



Published in final edited form as:

*Chem Biol.* 2011 October 28; 18(10): 1273–1283. doi:10.1016/j.chembiol.2011.07.018.

## Compound prioritization methods increase rates of chemical probe discovery in model organisms

Iain M Wallace<sup>\*1,2,3</sup>, Malene L Urbanus<sup>\*1,2,3</sup>, Genna M Luciani<sup>1,2</sup>, Andrew R Burns<sup>1,2</sup>, Mitchell KL Han<sup>4</sup>, Hao Wang<sup>6</sup>, Kriti Arora<sup>1,2</sup>, Lawrence E Heisler<sup>2,4</sup>, Michael Proctor<sup>5</sup>, Robert P St. Onge<sup>5</sup>, Terry Roemer<sup>6</sup>, Peter J Roy<sup>1,2</sup>, Carolyn L Cummins<sup>4</sup>, Gary D Bader<sup>1,2,3</sup>, Corey Nislow<sup>1,2,3</sup>, and Guri Giaever<sup>#1,2,4</sup>

<sup>1</sup>Department of Molecular Genetics, University of Toronto, 1 King's College Circle, Toronto, Ontario M5S 1A8, Canada

<sup>2</sup>Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, 160 College Street, Toronto, Ontario M5S 3E1, Canada

<sup>3</sup>Banting and Best Department of Medical Research, University of Toronto, 112 College Street, Toronto, Ontario M5G 1L6, Canada

<sup>4</sup>Department of Pharmaceutical Sciences, University of Toronto, 144 College Street, Toronto, Ontario M5S 3M2, Canada

<sup>5</sup>Stanford Genome Technology Center, 855 California Avenue, Palo Alto, California 94304, USA

<sup>6</sup>Center of Fungal Genetics, Merck-Frosst Canada Ltd., Montreal, Quebec H9H 3L1, Canada

### SUMMARY

Pre-selection of compounds that are more likely to induce a phenotype can increase the efficiency and reduce the costs for model organism screening. To identify such molecules, we screened ~81,000 compounds in *S. cerevisiae* and identified ~7,500 that inhibit cell growth. Screening these growth-inhibitory molecules across a diverse panel of model organisms resulted in an increased phenotypic hit-rate. This data was used to build a model to predict compounds that inhibit yeast growth. Empirical and *in silico* application of the model enriched the discovery of bioactive compounds in diverse model organisms. To demonstrate the potential of these molecules as lead chemical probes we used chemogenomic profiling in yeast and identified specific inhibitors of lanosterol synthase and of stearoyl-CoA 9-desaturase. As community resources, the ~7,500 growth-inhibitory molecules has been made commercially available and the computational model and filter used are provided.

---

© 2011 Elsevier Ltd. All rights reserved.

<sup>#</sup>Corresponding author Correspondence: guri.giaever@utoronto.ca phone: 416-978-7182, fax:416-978-8287 .

<sup>\*</sup>IM Wallace and ML Urbanus contributed equally to this work

**Website** All of the data used in this study is available at <http://chemogenomics.med.utoronto.ca/supplemental/bioactive/>

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## INTRODUCTION

Current methods for identifying lead chemical probes frequently rely on high-throughput screening against select targets of interest. This approach assumes that the *in vitro* high potency of small molecules will translate to low-dose efficacy *in vivo*. However, this is often not the case (Gleeson, et al., 2011). In contrast, *in vivo* model organism screening provides a direct measure of cellular potency, bypassing the bias of target pre-selection typically used in modern drug discovery. A growing number of academic labs are pursuing model organism screens to identify chemical probes for use as powerful molecular tools to probe biological function (Frearson and Collie, 2009). Chemical probes complement standard genetic approaches to elucidate gene function while offering distinct advantages. For example, when applied to a cell or whole organism, the effects induced by chemical probes are often rapid, reversible and tuneable (Morgan, et al., 2008; Oprea, et al., 2009; Workman and Collins, 2010). Moreover, chemical probes can often be transferred across model organisms, regardless of their genetic tractability (Specht and Shokat, 2002). One drawback of chemical and chemical-genetic screens is that the percentage of compounds that results in a desired phenotype is often small; for example, in a *C. elegans* study, only 2% of pharmacologically active compounds resulted in a phenotype (screened at 25 $\mu$ M) (Kwok, et al., 2006) and in a study using a hyperpermeable *E. coli* strain, only 3.5% of compounds (screened at 50 $\mu$ M) resulted in growth inhibition (Li, et al., 2004). These observations, combined with the fact that model organism screening can be both compound-intensive and time-consuming (Burns, et al., 2010; Wheeler and Brändli, 2009) places an emphasis on compound selection prior to screening in contrast to typical *in vitro* high-throughput screening campaigns (Agresti, et al., 2010; Lipinski and Hopkins, 2004) where the number of total hits is higher and compound consumption is lower. Such pre-screening compound selection strategies may include enriching for known active substructures against multiple targets (“privileged structures”) (Klekota and Roth, 2008) and/or enriching for compounds most likely to accumulate in the organism of interest (Burns, et al., 2010).

