results not shown). Although I was able to obtain informative CG profiles using data from the remaining 75% of the strains I assessed, it would be desirable to select additional strains to replace the absent 25% to ensure the diagnostic pool remains representative of most biological processes. To accomplish this, I propose employing the original gene selection strategy but restricting the algorithm to choose genes whose barcodes yield good sequencing counts, based on our labs' analysis of the complete deletion collection. The second issue encountered that must be addressed involves a potential bias with either the experiment-specific indexing tags or the position with which drugs are screened in a 96-well plate. Specifically, we found that the CG interaction profiles of different/unrelated drugs assayed using the same bar-seq indexing tags exhibited a greater correlation then the CG profiles of different/unrelated drugs assayed using different bar-seq indexing tags (results not shown). Because each indexing tag is used at the same position in each 96-well drug-plate screened, it is not currently possible to differentiate whether this outcome results from the specific indexing tags used or a possible plate-position bias. These two potential biases could be assessed separately by scrambling the position of the indexing tags used and re-screening a series of control compounds. However, one characteristic unique to the correlated experiments is they exhibited either high or low sequencing counts relative to the average, suggesting the observed bias may be due to the indexing tags and could be addressed by developing a new set of indexing tags that yield similar sequencing counts. Within the coming months it is likely that solutions to these issues will be identified and implemented, permitting work meeting the overarching goal of this project to begin.

Once the aforementioned issues have been addressed, the immediate goal of this project is to begin analysis of the 40,000+ NPDepo library compounds. Priority will be given initially to characterizing a subset of around ~15,000 of the more abundant NPs, which have been isolated from diverse sources. Following screening these NPs, there are three classes of compounds I think would be ideal to follow-up on. First, NPs that may possess either fungal-specific targets or targets that are known homologues of human disease-causing proteins would be beneficial to further explore. These NPs could present leads for antifungal or human drug development, respectively. The second interesting class of compounds I would like to investigate is NPs that are predicted to target either an uncharacterized biological process or protein. These compounds could serve as ideal biological probes to advance our understanding of the molecular biology of

yeast. Finally, I think it would be interesting to investigate NPs that exhibit MOAs similar to existing drugs, but possess completely novel or divergent chemical structures. To rapidly identify compounds with related MOAs, the CG profiles of well-characterized compounds could be used to develop an index to group the NPs into broad MOA classes based on profile similarities (e.g. sterol binding or DNA damaging agents). The MOA of the classified compounds could then be confirmed using function-specific biochemical assays, such as media supplementation with ergosterol to evaluate whether a given NP binds sterols (described in Ho et al., 2009). Using the known chemical structure of all NPs in the NPDepo, one could then to determine whether novel structures are present within each functionally defined class of compounds. The potential to identify novel compound classes is not only interesting, but could allow for better understanding the chemical structures required to target specific biological processes.

In conclusion, over the course of my M.Sc. thesis work, I have developed a modified CG assay for characterizing natural compounds in budding yeast. By selecting a diagnostic subset of yeast deletion mutants to screen and employing novel drug-hypersensitive background, I have created a powerful resource for assaying novel bioactive compounds. In using this strategy to evaluate the relatively untapped NPDepo compound library within the near future, I am hopeful that valuable biological insights and novel pharmaceutical leads will gained.

4.0 MATERIALS AND METHODS

4.1 Yeast strains used in this study

The yeast strains used in this study are summarized in **Table 5.**

4.2 Precision-recall analysis for quantifying the quality of query-query grouping by comparison to a GO co-annotation standard

Evaluating query-query grouping within a genetic-interaction clustergram was important for selecting a diagnostic subset deletion strains and assessing whether the strains chosen could informatively re-cluster genetic interaction data. This analysis was performed by comparing query-query grouping to a GO co-annotation gold standard and using a precision-recall analysis to quantify the comparison. A precision-recall analysis is a commonly used metric to evaluate the relevance or informative capacity of data contained within a list. For quantifying the comparison of query-query grouping to a GO co-annotation gold standard, the precision-recall analysis conceptually proceeded as follows:

1. Comparing query-query grouping in an interaction clustergram to the GO co-annotation gold standard yielded a binary quality score for each gene pair. A score of 1 was assigned to the gene pairs that were annotated to the same biological process based on the GO co-annotation gold standard, while a score of 0 was assigned to the gene pairs that were not annotated to the same biological process. From the ~1700 query gene by ~3890 array gene interaction dataset, 1699 gene pair scores were produced.

2. All gene pair scores were then ordered from greatest to smallest, based on their assigned score of 1 or 0. Precision was then calculated across the complete range of possible recalls. The value of the computed precision ranges from 0-1, where a precision of 0 would indicate there were no gene pair scores of 1 in the list for the recall applied, while a score of 1 would indicate all gene pair scores were 1 for the recall applied. Similarity, the value of the recall used ranges from 0-1, where a recall of 0 indicates no scores were used to calculate precision and a recall of 1 would indicate all scores present were used to calculate precision.

Table 5. *Saccharomyces cerevisiae* strains used in this study.

Strain	Genotype	Reference
9Δ	*MAT***a** *yor1Δ::HisG snq2Δ::HisG pdr5Δ::HisG pdr10Δ::HisG pdr11Δ::HisG ycf1Δ::HisG pdr3Δ::HisG pdr15Δ::HisG pdr1Δ::HisG*	Rogers et al., (2001)
pdr1Δ/pdr3Δ	*MATα pdr1Δ::natMX pdr3Δ::KI.URA3 can1Δ::STE2pr-Sp_his5 lyp1Δ his3Δ1 leu2Δ0 ura3Δ0 met15Δ0*	This study
SGA query strain	*MATα can1Δ::STE2pr-Sp_his5 lyp1Δ his3Δ1 leu2Δ0 ura3Δ0 met15Δ0*	Tong & Boone, (2006)
BY4741	*MAT***a** his3Δ1 leu2Δ0 LYS2 met15Δ0 ura3Δ0	Brachman *et al.*, (1998)
BY4743	*MAT***a**/α *his3Δ1/his3Δ1 leu2Δ0/leu2Δ0 LYS2/lys2Δ0 met15Δ0/MET15 ura3Δ0/ura3Δ0*	Brachman *et al.,* (1998)
pdr1Δ/pdr3Δ/xxxΔ	*MATα pdr1Δ::natMX pdr3Δ::KI.URA3 xxxΔ::KI.LEU2 can1Δ::STE2pr-Sp_his5 lyp1Δ his3Δ1 leu2Δ0 ura3Δ0 met15Δ0*	This study
pdr1Δ/pdr3Δ/snq2Δ	*MATα pdr1Δ::natMX pdr3Δ::KI.URA3 snq2Δ::KI.LEU2 can1Δ::STE2pr-Sp_his5 lyp1Δ his3Δ1 leu2Δ0 ura3Δ0 met15Δ0*	This study
pdr1Δ	*MATα pdr1Δ::natMX can1Δ::STE2pr-Sp_his5 lyp1Δ his3Δ1 leu2Δ0 ura3Δ0 met15Δ0*	This study
pdr3Δ	*MATα pdr3Δ::KI.URA3 can1Δ::STE2pr-Sp_his5 lyp1Δ his3Δ1 leu2Δ0 ura3Δ0 met15Δ0*	This study
***pdr1Δ/snq2* Δ**	*MATα pdr1Δ::natMX snq2Δ::KI.LEU2 can1Δ::STE2pr-Sp_his5 lyp1Δ his3Δ1 leu2Δ0 ura3Δ0 met15Δ0*	This study
***pdr3Δ/snq2* Δ**	*MATα pdr3Δ::KI.URA3 snq2Δ::KI.LEU2 can1Δ::STE2pr-Sp_his5 lyp1Δ his3Δ1 leu2Δ0 ura3Δ0 met15Δ0*	This study
***pdr1Δ/pdr3Δ/snq2Δ* diploid**	*MAT***a**/α *pdr1Δ::natMX/PDR1 pdr3Δ::KI.URA3/PDR3 snq2Δ::KI.LEU2/SNQ2 can1Δ::STE2pr-Sp_his5/CAN1 lyp1Δ/LYP1 his3Δ1/ his3Δ1 leu2Δ0/ leu2Δ0 ura3Δ0/ ura3Δ0 met15Δ0/met15Δ0*	This study
***rad52Δ* in 3Δ**	*MATα pdr1Δ::natMX pdr3Δ::KI.URA3 snq2Δ::KI.LEU2 rad52Δ::kanMX can1Δ::STE2pr-Sp_his5 lyp1Δ his3Δ1 leu2Δ0 ura3Δ0 met15Δ0*	This study
***vps8Δ* in 3Δ**	*MATα pdr1Δ::natMX pdr3Δ::KI.URA3 snq2Δ::KI.LEU2 vps8Δ::kanMX can1Δ::STE2pr-Sp_his5 lyp1Δ his3Δ1 leu2Δ0 ura3Δ0 met15Δ0*	This study
rad52Δ	*MAT***a** *his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 rad52Δ::kanMX*	Giaever et al., (2002)
vps8Δ	*MAT***a** *his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 vps8Δ::kanMX*	Giaever et al., (2002)
***cin1Δ* in 3Δ**	*MATα pdr1Δ::natMX pdr3Δ::KI.URA3 snq2Δ::KI.LEU2 cin1Δ::kanMX can1Δ::STE2pr-Sp_his5 lyp1Δ his3Δ1 leu2Δ0 ura3Δ0 met15Δ0*	This study
***tub3Δ* in 3Δ**	*MATα pdr1Δ::natMX pdr3Δ::KI.URA3 snq2Δ::KI.LEU2 tub3Δ::kanMX can1Δ::STE2pr-Sp_his5 lyp1Δ his3Δ1 leu2Δ0 ura3Δ0 met15Δ0*	This study

***bck1Δ* in 3Δ**	*MATα pdr1Δ::natMX pdr3Δ::KI.URA3 snq2Δ::KI.LEU2 bck1Δ::kanMX can1Δ::STE2pr-Sp_his5 lyp1Δ his3Δ1 leu2Δ0 ura3Δ0 met15Δ0*	This study
***rom2Δ* in 3Δ**	*MATα pdr1Δ::natMX pdr3Δ::KI.URA3 snq2Δ::KI.LEU2 rom2Δ::kanMX can1Δ::STE2pr-Sp_his5 lyp1Δ his3Δ1 leu2Δ0 ura3Δ0 met15Δ0*	This study
cin1Δ	*MAT***a** *cin1Δ::kanMX his3Δ1 leu2Δ0 ura3Δ0 met15Δ0*	Giaever et al., (2002)
tub3Δ	*MAT***a** *tub3Δ::kanMX his3Δ1 leu2Δ0 ura3Δ0 met15Δ0*	Giaever et al., (2002)
bck1Δ	*MAT***a** *bck1Δ::kanMX his3Δ1 leu2Δ0 ura3Δ0 met15Δ0*	Giaever et al., (2002)
rom2Δ	*MAT***a**, *rom2:kanMX his3Δ1 leu2Δ0 ura3Δ0 met15Δ0*	Giaever et al., (2002)

*YDC – yeast deletion collection

Generally, as recall increases the corresponding precision calculated decreases. One clustergram is said to have more informative query-query grouping relative to another clustergram if the precision remains high for a larger recall. Determining the area under a precision-recall curve is representative of the curve shape and can therefore be used to directly compare PR analyses performed on different datasets. To perform function-specific PR analysis, the GO co-annotation gold standard was subdivided into 19 functional categories and gene pair scores were obtained only for the genes present in the standard.