The pre-selection strategy described here is aimed at increasing the discovery rate of lead chemical probes in model organisms by first identifying small molecules that inhibit yeast growth. Growth is a comprehensive phenotype, combining multiple effects on cellular physiology into a single quantitative metric (Botstein and Fink, 1988). Moreover, growth measurements can be made in a rapid, high-throughput and low cost manner (Paixao, et al., 2008; Proctor, et al., 2011). Here, we first screened ~81,000 commercially available synthetic compounds and identified ~7,500 compounds that inhibit growth of *S. cerevisiae*. It is noteworthy that yeast screens often require significantly higher doses (approximately 5-10x) compared to typical mammalian cell culture screens or *in vitro* assays (Figure S1 and (Blackburn and Avery, 2003; Buurman, et al., 2005; Ericson, et al., 2008; Kwak, et al., 2011)). While our initial yeast screening concentrations are relatively high (maximum 200 $\mu$ M), this high dose does not sacrifice specificity (Blackburn and Avery, 2003; Botet, et al., 2007; Dias, et al., 2010; Dorer, et al., 2005; Ericson, et al., 2008; Giaever, et al., 2004; Khozoie, et al., 2009; Kwak, et al., 2011; Muren, et al., 2001). Several biological factors also contribute to yeast’s ability to resist chemical perturbation, including the physical barrier of the yeast cell wall (Dielbandhosing, et al., 1998) and a dynamic defense known

as the pleiotropic drug response (PDR). The PDR is comprised of efflux pumps that reduce the intracellular dose of a broad spectrum of diverse small molecules (Ernst, et al., 2010; Kolaczkowski, et al., 1998; Rogers, et al., 2001).

Once we had identified the ~7,500 yeast bioactives or “yactives”, we then tested the set on a diverse set of model organisms for bioactivity. We found that the yactives significantly increased phenotypic hit-rates compared to randomly selected compounds. Using the physicochemical properties of the yactives, we designed a two-property compound filter based on a simple modification of the Lipinski’s rule-of-five (Lipinski, et al., 1997) and in addition, built a Naïve Bayes model to identify substructures present in yactives. We demonstrate both empirically and *in silico* (using publicly available datasets) that application of the two-property filter and the Naïve Bayes model result in an enrichment for phenotype-inducing compounds in diverse model organisms. Finally, we address the question of whether growth inhibitory compounds have the potential to become specific chemical probe leads by testing twenty of the most potent growth-inhibitory compounds *in vivo* against all ~1100 essential yeast proteins using our well-validated HaploInsufficiency Profiling (HIP) assay (Baetz, et al., 2004; Giaever, et al., 2004; Giaever, et al., 1999; Lain, et al., 2008; Lum, et al., 2004; Xu, et al., 2007). Several of these compounds exhibit specific genome-wide profiles, identifying candidates for the most likely protein target(s). We pursued two of the most promising target candidates; one supporting lanosterol synthase (Erg7 in yeast, mammalian homolog LSS) as the primary target, and the second fatty acid desaturase (Ole1 in yeast, mammalian homolog SCD) as the potential target. We confirmed these two targets genetically and in independent secondary assays. Taken together, our results demonstrate that pre-selection and prioritization of compound libraries increase the likelihood of identifying specific chemical probe leads for model organisms while decreasing overall costs. To disseminate these tools, the yactives have been made available through Chembridge, Inc. and we provide a prioritized list of compounds generated by applying our model to all commercially available small molecules (Irwin and Shoichet, 2005) on our supplementary website (<http://chemogenomics.med.utoronto.ca/supplemental/bioactive/>).

## RESULTS

### Small molecules that inhibit yeast growth increase the phenotypic hit-rate in other model organisms

To identify small molecules that decrease yeast fitness or growth, we screened 81,320 commercially available synthetic compounds (Table S1) and identified 7,476 small molecules that inhibit wild-type *S. cerevisiae* growth by at least 30% (IC30)(Experimental Procedures). We next asked if this set of yeast bioactive compounds, or yactives, were enriched for molecules that induce a phenotype when tested across a diverse set of model organisms, spanning substantial evolutionary distance (Figure 1A). Subsets of the 7,476 yactives were screened against our panel of model organisms (as well as a human cell line) and the results compared to those obtained from screening random compounds (Figure 1B, Table S2). Yactives significantly enriched for compounds that inhibited growth (IC50 or greater) in human A549 non-small cell lung carcinoma cells, *S. pombe*, *C. neoformans*, *E. coli*, *B. subtilis*, and *C. albicans*. The increase in phenotypic hit-rate was independent of

evolutionary distance. Notably, in the model metazoan *C. elegans* (where hit-rate was determined by visual inspection) the yactives increased the discovery rate 6.6x over random compounds (Figure 1B).

### Physicochemical properties of growth-inhibitory compounds

Because the majority of the 81,320 synthetic compounds screened adhere to Lipinski's rule-of-five (intended to define chemical properties that reflect oral bioavailability (Lipinski et al. 1997)) we asked if a simple modification of these rules could be used as a yactive filter. Two of four physicochemical properties that comprise Lipinski's rule-of-five were significantly different ( $p < 1 \times 10^{-15}$ ) in yactive compounds versus inactive compounds (Table S3). First, Lipinski's rule states compounds should have a calculated octanol-water partition coefficient  $\text{LogP} \leq 5$ . In contrast, yactive compounds are more lipophilic (mean  $\text{LogP} = 4.0$ ) than inactive compounds (mean  $\text{LogP} = 3.1$ ) (Figure 2A). This observation suggests that they are more likely to be passively transported into the cell as their solubility in a lipid rich-environment would be expected to contribute to cell permeability (Al-Awqati, 1999; Gamo, et al., 2010; Hacker, et al., 2009). Second, Lipinski's rule includes compounds that have  $\leq 10$  hydrogen acceptors, while yactives are best described using a limit of  $\leq 6$  hydrogen acceptors (Figure 2B). This decreased number of hydrogen acceptors also reflects the likelihood that such compounds can be passively transported across the cell membrane (Muegge, 2003). The increased hit-rate achieved by applying a 2-property filter based on these observations (compounds pass if they have a  $\text{LogP} \leq 2$  and hydrogen acceptors  $\leq 6$ ) (12.7% compared to 9.2%, p-value of  $10^{-15}$ , Table S2) pre-purchase, would have reduced the number of compounds screened from 81,320 to 53,480 (a 30% cost savings) while still identifying 91% of the original 7,475 yactives, demonstrating that even such a modest increase in hit-rate can result in substantial cost savings.

### Application of a Naïve Bayes model allows prediction of yactives

Encouraged by the increased hit-rate resulting from our two-property modification of Lipinski's rule-of-five, we built a Naïve Bayes model to better enable prediction of yactives. Specifically, in this model, substructures in active compounds are weighted higher than those found in inactive compounds, resulting in a prioritized list of compounds for screening. The Naïve Bayes model was built using the data from our original 81,320 compounds as a training set. ECFP\_4 topological fingerprints (Rogers, 2005) was selected to represent substructures because it outperformed three other representational methods (Figure 3A). Five-fold cross validation was used, with 4/5ths of the original screening data used as the training set, and the remaining 1/5 used to test the model's performance. This procedure was repeated five times, and the model's performance reported as the average over the iterations. At a cutoff of the top 10% of ranked compounds, the ECFP\_4 model resulted in an enrichment factor of  $\sim 4.5$ , defined as  $\sim 4.5$ -fold the number of yactives compared to a random set of compounds (Figure 3A).

To address potential overestimation bias from using the same library for model building and testing, we assessed the performance of the model on an independent chemical library. Because this library has different structural property distributions than the training set, it better represents real world performance. To generate this dataset, the Spectrum Library of

~2,000 compounds was screened against *S. cerevisiae* to identify yactives followed by application of the Naïve Bayes model. The top 10% of compounds ranked by the model showed an enrichment factor of ~3.5 for the yactives (Figure 3B). Extending these tests, we applied our model to publically available chemical screening data. Specifically, the model was applied to the data from 29 yeast assays available from PubChem (Wang, et al., 2010); 23 of the assays were designed to identify modulators of yeast growth (in mutant backgrounds or in wild-type strains), and six relied on readouts other than growth inhibition (see Table S4 for assay details, Table S5 for 2-property filter results, Table S6 for Naïve Bayes model results). Our model performed well across nearly all of these diverse assays, achieving a median enrichment factor 1.85 in the top 10% of compounds ranked (Figure 3C).