4.3 Comparing interaction profiles using two-dimensional hierarchical clustering

To visualize the comparison of interaction profiles, two-dimensional agglomerative hierarchical clustering using Pearson's correlation and average linkage was applied to the data using Cluster 3.0 software. Clustering was restricted to genes exhibiting an interaction z-score less than 0.

4.4 Computing the correlation between interaction profiles to predict the targeted biological processes and the MOA of bioactive compounds

Identifying which genetic interaction profiles bear similarity to the CG interaction profile produced by a compound of interest can be used to predict the biological processes targeted. Similarly, MOA insights can be gained by identifying compounds that yielded similar CG interaction profiles. To perform these comparisons, the similarity between different interaction profiles was determined by computing cosine similarity coefficients using the following equation:

$$similarity\ coefficient = cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i\ x\ B_i}{\sqrt{\sum_{i=1}^{n}(A_i)^2}\ x\ \sqrt{\sum_{i=1}^{n}(B_i)^2}}$$

Where A is a data vector for one interaction profile and B is a data vector for the second interaction profile.

A similarity coefficient of 1 indicates the two interaction profiles compared are identical while a similarity coefficient of 0 indicates the compared interaction profiles exhibit no similarity. This metric resembles the commonly used Pearson correlation coefficient; the magnitude of correlation coefficients calculated using cosine similarity and Pearson correlation are usually very similar for comparing interaction profiles. The main difference between these approaches is that Pearson correlations are calculated by first centering data around an experimental mean, which may change the sign (+/-) of the data analyzed if the experimental mean is not zero. Because the sign of an interaction score indicates whether the interaction is negative or positive, it was preferred to maintain the correct sign of all scores analyzed. In addition, not centering the data enables cosine similarity coefficients to be computed faster then Pearson correlation coefficients. This characteristic is ideal given that the ultimate goal of this project is to screen the entire NPDepo library, which will yield +40,000 CG interaction profiles to analyze.

4.5 A high-throughput chemical-halo assay for drug-sensitivity testing

The chemical sensitivity of various deletion mutants was assessed using a high-throughput chemical-halo assay. After growing yeast strains overnight to saturation, cultures were standardized to an OD_{600} = 4.0 and 2 mL was added to a 50 mL stock of 2% YEPD + 2% agar. Seeded plates were prepared by pouring 10 mL of seeded culture into NUNC square plates and drying for 10 minutes to facilitate compound absorption. Robotic pinning with the Biotec ADS384 was used to transfer 0.2μL of each natural product to the seeded plates at a density of 96 compounds per plate; 440 natural products were evaluated in total. After incubating the plates for 18 – 24 hours at 30°C, plates were imaged and the widths of visible areas of growth inhibition were measured using JMicrovision software. A compound was deemed toxic if it generated an area of growth inhibition with a diameter greater than 1mm.

4.6 Staining and visualization of a yeast budding defect

To visualize and compare the gross morphological appearance of the 9Δ mutant to that of the *pdr1Δ pdr3Δ* mutant, the *pdr1Δ pdr3Δ snq2Δ* mutant and a wild-type strain, cells were grown to saturation overnight and sub-cultured the following day to early log-phase (OD_{600} = 0.2 – 0.3). Cells were then fixed with formaldehyde and stained with 100µg/mL of a concanavalin A, Texas Red® conjugate (Con A; Invitrogen) for 30 minutes at 30°C. Con A is a lectin that binds with strong affinity to α-linked mannose homopolymers that are found in the yeast cell wall and are concentrated in the bud scars of yeast cells following division (Tkacz et al., 1971). All cell images were captured at 63X magnification using a DMI 6000B fluorescence microscope (Leica Microsystems) equipped with a spinning-disk head, a 561nm laser (Cobalt AB) and an ImagEM charge-coupled device camera (Hamamatsu C9100-13, Hamamatsu Photonics). Images were processed using Volocity software (Improvision).

4.7 Deleting multi-drug resistance genes via PCR mutagenesis

A one-step PCR-directed mutagenesis strategy was used to create the *pdr1Δ pdr3Δ* double mutant, all 19 xxxΔ *pdr1Δ pdr3Δ* triple mutant strains, and the 5 single and double mutants with all possible combinations of *pdr1Δ, pdr3Δ, snq2Δ* deletions. *PDR1* was deleted in the SGA query strain by replacement with the *natMX* antibiotic resistance marker, which provides resistance to the drug nourseothricin (NAT). To create the *pdr1Δ pdr3Δ* double mutant, *PDR3* was then deleted in the *pdr1Δ* mutant by replacement with the *K.lactis URA3* autotrophic marker, which permits cells to grow on synthetic media lacking uracil. All 19 selected genes were deleted individually in the *pdr1Δ pdr3Δ* mutant by replacement with the *K.lactis LEU2* autotrophic marker, which permits cells to grow on synthetic media lacking leucine. The *pdr1Δ, pdr3Δ,* and *snq2Δ* single or double mutants were created by replacement with the *natMX*, *K.lactis URA3* or *K.lactis LEU2* markers, respectively. The *natMX*, *Kl.URA3* and *Kl.LEU2* markers were amplified from plasmids using primers designed with 50 base pairs of sequence homologous to regions upstream and downstream of the deleted genes (**Table 6**). PCR amplicons were transformed into the appropriate strains using lithium acetate and polyethylene glycol-based transformations (Gietz & Schiestl, 2007). Deletion of the native gene and integration of the marker at the correct

Table 6. A summary of primers used to create drug-sensitive deletion strains.

Targeted Gene	Deletion Marker	Primer Sequence	Plasmid
PDR1	*natMX*	**Forward**: CATCTCAGCCAAGAATATACAGA AAAGAATCCAAGAAACTGGAAG**ACATGG AGGCCCAGAATACCC**	pAG25 (Goldstein & McCusker, 1999)
		Reverse: CATCTCAGCCAAGAATATACAGA AAAGAATCCAAGAAACTGGAAG**AGATCT GTTTAGCTTGCCTTGTCC**	
PDR3	*KI.URA3*	**Forward:** ACTGCATCAGCAGTTTTATTAAT TTTTTCTTATTGCGTGACCGCA**CGGAGAC AATCATATGGGAG**	pUG72 (Guldener et al., 2002)
		Reverse: CCATTTACTATGGTTATGCTCTGC TTCCCTATTTCTTTTGCGTTT**TCTGGAGGA AGTTTGAGAGG**	
Each of the 19 selected genes	*KI.LEU2*	**Forward:** *50 bases US. - **ATGTCTGCCC CTAAGAAGAT**	pUG73 (Guldener et al., 2002)
		Reverse: *50 bases DS. - **TTAAGCAAGG ATTTTCTTAACTTCTTCGG**	

* 50 bases of sequence homologous to regions upstream (US) or downstream (DS) of the gene targeted for deletion.

loci was confirmed using a series of PCR confirmations. Confirmation primers were designed specific to regions both flanking the integration site and internal to the inserted marker to interrogate both the full length of the inserted marker and the 5' and 3' boundaries.

4.8 Comparing chemical sensitivity with dose-response testing

A dose-response assay was performed to compare the chemical sensitivity of the 19 triple mutant strains created. Each strain was grown in YEPD media overnight to saturation, diluted 1000 fold, and sub-cultured in fresh media at 30°C until an optical density of 0.0625 was reached. 100µL of each culture was then transferred to a 96-well round-bottom plate and wells were inoculated with one of eight compounds at the appropriate concentration. The eight compounds selected and the range of concentrations at which they were tested is listed in **Table 7**. Prepared plates were covered with an optically clear adhesive lid and the optical density was measured at 600 nm in a plate reader (BioHITS) for 24 hours. The final OD_{600} values for each strain were plotted against the $\log_{10}$ of the drug concentration tested and the data was fit with a four-parameter logistic regression model using SigmaPlot™ graphing software. This analysis was used to calculate the IC_{50} values for each drug, as the equation that defines the four-parameter logistic model estimates the value of the x-variable required to yield a 50% maximal response in the value of the y-variable:

$$f(x) = \left(\frac{A - D}{\left(1 + \left(\left(\frac{x}{C}\right)^{B}\right)\right)} \right) + D$$

where: A = the response measured at x = 0 or when [drug] = 0

B = the slope

C = the value of x where the response equals ½ max or the concentration of drug Inhibiting cell growth by 50% (**IC_{50}**).

D = the response measured at x = ∞ or when [drug] = ∞

x = independent variable (drug concentration)

$f(x)$ = dependent variable (Optical density)

Table 7. Drugs and drug-concentrations tested with a dose-response assay to evaluate the drug-sensitivity of 19 triple mutant strains.