### Predictive approaches enrich for phenotype-inducing compounds in diverse organisms

We next asked how well the Naïve Bayes model and the 2-property filter enriched for phenotype-inducing compounds when applied to our data for other model organisms (*S. pombe*, *B. subtilis*, *E. coli*, and *C. elegans*). Comparison of the performance of three approaches (yactives/Naïve Bayes/2-property filter) revealed that empirical screening of the yactives gave the best performance (median enrichment factor 5.95), the Naïve Bayes model performed nearly as well (median enrichment factor 4.30), while the 2-property filter performed appreciably lower (median enrichment factor 1.64) (Figure 4A, see also Tables S2 and S7). To avoid overestimating the level of performance of the Naïve Bayes model due to the model being tested on the same library as the training set, we tested the performance using results from nine publically available small molecule screens performed in four organisms (*E. coli*, *C. elegans*, *C. reinhardtii*, and *D. rerio*) from PubChem (Wang, et al., 2010). As was the case in the yeast assays, the Naïve Bayes model performed best (median enrichment factor 2.10), while the 2-property filter exhibited only modest improvement (enrichment factor 1.28) (Figure 4B, see also Tables S8 and S9 for individual results). The increase in enrichment factors observed across such diverse model organisms (see also Figure 4A and Figure 4B) demonstrates that these approaches are broadly generalizable across a very wide range of model organisms and are therefore valuable methods for compound selection and prioritization.

### Yactives are a rich source of lead probes

To be useful as a chemical probe, a compound should act in a specific manner to inhibit a protein or cellular activity. We therefore tested the twenty most potent yactives using our well-validated genome-wide HaploInsufficiency Profiling (HIP) assay (Baetz, et al., 2004; Giaever, et al., 2004; Giaever, et al., 1999; Lain, et al., 2008; Lum, et al., 2004; Xu, et al., 2007) to identify candidate protein targets. The HIP assay allows an unbiased, *in vivo* quantitative measure of the relative drug sensitivity of all ~1100 essential yeast proteins in a single assay, and results in a list of candidate protein targets ranked in order of compound sensitivity. The profiles of the 20 yactives revealed that 13 of the 20 tested exhibited a degree of specificity for an essential protein or protein(s) in the HIP profile while the remaining 7 compounds did not (Figure S2). We chose the two compounds that exhibited the highest degree of specificity for detailed follow-up studies. Our data suggest these compounds target Erg7, lanosterol synthase, and Ole1, fatty acid desaturase, respectively.

The HIP profile of Chembridge 95809153 (ERG7.153, Figure 5A) supports Erg7 as the most likely target. *ERG7* encodes lanosterol synthase, an essential protein involved in ergosterol biosynthesis (Lees, et al., 1995), a pathway exhibiting structural and functional conservation with the biosynthesis of cholesterol in human. Erg7 performs an essential step in ergosterol biosynthesis and holds promise as an antifungal target based on the success of antifungal agents that target other steps of this pathway (Jolidon, et al., 1990; Voyron, et al., 2010). In addition, the human homolog of Erg7 (LSS, lanosterol synthase BLASTP e-value 5e-148) has potential therapeutic relevance as a cholesterol-lowering agent (Charlton-Menys and Durrington, 2007). Two supporting studies demonstrated that compounds sharing structural similarity to ERG7.153 inhibit lanosterol synthase (Figure S3A). One of these compounds was demonstrated to inhibit lanosterol synthase (Erg7) in *C. albicans* (Buurman, et al., 2005), while the other was shown to inhibit the human lanosterol synthase, LSS (Fouchet, et al., 2008) (Figure S3A). To genetically confirm that ERG7.153 inhibits Erg7, we tested the individual *erg7* heterozygous deletion for the expected compound hypersensitivity to the wild type (Figure S3B). Because ERG7.153 was not available for resupply, we carried out further testing with a close analog, CB 83425298 (ERG7.298), which induced similar hypersensitivity in the *S. cerevisiae erg7* heterozygous deletion strain (Figure 5B). Analogous growth assays of an *erg7* heterozygous deletion mutant and a conditional promoter shut-off allele in the human fungal pathogen *C. albicans* also exhibited hypersensitivity to compound, providing several lines of gene-dose support for Erg7 as the drug target of ERG7.153 and ERG7.298 (Figure 5B). Two additional heterozygous deletion strains, *neol* and *pik1*, encoding a putative aminophospholipid translocase (flippase) (Paulusma and Oude Elferink, 2005) and a phosphatidylinositol 4-kinase (Flanagan, et al., 1993), respectively, are also sensitive to ERG7.153. Both of these genes have been previously classified as multi-drug resistant (MDR) (Hillenmeyer, et al., 2008). In this study, *neol* heterozygous deletion strain was sensitive in 14 out of 20 profiles (70%) in this study and the *pik1* heterozygous deletion strain sensitive to 7 of 20 (35%) compounds tested.