Drug (acronym)	Concentrations tested (μg/mL)
Amphotericin B (AMB)	0, 0.01, 0.02, 0.05, 0.1, 0.12, 0.2, 0.4, 0.8
Anisomycin (ANIS)	0, 0.05, 0.1, 0.25, 0.5, 0.75, 1, 2, 4
Cyclohexamide (CYX)	0, 0.00075, 0.0015, 0.0037, 0.0075, 0.0087, 0.01, 0.02, 0.04
Daunorubicin (DAU)	0, 0.1, 0.2, 0.5, 1, 1.25, 1.5, 3, 6
Ketoconazole (KETO)	0, 005, 0.01, 0.025, 0.05, 0.075, 0.1, 0.2, 0.4
Micafungin (MICA)	0, 0.0025, 0.005, 0.0125, 0.025, 0.0275, 0.03, 0.06, 0.12
Terbinafine (TERB)	0, 0.05, 0.1, 0.25, 0.5, 0.75, 1, 2, 4
Nystatin (NYS)	0, 0.05, 0.1, 0.25, 0.5, 0.55, 0.6, 1.2, 2.4

A strain's fold-change in sensitivity to each of the eight drugs relative to wild-type was calculated by dividing the IC^{50} value obtained for the BY4741 wild-type strain by the IC^{50} value obtained for the mutant. These fold-change values were compared among the 19 mutants to determine which strain(s) exhibited the greatest increase in sensitivity to the widest range of compounds. This strategy was also used to compare the chemical hypersensitivity of the *pdr1Δ pdr3Δ snq2Δ* strain when culturing in YPD vs. YPGal media. The procedure used for this analysis is identical to described above, except cells were cultured in YP media supplemented with either 2% dextrose or 2% galactose as the carbon source. The concentrations of the six drugs tested are listed in **Table 8**.

4.9 Tetrad analysis

To evaluate whether the *pdr1Δ pdr3Δ snq2Δ* drug-hypersensitive mutant possessed the sporulation defect observed with the 9Δ strain, it was first crossed to the BY4741 *MAT***a** haploid strain to create a *pdr1Δ/PDR1, pdr3Δ/PDR3, snq2Δ/SNQ2* homozygous diploid. After culturing in YEPD to an OD_{600} of 2.5-3.0, diploid cells were washed with sterile water and suspended in SPO media supplemented with 0.25X the standard quantity of histidine and methionine. After incubating for 3-5 days at room temperature, cultures were inspected microscopically for the presence of asci. To ensure the deletion markers segregated 1:1 among the spores in each ascus, cells were washed with sterile water and suspended in 50μL of 0.5 mg/mL zymolyase (in 1M sorbitol) to digest the wall of the ascus. Individual spores were then separated using a Singer MMS micromanipulation system (Singer) on YEPD plates and were grown for 2 days at 30°C. Finally, spores were replica plated onto YEP+NAT, SD-URA and SD-LEU media to evaluate segregation of the *pdr1Δ::NAT, pdr3Δ::KI.URA3* and *snq2Δ::KI.LEU2* deletions.

Table 8. Drugs and drug-concentrations tested with a dose-response assay to evaluate drug-sensitivity when culturing in YPD vs. YPGal media.

Drug	Concentrations tested (μg/mL)
Amphotericin B	10, 5, 1, 0.75, 0.5, 0.25, 0.1, 0.075
Cyclohexamide	1, 0.5, 0.25, 0.125, 0.0625, 0.03125, 0.0156, 0.0078
Ketoconazole	2, 1, 0.5, 0.125, 0.0625, 0.03125, 0.0156
Daunorubicin	50, 10, 7.5, 5, 2.5, 1.5, 1, 0.5
Rapamycin	0.1, 0.05, 0.025, 0.01, 0.005, 0.0025, 0.001, 0.0005
Anisomycin	20, 10, 8, 6, 4, 2, 1

4.10 Synthetic genetic array strategy for transferring yeast deletion cassettes to the drug hypersensitive background

The created *pdr1Δ::natMX*; *pdr3Δ::KI.URA3*; *snq2Δ::KI.LEU2 MATα* query strain carried the *can1Δ::STEpr-SP_his5* and *lypΔ* SGA reporters. *STEpr-SP_his5* is an auxotrophic marker that allows only *MAT***a** cells to grow in the absence of histidine, while the *can1Δ* and *lypΔ* deletions allow haploid cells to grow in the presence of the drugs canavanine and thialysine, respectively. The *MATα* query strain was crossed to an ordered array of the 491 *MAT*a *xxxΔ::KanMX* deletion mutants and the resulting zygotes were selected for on synthetic media lacking uracil and leucine with NAT and geniticin (G418). The heterozygous diploids were transferred to media with reduced carbon and nitrogen to induce sporulation and the formation of haploid meiotic progeny. The resulting spores were transferred to synthetic media lacking histidine and containing canavanine and thialysine to select for the *MAT***a** meiotic progeny. Cells were then transferred to synthetic media lacking uracil and containing NAT to select for growth of cells carrying both the *pdr3Δ::KI.URA3* and *pdr1Δ::NAT* deletions. Finally, these cells were transferred to synthetic media lacking uracil & leucine and containing G418 & NAT to select for the desired *pdr1Δ, pdr3Δ, snq2Δ, xxxΔ* triple mutants. Mating, diploid selection, sporulation and haploid selections steps were carried out in a 96-colony format using robotic replica pinning on a Virtek Colony Arrayer (Biorad Laboratories). This protocol was adapted from Tong et al. (2001).

4.11 Quality control of SGA output strains

Verifying strain ploidy using flow cytometry analysis.

Flow cytometry was used to verify the integrity of the adapted SGA pipeline by confirming all 491 SGA output strains were haploid. After growing strains overnight to saturation in 96-deep-well plates, cells were transferred to fresh YEPD media and subcultured at 30°C to mid-log phase. After fixing cells in 70% ethanol at 4°C overnight, cells were washed with 0.2 mg/mL RNase A (Qiagen) for 3 hours at 37°C, followed by 2 mg/mL Proteinase K (Roche Diagnostics) for 1 hour at 50°C. Cells were then stained with 4.5nM of the light sensitive Sybrgreen nucleic acid stain (Invitrogen), sonicated at low intensity for 1 – 2 seconds per well and covered in foil until analysis. After passing samples through the Guava Easycyte flow cytometer (Millipore)

using the Guava Express Plus program, results were analyzed with FlowJo™ software. Strains were classified as haploid if cells possessed 1n or 2n DNA peaks comparable those produced by the haploid control strain, and did not exhibit the 4n peak produced by the diploid control strain.

Identifying the mating type of yeast strains using a standard pheromone halo test

Mating type tests were performed on single colonies streaked from the SGA output array to verify integrity of the SGA pipeline. Five colonies were randomly selected for testing from each of the six array plates. Strains were streaked onto YEPD media, grown for two days at 30°C and then replica plated onto YEPD media overlaid with either *MAT***a** *bar1Δ* or *MAT*α *sst2Δ* mutant strains. Replica plated strains were grown for 1 day at 30°C. Colonies were identified as *MAT***a** or *MAT*α based on the presence or absence of a pheromone-induced halo of G1 arrested cells surrounding the replica-plated strains. The strains that induced the formation of a halo on *sst2Δ* lawns were classified as *MAT***a**, while the strains that induced the formation of a halo on *bar1Δ* lawns were classified as *MAT*α.

Testing for presence of four deletion markers

Single colonies streaked from the SGA output array were evaluated for the presence of the *natMX*, *KI.URA3*, *KI.LEU2* and *kanMX* deletion markers to further verify the integrity of the SGA pipeline. The same five randomly selected colonies used for the mating type test were evaluated. Strains were streaked onto YEPD media, grown for two days at 30°C and then replica plated onto the following selection media to interrogate for the presence of a deletion marker: YEPD + clonNAT to test for *natMX*, SD – URA to test for *KI.URA3*, SD – LEU to test for *KI.LEU2* and YEPD + G418 to test for *kanMX*. Strains that were able to grow on selection media carried the appropriate deletion marker.

4.12 Liquid growth curve analysis

Liquid growth curve analysis was performed to compare the fitness of strains harboring *RAD52* and *VPS8* deletions in the drug-hypersensitive background to strains harboring these deletions in a WT background. This analysis was also used to confirm the top benomyl and micafungin drug-

sensitive strains. Yeast cultures were grown to saturation overnight in YEPD, diluted 1000-fold, and then grown at 30°C until an optical density of 0.06 was reached. 100 µl of each sample was then transferred to a round-bottomed 96-well plate and covered with an optically clear adhesive lid. Optical density was measured at 600 nm in a TECAN™ plate reader every 15 min for 24 h.

4.13 Measuring the sensitivity of yeast strains with an adapted barcode sequencing approach

A pool of the 491 drug-hypersensitive deletion mutants was constructed by first pinning frozen 96-well glycerol stocks of each strain onto Nunc Omni Tray plates containing YEPD + G418 solid media and incubating for 2 days at 30°C. Each plate was then flooded with 10mL of YEPD liquid media and a cell spreader was used to re-suspend grown colonies. The resulting cell suspensions were then transferred to a 50mL conical tube where glycerol was added to a 15% final concentration. Finally, the pool was adjusted to a final concentration of 50 OD_{600}/mL by dilution or centrifugation and stored at -80°C until required.

To assay the mutant pool for drug-hypersensitivity, cells were thawed and diluted to an OD_{600} of 0.0625 in YEP + 2% galactose in a 96-well round-bottom plate. Cultures were then spiked with the desired concentration of drug or with a 2% DMSO control. After growing for 5 generations at 30°C, cells from each well were harvested by centrifugation. Genomic DNA was purified from the harvested cells by re-suspending in 125 µL of zymolyase buffer (1 mg/mL) and using the QIAextractor (Qiagen) as per manufacturers instructions, with a 100 µL elution volume. Individual barcodes were then PCR amplified using the following PCR mix and protocol:

PCR Mix with a 50 µL final volume:

- 44 µL of Platinum PCR supermix (Invitrogen)
- 2 µL of extracted genomic DNA
- 2 µL U1 primer, 5µM (see **Table 9**)
- 2 µL U2 primer, 5µM (see **Table 9**)

PCR cycle:

- 1 cycle – 4 minutes at 95°C
- 30 cycles – 30 seconds at 95°C
  30 seconds at 55°C
  30 seconds at 72°C
- 1 cycle – 7 minutes at 72°C

The resulting 150 bp amplicons from each well in a 96-well plate were then pooled and run on a 12% acrylamide gel to purify the PCR products. The appropriate band was excised and DNA was extracted using the crush and soak method (Sambrook & Russell, 2001). Sequencing was then performed on the Genome Analyzer IIx (Illumina) as described previously (Smith et al., 2011), except using 21 cycles to capture the barcode and 8 cycles to capture the indexing tag. After computing the $\log_2$ of the absolute barcode count for each strain, lowess normalization was performed to average counts with respect to the DMSO control. The DMSO control used was an average of 16 total DMSO experiments, as 4 DMSO control wells were present in each of the four 96-well plates screened. An interaction z-score was then calculated for all strains tested in each experiment to identify the significance of barcode-count deviation from the control using the following formula:

$$z - score = \frac{barcode\ count - average\ control\ count}{standard\ deviation}$$

where the standard deviation used in the calculation accounted for the maximal variation introduced by both the DMSO control and the drug screened. Z-scores calculated for each strain were then compiled to create CG interaction profiles for the five compounds described.

Table 9. Primers used for barcode amplification

The sequence of primers used to amplify barcodes for 96-plex next generation sequencing were obtained from Smith et al. (2011).

Primer Name	Primer Sequence (5' – 3')
U1 (forward)	*CAAGCAGAAGACGGCATACGAGCTCTTCCGATCT***GATGTCCACGAGGTCTCT**
U2 (reverse)	*AATGATACGGCGACCACCGACACTCTTTCCCTACACGACGCTCTTCCGATCT*XXXXXXXX **GTCGACCTGCAGCGTACG**

Italic – Illumina sequence required for cluster formation on flow cell
Underlined – one of 96 8-mer multiplex tags
Bold – common priming site sequence

5.0 References

Arita, Y., Nishimura, S., Matsuyama, A., Yashiroda, Y., Usui, T., Boone, C., Yoshida, M. (2011). Microarray-based target identification using drug hypersensitive fission yeast expressing ORFeome. *Molecular Biosystems,* **7**: 1463-1472.

Babu, M., Diaz-Mejia, J., Vlasblom, J., Gagarinova, A., Phanse, S., Graham, C., Yousif, F., Ding, H., Xiong, X., Nazarians-Armavil, A., Alamgir, M., Ali, M., Pogoutse, O., Pe'er, A., Arnold, R., Michaut, M., Parkinson, J., Golshani, A., Whitfield, C., Wodak, S.J., Moreno-Hagelsieb, G., Greenblatt, J.F., Emili, A. (2011). Genetic interaction maps in Escherichia coli reveal functional crosstalk among cell envelope biogenesis pathways. *Plos Genetics,* **7**: 1-15.

Balzi, E., Wang, M., Leterme, S., Van Dyck, L., Goffeau, A. (1994). PDR5, a novel yeast multidrug resistance conferring transporter controlled by the transcription regulator PDR1. *Journal of Biological Chemistry,* **269**: 2206-2214.

Bauer, B.E., Wolfger, H., Kuchler, K. (1999). Inventory and function of yeast ABC proteins: about stress, sex, pleiotropic drug and heavy metal resistance. *Biochimica et Biophysica Acta,* **1461**: 217-236.

Bharucha, N., Chabrier-Rosello, Y., Xu, T., Johnson, C., Sobczynski, S., Song, Q., Dobry, C.J., Eckwahl, M.J., Anderson, C.P., Benjamin, A.J., Kumar, A., Krysan, D.J. (2011). A large-scale complex haploinsufficiency-based genetic interaction screen in Candida albicans: analysis of the RAM network during morphogenesis. *Plos Genetics,* **7**: 1-12.

Birrell, G.W., Giaever, G., Chu, A.M., Davis, R.W., Brown, J.M. (2001). A genome-wide screen in Saccharomyces cerevisiae for genes affecting UV radiation sensitivity. *Proceedings of the National Academy of Sciences,* **98**: 12608-12613.

Borowski, E., Bontemps-Gracz, M.M., Piwkowska, A. (2005). Strategies for overcoming ABC-transporters-mediated multidrug resistance (MDR) if tumor cells. *Acta Biochimica Polonica,* **52**: 609-627.

Botstein, D., Chervitz, S.A., Cherry, J.M. (1997). Yeast as a Model Organism. *Science,* **277**: 1259-1260.

Brachmann, C.B., Davies, A., Cost, G. J., Caputo, E., Li, J., Hieter, P., Boeke, J.D. (1998). Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* **14**: 115–132.

Brown, J.A., Sherlock, G., Myers, C.L., Burrows, N.M., Deng, C., Wu, H.I., McDann, K.E., Troyanskaya, O., Brown, J.M. (2006). Global analysis of gene function in yeast by quantitative phenotypic profiling. *Molecular Systems Biology*, **2**: 1-9.

Bu, L., Gao, X., Jiang, X., Chien, K., Wang, Z. (2010). Targeted conditional gene knockout in human embryonic stem cells. *Cell Research,* **20**: 379-382.

Byrne, A.B., Weirauch, M.T., Wong, V., Kovea, M., Dixon, S.J., Stuart, J.M., Roy, P.J. (2007). A global analysis of genetic interactions in Caenorhabditis elegans. *Journal of Biology,* **6**:

Cardenas, M.E., Cruz, M.C., Del Poeta, M., Chung, N., Perfect, J.R., Heitman, J. (1999). Antifungal activities of antineoplastic agents: Saccharomyces cerevisiae as a model system to study drug action. *Clinical Microbiology Reviews,* **12**: 583-611.

Chan, J.N.Y., Nislow, C., Emili, A. (2009). Recent advances and method development for drug target identification. *Trends in Pharmacological Sciences*, **31:** 82-88.

Chang, M., Bellaoui, M., Boone, C., 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. *Proceedings of the National Academy of Sciences,* **99**: 16934-16939.

Chang, G. (2003). Multidrug resistance ABC transporters. *FEBS Letters,* **555**: 102-105.

Chen, L-m., Liang, Y-J., Ruan, J-W., Ding, Y., Wang, X-W., Shi, Z., Gu, L-Q., Fu, L-W. (2004). Reversal of P-gp mediated multidrug resistance in-vitro and in-vivo by FG020318. *Journal of Pharmacy and Pharmacology,* **56**: 1061-1066.

Coburger, C., Lage, H., Molnar, J., Hilgeroth, A. (2009). Impact of novel MDR modulators on human cancer cells: reversal activities and induction studies. *Pharmaceutical Research,* **26**: 182-188.

Colegate, S.M., & Molyneux, R.J. (2008). Bioactive natural products: detection, isolation and structural determination. *CRC* Press: 11-76

Costanzo et al. (2010). The genetic landscape of a cell. *Science,* **327**: 425-431.

Cullen, L.M, & Arndt, G.M. (2005). Genome-wide screening for gene function using RNAi in mammalian cells. *Immunology & Cell Biology,* **83**: 217-223.

Decottignies, A., Lambert, L., Catty, P., Degand, H., Epping, E.A., Moye-Rowley, W.S., Balzi, E., and Goffeau, A. (1995). Identification and characterization of SNQ2, a new multidrug ATP binding cassette transporter of the yeast plasma membrane. *Journal of Biological Chemistry*, **270**: 150-157.

Dolle, R.E., Le Bourdonnec, B., Morales, G.A., Moriart, K.J., Salvaro, J.M. (2005). Comprehensive survey of combinatorial library synthesis: 2005. *Journal of Combinatorial Chemistry,* **8**: 597-635.

Douglas, C.M., Marrinan, J.A., Kurtz, M.B. (1994). A saccharomyces cerevisiae mutant with echinocandin-resistant 1,3-beta-D-glucan synthase. *Journal of Bacteriology,* **176**: 5686-5696.

Douglas, C.M., D'Ippolito, J.A., Shei, G.J., Meinz, M., Onishi, J., Marrinan, J.A., Li, W., Abruzzo, G.K., Flattery, A., Bartizal, K., Mitchell, A., Kurtz, M.B. (1997). Identification of the FKS1 gene of Candida albicans as the essential target of 1,3-beta-D-glucan synthase inhibitors. *Antimicrobial Agents and Chemotherapy,* **41**: 2471-2479.

Dwight, S.S. (2004). Saccharomyces genome database: underlying principles and organization. *Briefings in Bioinformatics,* **5**: 9-22.

Enserink, J.M., Hombauer, H., Huang, M-R., Kolodner, R.D. (2009). Cdc28/Cdk1 positively and negatively affects genome stability in S. cerevisiae. *Journal of Cell Biology*. **185**: 423-437.

Ericson, E., Gebbia, M., Heisler, L.E., Wildenhain, J., Tyers, M., Giaever, G., Nislow, C. (2008). Off-target effects of psychoactive drugs revealed by genome-wide assays in yeast. *Plos Genetics,* **4**: e100151.

Fang, G., Wang, W., Paunic, V., Oately, B., Haznadar, M., Steinbach, M., Van Ness, B., Myers, C., Kumar, V. (2011). Construction and functional analysis of human genetic interaction networks with genome-wide association data.

Feher, M. & Schmidt, J.M. (2003). Property distributions: differences between drugs, natural products, and molecules from combinatorial chemisty. *J. Chem. Inf. Comput. Sci.*, **43**: 218-277.

Galm, U. & Shen, B. 2007. Natural product drug discovery: the times have never been better. *Chemistry & Biology,* **14**: 1098-1104.

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

Giaever, G., et al. (2002). Functional profiling of the Saccharomyces cerevisiae genome. *Nature*, **418:** 387-91.

Giaever, G., Flaherty, P., Kumm, J., Proctor, M., Nislow, C., Jaramillo, D., Chu, A., Jordan, M., Arkin, A., Davis, R. (2004). Chemogenomic profiling: identifying the functional interactions of small molecules in yeast. *Proceedings of the National Academy of Sciences USA*, **101**: 793-798.

Gietz, R.D., Schiestl, R.H. (2007). Quick and easy yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nature Protocols*, **2**: 35-37.

Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., Louis, E.J., Mewes, H.W., Murakami, Y., Philippsen, P., Tettelin, H., Oliver, S.G. (1996). *Life with 6000 genes. Science*. **274**: 546

Goldstein, A.L., & McCusker, J.H. (1999). Three new dominant drug resistance cassettes for gene disruption in Saccharomyces cerevisiae. *Yeast,* **15**: 1541-1553.

Guldener, U., Heinisch, J., Kohler, G.J., Voss, D., Hegemann, J.H. (2002). A second set of loxP marker cassettes for Cre-mediated multiple gene knockouts in budding yeast. *Nucleic Acid Research,* **30**: e23.

Harvey, A.L. (2008). Natural products in drug discovery. *Drug Discovery Today,* **13**: 894-901.

Hayase, M., Hashitani, H., Suzuki, H., Kohri, K., Brading, A.F. (2007). Evolving mechanisms of action of alverine citrate on phasic smooth muscles. *British Journal of Pharmacology,* **152**: 1228-1238.

Higgins, M.J., & Graham, S.J. (2009). Intellectual property. Balancing motivation and access: Patent challenges tip the scales. *Science,* **326**: 370-371.

Hikkel, I., Lucau-Danila, A., Delaveau, T., Marc, P., Devaux, F., Jacq, C. (2003). A general strategy to uncover transcription factor properties identifies a new regulator of drug resistance in yeast. *The Jornal of Biological Chemistry*, **278**: 11427-11432.

Hillenmeyer, M.E., Fung, E., Wildenhain, J., Pierce, S.E., Hoon, S., Lee, W., Proctor, M., St Onge, R.P., Tyers, M., Koller, D., Altman, R.B., Davis, R.W., Nislow, C., Giaever, G. (2008). The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science,* **320**: 362-365.

Ho, C-H., Magtanong, L., Barker, S., Gresham, D., Nishimura, S., Natarajan, P., Porter, J., Gray, C., Andersen, R., Giaever, G., Nislow, C., Andrews, B., Botstein, D., Graham, T., Yoshida, M., Boone, C. (2009). A molecular barcoded yeast ORF library enables mode-of-action analysis of bioactive compounds. *Nature Biotechnology,* **4**: 369-377.

Ho, C.H., Piotrowski, J., Dixon, S.J., Baryshnikova, A., Costanzo, M., Boone, C. (2011). Combining functional genomics and chemical biology to identify targets of bioactive compounds. *Current Opinion in Chemical Biology*, **15**: 66-78.

Hoon, S., St.Onge, R.P., Giaever, G., Nislow, C. (2008). Yeast chemical genomics and drug discovery: an update. *Trends in Pharmacological Sciences*, **29**: 499-504.

Hopkins, A.L., & Groom, C.R. (2002). The druggable genome. *Nature Reviews Drug Discovery*, **1**: 727-730.

Hu, w., Sillaots, S., Lemieux, S., Davison, J., Kauffman, S., Breton, A., Linteau, A., Xin, C., Bowman, J., Becker, J., Jiang, B., Roemer, T. (2007). Essential gene identification and drug target prioritization in aspergillus fumigatus. *Plos Pathogens,* **3**: 1-15.

Huang, T., Barclay, B.J., Kalman, T.I., von Borstel, R.C., Hastings, P.J. (1992). The phenotype of a dihydrofolate reductase mutant of Saccharomyces cerevisiae. *Gene,* **121**: 167-171.

Hughes, T.R. (2002). Yeast and drug discovery. *Functional & Integrative Genomics,* **2**: 199-211.

Jelinsky, S.A. & Samson, L.D. (1999). Global response of *Saccharomyces cerevisiae* to an alkylating agent. *Proceedings of the National Academy of Sciences.* **96**: 1486-1491.

Jones, T., Federspiel., N.A., Chibana, H., Dungan, J., Kalman, S., Magee, B.B., Newport, G., Thorstenson, Y.R., Agabian, N., Magee, P.T., Davis, R.W., Scherer, S. (2004). The diploid genome sequence of Candida albicans. *PNAS*, **101**: 7329–7334.

Jungwirth, H., & Kuchler, K. (2006). Yeast ABC transporters – a tale of sex, stress, drugs and aging. *FEBS Letters,* **580**: 1131-1138.

Kawasumi, M. & Nghiem, P. (2007). Chemical genetics: elucidating biological systems with small-molecule compounds. *Journal for Investigative Dermatology,* **127**: 1577-1584.

Kerns, E.H. & Di, L. (2008). Drug-like properties: concepts, structure design and methods. *Academic Press.*

Knight, Z.A. & Shokat, K.M. (2007). Chemical genetics: where genetics and pharmacology meet. *Cell,* **128**: 425-430.

Kohen, F.E., & Carter, G.T. (2005). The evolving role of natural products in drug discovery. *Nature Reviews Drug Discovery,* **4**: 206-220.

Kolaczkowski, M., Kolaczowska, A., Luczynski, J., Witek, S., Goffeau, A. (1998). In vivo characterization of the drug resistance profile of the major ABC transporters and other components of the yeast pleiotropic drug resistance network. *Microbial Drug Resistance,* **4**: 143-158.

Kuchler, K., & Schuller, C. (2007). ABC transporters in yeast – drug resistance and stress response in a nutshell. in *Yeast as Tool in Cancer Research* (Springer), 289-314.

Lander, E.S. et al., (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**: 860-921.

Lam, K.S. (2007). New aspects of natural products in drug discovery. *Trends in Microbiology,* **15**: 279-289.

Le Crom, S., Devaux, F., Zhang, Z., Moye-Rowley, W.S., Jacq, C. (2002). New insights into the pleiotropic drug resistance network from genome-wide characterization of the YRR1 transcription factor regulation system. *Molecular and Cellular Biology,* **22**: 2642-2649.

LeDoux, J.E. (2003). Synaptic self: how our brains become who we are. *Penguin.* 1921.

Lee, W., St.Onge, R.P., Proctor, M., Flaherty, P., Jordan, M.I., Arkin, A.P., Davis, R.W., Nislow, C., Giaever, G. (2005). Genome-wide requirements for resistance to functionally distinct DNA-damaging agents. *Plos Genetics*, **1**: 235-246.

Lehner, B., Crombie, C., Tischler, J., Fortunato, A., Fraser, A.G. (2006). Systematic mapping of genetic interactions in Caenorhabditis elegans identifies common modifiers of diverse signaling pathways. *Nature Genetics,* **38**: 896-903.

Li, R. & Murray, A.W. (1991). Feedback control of mitosis in budding yeast. *Cell,* **66**: 519-531.

Li, J.W-H., & Vederas, J.C. (2009). Drug discovery and natural products: end of an era or an endless frontier? *Science,* **325**: 161-165.

Lin, A., Wang, R.T., Ahn, S., Park, C.C., Smith, D.J. (2010). A genome-wide map of human genetic interactions inferred from radiation hybrid genotypes. *Genome Research,* **20**: 1122-1132.

Lipinski, C.A. (2000). Drug-like properties and the causes of poor solubility and poor permeability. *Journal of Pharmacological and Toxicological Methods,* **44**: 235-249.

Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews,* **46**: 3-26.

Liu, L.F., Desai, S.D., Li, T.K., Sun, M., Sim, S.P. (2000). Mechanism of action of camptothecin. *Annals of the New York Academy of Sciences,* **922**: 1-10.

Lomovskaya, O., & Watkins, W. (2001). Inhibition of efflux pumps as a novel approach to combat drug resistance in bacteria. *Journal of Molecular Microbiology and Biotechnology,* **3**: 225-236.

Mallat, S.G., & Zhang, Z. (1993). Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing,* **41**: 3397-3415.

Mani, R., St Onge, R.P., Hartman, J.L., Giaever, G., Roth, F.P. (2008) Defining genetic interaction. *Proceedings of the National Academy of Sciences*, **105**: 3461-3466.

Martelli, C., Alderighi, D., Coronnello, M., Dei, S., Frosini, M., Le Bozec, B., Manetti, D., Neri, A., Romanelli, M.N., Salerno, M., Scapecchi, S., Mini, E., Sgaragli, G., Teodori, E. (2009). N,N-bis(cyclohexanl)amine aryl esters: a new class of highly potent transporter-dependent multidrug resistance inhibitors. *Journal of Medicinal Chemistry,* **52**: 807-817.

Mayer, T.U. (2003) Chemical genetics: tailoring tools for cell biology. *Trends in Cell Biology*, **13**: 270-277.

Michalkova-papajova, D., Obemauerova, M., Subik, J. (2000). Role of the PDR gene network in yeast susceptibility to the antifungal antibiotic mucidin. *Antimicrobial Agents and Chemotherapy,* **44**: 418-420.

Mishra, H. Singh, N., Lahiri, T., Misra, K. (2009). Bioinformation for predicting drug-likeness of small molecules *Bioinformation*. **2063**: 384-388.

Moye-Rowley, W.S. (2003). Transcriptional control of multidrug resistance in the yeast Saccharomyces. *Progress in Nuclear Acid Research and Molecular Biology,* **73**: 251-279.

Mukhopadhyay, K., Kohli, A., Prasad, R. (2002). Susceptibilities of yeast cells are affected by membrane lipid composition. *Antimicrobial Agents and Chemotherapy.***46**: 3695-3705.

Mullen, S., Mani, N., Grossman, T.H. (2004). Inhibition of antibiotic efflux in bacteria by the novel multidrug resistance inhibitors biricodar (VX-710) and timcodar (VX-853). *Antimicrobial Agents and Chemotherapy,* **48**: 4171-4176.

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

Newman, D.J., & Cragg, G.M. (2007). Natural products as sources of new drugs over the last 25 years. *Journal of Natural Products,* **70**: 461-477.

Nierman, W.C., et al. (2005). Genomic sequence of the pathogenic and allergenic filamentous fungus Aspergillus fumigatus. *Nature*, **438**: 1151-1156.

Onyewu, C. & Heitman, J. (2007). Unique applications of novel antifungal drug combinations. *Anti-Infective Agents in Medicinal Chemistry,* **6**: 3-15.

Pannuzio, V.G., Burgos, H.I., Alonso, M., Mattoon, J.R., Ramos, E.H., Stella, C.A. (2004) A simple chemical method for rending yeast permeable to brefeldin A that does not require the presence of an erg6 mutation. *Journal of Biomedicine and Biotechnology,* **3**: 150-155.

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

Parsons, A.B., Lopez, A., Givoni, I.E., Williams, D.E., Gray, C.A., Porter, J., Chua, G., Sopko, R., Brost, R.L., Ho, C.H., Wang, J., Ketela, T., Brenner, C., Brill, J.A., Fernandez, G.E., Lorenz, T.C., Payne, G.S., Ishihara, S., Ohya, Y., Andrews, B., Hughes, T.R., Frey, B.J., Graham, T.R., Andersen, R.J., Boone, C. (2006). Exploring the mode-of-action of bioactive compounds by chemical-genetic profiling in yeast. *Cell*, **126**: 611-625.

Peter, G.J., During, L., Ahmed, A. (2006). Carbon catabolite repression regulates amino acid permeases in Saccharomyces cerevisiae via the TOR signaling pathway. *Journal of Biological Chemistry,* **281**: 5546-5552.

Pierce S., Davis R., Nislow C., Giaever G. (2007). Genome-wide analysis of barcoded Saccharomyces cerevisiae gene-deletion mutants in pooled cultures. Nature Protocols. **2**: 2958-74.

Qi, H., Chen, H.Z., Jin, Z.J. (2002). Caspase 3 gene expression and [Ca2+]i homeostasis underlying desipramine-induced C6 glioma cell apoptosis. *Acta Pharmacologica Sinica,* **23**: 803-807.

Richardson, M.D. (2005) Changing patterns and trends in systemic fungal infections. *Journal of Antimicrobial Chemotherapy,* **56**: Suppl. i5-i11.

Roemer, T., Xu, D., Singh, S.B., Parish, C.A., Harris, G., Wang, H., Davies, J.E., Bills, G.F. (2011). Confronting the challenges of natural product-based antifungal discovery. *Chemistry & Biology,* **18**: 148-164.

Rogers, B., Decottignies, A., Kolaczkowski, M., Carvajal, E., Balzi, E., Goffeau, A. (2001). The pleitropic drug ABC transporters from Saccharomyces cerevisiae. *Journal of Molecular Microbiology and Biotechnology,* **3**: 207-214.

Rooney, J.P., George, A.D., Patil, A., Begley, U., Bessette, E., Zappala, M.R., Huang, X., Conklin, D.S., Cunningham, R.P., Begley, T.J. (2009). Systems based mapping demonstrated that recovery from alkylation damage requires DNA repair, RNA processing, and translation associated networks. *Genomics*, **93**: 42-51.

Rouhi, A.M., & Washington, C.E.N. (2003). Rediscovering natural products. *Chemical & Engineering News,* **81**: 77-78, 82-82, 86, 88-91.

Sambrook, J., Russell, D.W. (2001). Molecular cloning : a laboratory manual, 3rd. edition. *Cold Spring Harbor Laboratory.*

Sams-Dodd, F. (2005). Target-based drug discovery: is something wrong? *Drug Discovery Today*, **10**: 139-147.

Schreiber, S.L. (2000). Target-oriented and diversity-oriented organic synthesis in drug discovery. *Science,* **287**: 1964-1969.

Shogren-Knaak, M.A., Alaimo, P.J., & Shokat, K.M. (2001). Recent advances in chemical approaches to the study of biological systems. *Annual review of cell and developmental biology*, **17**: 405-33.

Sikic, B.I., Fisher, G.A., Lum, B.L., Halsey, J., Beketic-Oreskovic, L., Chen, G. (1997). Modulation and prevention of multidrug resistance by inhibitors of P-glycoprotein. *Cancer Chemotherapy and Pharmacology,* **40**: S13-S19.

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

Skarnes, W.C et al. (2011) A conditional knockout resource for the genome-wide study of mouse gene function. *Nature,* **474**: 337-342.

Smith, A.M., Heisler, L.E., Mellor, J., Kaper, F., Thompson, M., Chee, M., Roth, F., Giaever, G., Nislow, C. (2009). Quantitative phenotypic via deep barcode sequencing. *Genome Research,* **19**: 1836-1842.

Smith, A.M., Ammar, R., Nislow, C., Giaever, G. (2010). A survey of yeast genomic assays for drug and target discovery. *Pharmacology and Therapeutics,* **127**: 156-164.

Smith, A.M., Heisler, L.E., St Onge, R.P. Farias-Hesson, E., Wallace, I.M., Bodeau, J., Harris, A.N., Perry, K.M., Giaever, G., Pourmand, N., Nislow, C. (2010b). Highly-multiplexed barcode sequencing: an efficient method for parallel analysis of pooled samples. *Nucleic Acids Research,* **38**: e142

Smith, A.M., Durbic, T., Oh, J., Urbanus, M., Proctor, M., Heisler, L., Giaever, G., Nislow, C. (2011). Competitive genomic screens of barcoded yeast libraries. *Journal of Visualized Experiments,* **54**.

Spring, D.R. (2004). Chemical genetic to chemical genomics: small molecules offer big insights. *Chemical Society Reviews,* **34**: 472-482.

Sterns, T., Hoyt, M.A., Botstein, D. (1990). Yeast mutants sensitive to antimicrotubule drugs define three genes that affect microtubule function. *Genetics,* **124**: 251-262.

Stockwell, B.R. (2000) Chemical genetics: ligand-based discovery of gene function. *Nature Reviews Genetics,* **1**: 116-125.

Strand, M., Prolla, T.A., Liskay, R.M., Petes, T.D. (1993). Destabilization of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature*, **365**: 274-276.

Symington, L.S. (2002). Role of RAD52 epistasis group genes in homologous recombination and double-strand break repair. *Microbiologu and Molecular Biology Reviews,* **66**: 630-670.

Tan, D. (2005) Diversity-oriented synthesis: exploring the intersections between chemistry and biology. *Nature Chemical Biology*, **1**: 74-84.

Tkacz, J., Cybulska, E., Lampen, J.O. (1971). Specific staining of well mannan in yeast cells with fluorescent-conjugated concanavalin A. *Journal of Bacteriology,* **105**: 1-5.

Tomiki, T., Saito, T., Ueki, M., Konno, H., Asaoka, T., Suzuki, R., Uramoto, M., Hideaki, K., Osada, H. (2006). RIKEN natural products encyclopedia (RIKEN NPEdia), a chemical database of the RIKEN natural products depository (RIKEN NPDepo). *Journal of Computer Aided Chemistry,* **7**: 157-162.

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

Tong, A.H., et al., (2004). Global mapping of the yeast genetic interaction network. *Science,* **303**: 808-813.

Tong, A.H. & Boone, C. (2006) Synthetic genetic array analysis in Saccharomyces cerevisiae. *Methods in Molecular Biology,* **313**: 171-192.

Tu, Zhidong, Wang, Li, Xu, Min, Zhou, Xianghong, Chen, Ting, Sun, Fengzhu. (2006). Further understanding human disease genes by comparing with housekeeping genes and other genes. *BMC Genomics*, **7**: e1-13.

Venter, J.C et al., (2001) The sequence of the human genome. *Science*, **291**: 1304 -1351.

Wang, Z., Ravula, R., Cao, M., Chow, M., Huang, Y. (2010). Transporter-mediated multidrug resistance and its modulation by Chinese medicines and other herbal products. *Current Drug Discovery Technologies,* **7**: 54-66.

Workman, C.T., Mak, H.C., McCuine, S., Tange, J.-B., Agarwal, M., Ozier, O., Begley, T.J., Samson, L.D., Ideker, T. (2006). A systems approach to mapping DNA damage response pathways. *Science,* **312**: 1054-1059.

Yildirim, M.A., Goh, K.-I., Cusick, M.E., Barabasi, A.-L., Vidal, M. (2007). Drug-target network. *Nature Biotechnology,* **25**: 1119-1126.

Yu, L., Lopez, A., Anaflous, A., El Bali, B., Hamal, A., Ericson, E., Heisler, L.E., McQuibban, A., Giaever, G., Nislow, C., Boone, C., Brown, G.W., Bellaoui, M. (2008). Chemical-genetic profiling of imidazo[1,2-a] pyridines and –pyrimidines reveals target pathways conserved between yeast and human cells. *Plos Geneics,* **4**: 1-11.

Zhao, J.H., Tohda, H., Oikawa, A. (1995). Camptothecin-induced sister-chromatic exchange dependent on the presence of bromodeoxyuridine and the phase of the cell cycle. *Mutation Research,* **282**: 49-54.

Zheng, X.F.S. & Chan, T.-F. (2002). Chemical genomics in the global study of protein functions. *Drug discovery today*, **7**: 97-205.

Zimmermann, G.R., Lehár, J., Keith, C.T. (2007). Multi-target therapeutics: when the whole is greater than the sum of the parts. *Drug Discovery Today*, **12**: 34-42.

6.0 Appendices

Appendix A. Diagnostic subset of 491 yeast deletion strains selected

ORF	Common Name	Manually Selected	Computationally Selected
YKL101W	*HSL1*	√	
YHR200W	*RPN10*	√	
YDR150W	*NUM1*	√	
YDR318W	*MCM21*	√	
YJL030W	*MAD2*	√	
YML124C	*TUB3*	√	
YMR307W	*GAS1*	√	
YHR142W	*CHS7*	√	√
YLR319C	*BUD6*	√	√
YLR342W	*FKS1*	√	
YPL047W	*SGF11*	√	
YER155C	*BEM2*	√	
YHR030C	*SLT2*	√	
YIL034C	*CAP2*	√	
YOL067C	*RTG1*	√	
YML121W	*GTR1*	√	
YPR060C	*ARO7*	√	
YDR158W	*HOM2*	√	
YFL031W	*HAC1*	√	
YMR238W	*DFG5*	√	
YJR075W	*HOC1*	√	
YGL227W	*VID30*	√	
YDL065C	*PEX19*	√	
YLR038C	*COX12*	√	
YDR181C	*SAS4*	√	
YKR084C	*HBS1*	√	
YNL001W	*DOM34*	√	√
YNL041C	*COG6*	√	
YOR070C	*GYP1*	√	
YFL013C	*IES1*	√	
YAL002W	*VPS8*	√	
YPL120W	*VPS30*	√	
YOR196C	*LIP5*	√	

YGL174W	*BUD13*	√	√
YGR129W	*SYF2*	√	
YHR034C	*PIH1*	√	
YGR157W	*CHO2*	√	
YGL077C	*HNM1*	√	√
YNR020C	*ATP23*	√	√
YKR082W	*NUP133*	√	√
YCR077C	*PAT1*	√	
YDR378C	*LSM6*	√	
YAL011W	*SWC3*	√	
YPL055C	*LGE1*	√	√
YOL004W	*SIN3*	√	
YOR038C	*HIR2*	√	
YNR010W	*CSE2*	√	
YOR123C	*LEO1*	√	
YML102W	*CAC2*	√	√
YMR078C	*CTF18*	√	
YCL061C	*MRC1*	√	
YOR368W	*RAD17*	√	√
YLR154C	*RNH203*	√	
YGL163C	*RAD54*	√	
YLR320W	*MMS22*	√	
YHR206W	*SKN7*	√	
YDR363W-A	*SEM1*	√	
YGR135W	*PRE9*	√	
YPR119W	*CLB2*	√	
YDR260C	*SWM1*	√	√
YBR082C	*UBC4*	√	
YBR057C	*MUM2*	√	
YMR067C	*UBX4*	√	√
YHR129C	*ARP1*	√	
YJR135C	*MCM22*	√	
YPL018W	*CTF19*	√	
YJL013C	*MAD3*	√	
YPL269W	*KAR9*	√	
YOR349W	*CIN1*	√	
YOR265W	*RBL2*	√	
YBR229C	*ROT2*	√	
YML019W	*OST6*	√	
YBL061C	*SKT5*	√	
YDR310C	*SUM1*	√	
YMR100W	*MUB1*	√	
YCR088W	*ABP1*	√	√

YMR109W	*MYO5*	√	√
YDL106C	*PHO2*	√	√
YBL103C	*RTG3*	√	
YKL176C	*LST4*	√	√
YLR436C	*ECM30*	√	√
YOR002W	*ALG6*	√	
YMR214W	*SCJ1*	√	
YLR292C	*SEC72*	√	
YOR216C	*RUD3*	√	
YMR135C	*GID8*	√	
YLR191W	*PEX13*	√	
YBR164C	*ARL1*	√	
YLR039C	*RIC1*	√	
YDL002C	*NHP10*	√	
YPR179C	*HDA3*	√	
YHR012W	*VPS29*	√	
YGL212W	*VAM7*	√	
YOR106W	*VAM3*	√	
YBR221C	*PDB1*	√	
YHR067W	*HTD2*	√	√
YKL074C	*MUD2*	√	
YPL178W	*CBC2*	√	
YJR050W	*ISY1*	√	√
YHR066W	*SSF1*	√	
YOL041C	*NOP12*	√	
YDR123C	*INO2*	√	
YJR073C	*OPI3*	√	
YLR133W	*CKI1*	√	
YLR393W	*ATP10*	√	
YML081C-A	*ATP18*	√	
YGR081C	*SLX9*	√	√
YLR018C	*POM34*	√	
YNL147W	*LSM7*	√	
YML041C	*VPS71*	√	
YMR075W	*RCO1*	√	
YBR175W	*SWD3*	√	
YGL244W	*RTF1*	√	
YMR263W	*SAP30*	√	
YGL043W	*DST1*	√	
YCR033W	*SNT1*	√	
YKR029C	*SET3*	√	
YBR195C	*MSI1*	√	
YCL016C	*DCC1*	√	

YPR135W	*CTF4*	√	√
YOR144C	*ELG1*	√	√
YPL194W	*DDC1*	√	
YDR386W	*MUS81*	√	
YBR228W	*SLX1*	√	
YML032C	*RAD52*	√	
YMR224C	*MRE11*	√	
YLL002W	*RTT109*	√	
YJL047C	*RTT101*	√	
YML007W	*YAP1*	√	
YIR037W	*HYR1*	√	
YLR021W	*IRC25*	√	
YPL144W	*POC4*	√	
YLR199C	*PBA1*	√	
YLR102C	*APC9*	√	√
YHR152W	*SPO12*	√	
YOR195W	*SLK19*	√	√
YOR014W	*RTS1*	√	√
YFR010W	*UBP6*	√	
YKL213C	*DOA1*	√	
YDL020C	*RPN4*	√	√
YLL049W	*LDB18*	√	
YKR054C	*DYN1*	√	
YDR488C	*PAC11*	√	√
YMR294W	*JNM1*	√	
YMR299C	*DYN3*	√	
YDR254W	*CHL4*	√	
YBR107C	*IML3*	√	
YLR381W	*CTF3*	√	
YPR046W	*MCM16*	√	
YPL017C	*IRC15*	√	
YGL086W	*MAD1*	√	
YPL253C	*VIK1*	√	√
YOR026W	*BUB3*	√	
YCR065W	*HCM1*	√	
YGL216W	*KIP3*	√	√
YLR210W	*CLB4*	√	√
YMR138W	*CIN4*	√	
YPL241C	*CIN2*	√	
YLR200W	*YKE2*	√	
YNL153C	*GIM3*	√	√
YCR082W	*AHC2*	√	
YOR023C	*AHC1*	√	

YGL094C	*PAN2*	√	
YKL025C	*PAN3*	√	
YOR264W	*DSE3*	√	
YAL058W	*CNE1*	√	√
YJL139C	*YUR1*	√	
YGL027C	*CWH41*	√	
YCR017C	*CWH43*	√	
YDR525W	*API2*	√	
YJL062W	*LAS21*	√	
YPR095C	*SYT1*	√	
YLR371W	*ROM2*	√	
YDR351W	*SBE2*	√	
YOR188W	*MSB1*	√	
YKL164C	*PIR1*	√	
YER149C	*PEA2*	√	
YKL127W	*PGM1*	√	
YOL003C	*PFA4*	√	
YOL070C	*NBA1*	√	
YML117W	*NAB6*	√	
YLR443W	*ECM7*	√	
YBR023C	*CHS3*	√	√
YJL099W	*CHS6*	√	√
YLR330W	*CHS5*	√	
YKL079W	*SMY1*	√	√
YLL021W	*SPA2*	√	
YNL079C	*TPM1*	√	
YGR229C	*SMI1*	√	
YDR146C	*SWI5*	√	
YPL158C	*AIM44*	√	
YGL066W	*SGF73*	√	
YER164W	*CHD1*	√	
YMR223W	*UBP8*	√	
YOL068C	*HST1*	√	
YOR279C	*RFM1*	√	
YLR024C	*UBR2*	√	
YJL095W	*BCK1*	√	
YPL089C	*RLM1*	√	
YGR080W	*TWF1*	√	
YIR003W	*AIM21*	√	
YJL020C	*BBC1*	√	√
YBR108W	*AIM3*	√	
YKL007W	*CAP1*	√	
YHR114W	*BZZ1*	√	

YMR092C	*AIP1*	√	
YDL006W	*PTC1*	√	
YDR162C	*NBP2*	√	
YDL106C	*BAS2*	√	
YBR291C	*CTP1*	√	
YMR241W	*YHM2*	√	
YGL252C	*RTG2*	√	
YNL076W	*MKS1*	√	
YLR089C	*ALT1*	√	
YDR128W	*MTC5*	√	
YBR076W	*ECM8*	√	
YKR007W	*MEH1*	√	
YJL036W	*SNX4*	√	
YDL173W	*PAR32*	√	
YDR508C	*GNP1*	√	√
YMR304W	*UBP15*	√	√
YDR127W	*ARO1*	√	
YGL148W	*ARO2*	√	
YBR068C	*BAP2*	√	√
YER052C	*HOM3*	√	
YGR208W	*SER2*	√	
YOR184W	*SER1*	√	
YJR139C	*HOM6*	√	
YGL226C-A	*OST5*	√	√
YHR079C	*IRE1*	√	
YBL082C	*ALG3*	√	
YGR227W	*DIE2*	√	√
YOR067C	*ALG8*	√	
YPL227C	*ALG5*	√	
YNR030W	*ALG12*	√	
YOR085W	*OST3*	√	
YDR349C	*YPS7*	√	
YGL203C	*KEX1*	√	
YDR389W	*SAC7*	√	
YAL053W	*FLC2*	√	
YFL025C	*BST1*	√	
YGL084C	*GUP1*	√	
YLR087C	*CSF1*	√	
YAL023C	*PMT2*	√	
YDL095W	*PMT1*	√	
YBR171W	*SEC66*	√	
YBR162C	*TOS1*	√	
YDL240W	*LRG1*	√	

YKR042W	*UTH1*	√	
YGL228W	*SHE10*	√	
YDR372C	*VPS74*	√	
YDR414C	*ERD1*	√	
YBR015C	*MNN2*	√	
YLR110C	*CCW12*	√	√
YER118C	*SHO1*	√	
YGR014W	*MSB2*	√	
YPR075C	*OPY2*	√	
YNL044W	*YIP3*	√	
YBR105C	*VID24*	√	
YDR255C	*RMD5*	√	
YIL017C	*VID28*	√	
YGL153W	*PEX14*	√	
YGR077C	*PEX8*	√	
YGR133W	*PEX4*	√	
YDR244W	*PEX5*	√	
YDR329C	*PEX3*	√	
YAL055W	*PEX22*	√	
YDR265W	*PEX10*	√	
YJL210W	*PEX2*	√	
YKL197C	*PEX1*	√	
YMR026C	*PEX12*	√	
YMR035W	*IMP2*	√	
YOL044W	*PEX15*	√	
YBR255W	*MTC4*	√	
YHR151C	*MTC6*	√	
YKL098W	*MTC2*	√	
YMR126C	*DLT1*	√	
YER153C	*PET122*	√	
YOR065W	*CYT1*	√	
YPR191W	*QCR2*	√	
YHR116W	*COX23*	√	
YMR256C	*COX7*	√	
YPL172C	*COX10*	√	
YOR350C	*MNE1*	√	
YPR134W	*MSS18*	√	
YOR213C	*SAS5*	√	
YMR127C	*SAS2*	√	
YOR109W	*INP53*	√	√
YPL259C	*APM1*	√	
YDL226C	*GCS1*	√	
YJR033C	*RAV1*	√	

YDR202C	*RAV2*	√	
YBR288C	*APM3*	√	
YGR261C	*APL6*	√	
YPL195W	*APL5*	√	
YJL024C	*APS3*	√	
YGL020C	*GET1*	√	
YDL100C	*GET3*	√	
YOR164C	*GET4*	√	
YOL111C	*GET5*	√	
YOR007C	*SGT2*	√	
YJL004C	*SYS1*	√	
YPL051W	*ARL3*	√	
YGL054C	*ERV14*	√	
YDR108W	*GSG1*	√	
YOL018C	*TLG2*	√	
YNL051W	*COG5*	√	
YGL005C	*COG7*	√	
YML071C	*COG8*	√	
YLR261C	*VPS63*	√	
YLR262C	*YPT6*	√	
YER092W	*IES5*	√	√
YNL021W	*HDA1*	√	
YBR126C	*TPS1*	√	
YGL253W	*HXK2*	√	
YDR345C	*HXT3*	√	
YML048W	*GSF2*	√	
YKL032C	*IXR1*	√	
YML128C	*MSC1*	√	
YLL029W	*FRA1*	√	
YJL101C	*GSH1*	√	
YOL049W	*GSH2*	√	
YKL041W	*VPS24*	√	
YKR035W-A	*DID2*	√	
YML097C	*VPS9*	√	
YOR089C	*VPS21*	√	
YJL154C	*VPS35*	√	
YJL053W	*PEP8*	√	
YOR069W	*VPS5*	√	
YOR068C	*VAM10*	√	
YOR132W	*VPS17*	√	
YLL040C	*VPS13*	√	
YLR360W	*VPS38*	√	
YGR193C	*PDX1*	√	

YER178W	*PDA1*	√	
YNL071W	*LAT1*	√	√
YER061C	*CEM1*	√	
YKL055C	*OAR1*	√	
YJL046W	*AIM22*	√	
YOR196C	*LIP5*	√	
YBR119W	*MUD1*	√	
YHR086W	*NAM8*	√	
YIR005W	*IST3*	√	
YOR308C	*SNU66*	√	√
YPR057W	*BRR1*	√	
YPL157W	*TGS1*	√	
YDR163W	*CWC15*	√	
YBR188C	*NTC20*	√	
YDR482C	*CWC21*	√	√
YNR032C-A	*HUB1*	√	
YPR152C	*URN1*	√	
YDL051W	*LHP1*	√	
YGR276C	*RNH70*	√	
YHR156C	*LIN1*	√	√
YHR157W	*REC104*	√	
YBR278W	*DPB3*	√	
YOL080C	*REX4*	√	√
YKR092C	*SRP40*	√	√
YLR221C	*RSA3*	√	
YGL231C	*EMC4*	√	
YCL045C	*EMC1*	√	
YKL207W	*LRC3*	√	
YLL014W	*EMC6*	√	
YJR088C	*EMC2*	√	
YIL090W	*ICE2*	√	
YBR283C	*SSH1*	√	
YOR311C	*DGK1*	√	
YGR202C	*PCT1*	√	
YKL020C	*SPT23*	√	√
YDR512C	*EMI1*	√	
YMR282C	*AEP2*	√	
YDR192C	*NUP42*	√	
YKL068W	*NUP100*	√	
YDR101C	*ARX1*	√	
YAL059W	*ECM1*	√	
YAR002W	*NUP60*	√	
YDL088C	*ASM4*	√	√

YML103C	*NUP188*	√	√
YJR074W	*MOG1*	√	
YLR335W	*NUP2*	√	
YMR129W	*POM152*	√	
YMR153W	*NUP53*	√	
YCR079W	*PTC6*	√	
YJL124C	*LSM1*	√	
YGL014W	*PUF4*	√	√
YHR167W	*THP2*	√	√
YBR231C	*SWC5*	√	
YDR334W	*SWR1*	√	
YDR485C	*VPS72*	√	
YLR085C	*ARP6*	√	
YJL168C	*SET2*	√	√
YPR023C	*EAF3*	√	√
YAR003W	*SWD1*	√	
YDR469W	*SDC1*	√	
YLR015W	*BRE2*	√	
YDL074C	*BRE1*	√	
YAL013W	*DEP1*	√	
YIL084C	*SDS3*	√	
YBR095C	*RXT2*	√	√
YNL097C	*PHO23*	√	
YPL181W	*CTI6*	√	
YPL139C	*UME1*	√	
YBL008W	*HIR1*	√	√
YJR140C	*HIR3*	√	
YGL151W	*NUT1*	√	
YPR070W	*MED1*	√	
YBR103W	*SIF2*	√	
YGL194C	*HOS2*	√	
YKR028W	*SAP190*	√	
YLR418C	*CDC73*	√	
YER161C	*SPT2*	√	
YKL160W	*ELF1*	√	
YDR225W	*HTA1*	√	
YBR010W	*HHT1*	√	
YPR018W	*RLF2*	√	
YHR191C	*CTF8*	√	√
YPL008W	*CHL1*	√	
YMR048W	*CSM3*	√	
YJR043C	*POL32*	√	√
YDR217C	*RAD9*	√	

YER173W	*RAD24*	√	
YOR025W	*HST3*	√	
YGL175C	*SAE2*	√	
YBR098W	*MMS4*	√	
YLR135W	*SLX4*	√	
YJL092W	*SRS2*	√	
YDR363W	*ESC2*	√	
YMR190C	*SGS1*	√	√
YHR031C	*RRM3*	√	
YHR154W	*RTT107*	√	
YDR279W	*RNH202*	√	
YNL072W	*RNH201*	√	
YKL113C	*RAD27*	√	
YML028W	*TSA1*	√	
YLR032W	*RAD5*	√	√
YDR004W	*RAD57*	√	
YDR076W	*RAD55*	√	
YER095W	*RAD51*	√	
YDR369C	*XRS2*	√	
YJL115W	*ASF1*	√	
YPR164W	*MMS1*	√	
YBR216C	*YBP1*	√	
YOR208W	*PTP2*	√	√
YNL053W	*MSG5*		√
YOR127W	*RGA1*		√
YBR118W	*TEF2*		√
YGR034W	*RPL26B*		√
YFL001W	*DEG1*		√
YGL078C	*DBP3*		√
YDR083W	*RRP8*		√
YPL003W	*ULA1*		√
YDR139C	*RUB1*		√
YLL046C	*RNP1*		√
YMR276W	*DSK2*		√
YIL140W	*AXL2*		√
YBR260C	*RGD1*		√
YMR036C	*MIH1*		√
YMR273C	*ZDS1*		√
YIL095W	*PRK1*		√
YJR066W	*TOR1*		√
YDR159W	*SAC3*		√
YIL040W	*APQ12*		√
YNL141W	*AAH1*		√

YPL180W	*TCO89*		√
YDR253C	*MET32*		√
YDL066W	*IDP1*		√
YML038C	*YMD8*		√
YJL134W	*LCB3*		√
YPR079W	*MRL1*		√
YJR060W	*CBF1*		√
YJR125C	*ENT3*		√
YOR327C	*SNC2*		√
YMR123W	*PKR1*		√
YDL056W	*MBP1*		√
YGR092W	*DBF2*		√
YPR141C	*KAR3*		√
YMR292W	*GOT1*		√
YDR517W	*GRH1*		√
YGR089W	*NNF2*		√
YML012W	*ERV25*		√
YNR051C	*BRE5*		√
YOR115C	*TRS33*		√
YMR274C	*RCE1*		√
YJL092W	*SRS2*		√
YMR055C	*BUB2*		√
YLR254C	*NDL1*		√
YIR025W	*MND2*		√
YDL070W	*BDF2*		√
YHR006W	*STP2*		√
YOL020W	*TAT2*		√
YKR099W	*BAS1*		√

Appendix B. Biological function of compounds listed in **Table 4**.

Compound	Function
Benomyl	Microtubule -depolymerizing agent
Camptothecin	DNA damaging agent: inhibits DNA enzyme topoisomerase I
Tamoxifen	Estrogen receptor antagonist
Micafungin	Inhibits β-1,3-glucan synthesis – an essential component of fungal cell walls
Mitomycin C	DNA damaging agent: crosslinks DNA
Nocodazole	Prevents microtubule polymerization
Hydroxyurea	DNA damaging agent; inhibits ribonuclease reductase
Desipramine	Inhibits synaptic reuptake of norepinephrine and serotonin
Caspofungin	Inhibits β-1,3-glucan synthesis – an essential component of fungal cell walls
Extract 95-97	NP: unknown MOA
Oligomycin	Inhibits ATP synthase by blocking its proton channel
Menthol	Terpene alcohol used as a flavorant
Mebendazole	Antiparasitic drug: Thought to inhibit microtubule synthesis
MMS	DNA damaging agent: methylates DNA
Wikostatin	N-WASP inhibitor
Alverine citrate	Smooth muscle relaxant
Nystatin	Binds ergosterol (a major component of the cell membrane) to cause toxic pores
Cantharidin	Induces cell cycle arrest in G2/M phase, apoptosis, and DNA damage
Methotrexate	DNA damaging agent: inhibits dihydrofolate reductase

Appendix C. Top 50 interacting genes interacting with benomyl or micafungin

Top 50 genes found to interact with benomyl

Gene	z-score
CIN1	-7.89909
TUB3	-5.38247
CIN4	-3.26071
HCM1	-1.85338
RBL2	-1.45043
SWR1	-1.25259
VPS72	-1.23563
BEM2	-1.235
ALG6	-1.16714
RIC1	-1.07483
ROM2	-1.07129
SET2	-0.991224
POC4	-0.962186
RAD54	-0.877599
MUD1	-0.853493
ARP1	-0.83057
SYS1	-0.830161
PAN2	-0.817209
ISY1	-0.802578
PAT1	-0.77695
DEP1	-0.77616
IDP1	-0.768966
RLM1	-0.758507
ARP6	-0.733688
HHT1	-0.711021
GNP1	-0.708673
SAC3	-0.69212
INO2	-0.679163
VAM10	-0.658327
RGD1	-0.64897
BRE2	-0.641851
SEC66	-0.630684
RTG1	-0.609855
RNH201	-0.598024
SWC5	-0.577148
YAP1	-0.563866
MAD2	-0.560774
IXR1	-0.560239
SNT1	-0.560132
NUP100	-0.556307

POM152	-0.551477
IES5	-0.533916
UBP8	-0.533267
CAC2	-0.523036
UBX4	-0.520465
NUP42	-0.517691
MKS1	-0.513716
CTI6	-0.512293
MMS4	-0.510808
CTF19	-0.496272

Top 50 genes found to interact with micafungin

Gene	z-score
BCK1	-5.3555375
ROM2	-3.45649
SMI1	-2.747405
ALG8	-2.57822
SKT5	-1.5669275
BEM2	-1.8994925
ROT2	-1.04351625
AIM44	-1.415302
INO2	-1.30704375
PAT1	-0.75070875
PFA4	-1.7681225
VPS21	-0.57113675
VPS72	-0.734783075
RLM1	-1.125246
COG7	-0.5599325
ECM1	-0.8371625
SWD3	-1.2520075
ASF1	-0.22414325
SER2	-0.36915
RNH203	-0.41623915
LGE1	-1.4834425
MRL1	-0.96613625
TOR1	-0.39588075
PEX2	-0.80091375
YHM2	-0.318591725
PIR1	-0.5751209
SYS1	-0.1012485
IXR1	-0.66089075
LSM1	-0.3488
YPS7	-1.2184555
RAD17	-0.098794
TRS33	-0.81937

NUP100	0.214489
SWC5	-0.51063125
RAD5	-0.5248175
NUP188	-0.547783075
VPS35	-1.08139825
MAD1	-0.5250505
RRP8	-0.38280525
RAD52	-0.801045
PEA2	-0.4026155
CTF8	-0.51788625
RMD5	-0.54324825
SWI5	-0.609976675
GNP1	0.45403875
STP2	-0.848168
CKI1	0.5860785
RAD27	-0.702155518
VPS17	0.11202125
CIN4	-0.315549608