To independently test if Erg7 is the target of ERG7.298, we analyzed the lipid metabolites from cells grown in the presence of this inhibitor (Figure 5C) by mass spectrometry. As predicted for a *bona fide* Erg7 inhibitor, the substrate of Erg7 (oxidosqualene) showed significant accumulation in the presence of inhibitor compared to vehicle alone. As a positive control, the level of oxidosqualene was measured in cells treated with cerivastatin which inhibits HMG-CoA reductase, the rate-limiting step of the ergosterol and cholesterol biosynthetic pathways (Endo, 1988). As expected, cells treated with cerivastatin did not accumulate oxidosqualene. Relative measurements of ergosterol, the end product of the ergosterol biosynthetic pathway, showed depletion in cells treated with the Erg7 inhibitor and with cerivastatin. Finally, we were able to partially rescue the growth defect caused by cerivastatin and ERG7.298 by adding ergosterol to the growth medium. Although yeast do not typically incorporate exogenously supplied lipids, we used a *S. cerevisiae* strain carrying the *upc2-1* mutation (Li and Prinz, 2004) (Figure 5D, Figure S3C) which allows cells to take up exogenously supplied sterols under aerobic growth conditions (Crowley, et al., 1998).

A second HIP profile supports Ole1 as a likely protein target (Figure 6A) of a novel compound. Ole1 encodes the yeast delta(9) fatty acid desaturase, which converts stearic acid



to oleic acid and which has previously been proposed as a potential antifungal target (Hu, et al., 2007). Furthermore, the human homolog of Ole1, SCD (stearoyl-CoA desaturase, BLASTP e-value 3e-52), has attracted interest for its potential modulation for the treatment of diabetes (Lenhard, 2011; Ntambi, et al., 2002). Two other heterozygous deletion strains unrelated to fatty acid desaturase inhibition, (*lsg1* and *rpb8*) are sensitive to this compound. Given that both of these genes encode ribosomal components which frequently come up as sensitive in diverse chemical screens (unpublished data), they were not further pursued in this study. We used a *S. cerevisiae ole1* DAmP loss-of-function allele (Schuldiner, et al., 2005; Yan, et al., 2008) to confirm compound hypersensitivity. Hypersensitivity was also seen with an *ole1* conditional promoter shut-off allele in *C. albicans*, further supporting Ole1 as the target of OLE1.041(CB 11119041)(Xu, et al., 2009) (Figure 6B). This compound was also effective *in vitro*, inhibiting the enzymatic activity of Ole1 in *S. cerevisiae*, *C. albicans* and human HepG2 cells (Figure 6C). Finally, we found that two *C. elegans* mutants with reduced stearoyl-CoA desaturase activity (*fat-5;fat-7* and *fat-5;fat-6*) (Brock, et al., 2006) are hypersensitive to OLE1.041 (Figure 6D). These two examples highlight the ability of the HIP assay to identify additional effects that can be monitored during compound optimization.

## DISCUSSION

Compounds that inhibit yeast growth are more likely to induce phenotypes in other model organisms. Modeling the properties of the subset of drug-like compounds that inhibit yeast growth allows prioritization of compounds for model organism screening, reducing screening costs and increasing efficiency. Over time, as the research community accumulates compound screening data, these models can be refined to be both organism and phenotype-specific resulting in increasing the predictability and accuracy. As an important first step, we have demonstrated that compounds that inhibit yeast growth are more likely to induce phenotypes of interest in other model organisms and in mammalian cell culture assays. In order to address whether the inhibitory compounds we identified could act in a specific manner, we followed up on two compounds that looked particularly promising based on their genome-wide profile of drug sensitivity. The mechanism of action of these compounds (ERG7.298, OLE1.041) was shown to be consistent with inhibition of their presumed molecular targets in both genetic and cellular biochemical assays. Both *ERG7* and *OLE1* are highly conserved with human *LSS* and *SCD*, and represent recognized targets of medical relevance. *ERG7* and *LSS* inhibitors have clinical relevance as potential antifungal and anti-cholesterol lowering agents, respectively, while the human homolog of *OLE1*, *SCD*, may represent a potential target for diabetes treatment.

The identification of two compounds that act with a high degree of specificity in a relatively short experimental time frame underscores the benefits of prioritizing compounds. While no filtering or prioritization method can trump an exhaustive screening campaign and perfectly predict all compounds of interest, our results clearly indicate that pre-selection methods, when applied across diverse assays and organisms, can identify and prioritize those compounds most likely to induce a phenotype. The advantage of using such compounds as starting points for chemical probe discovery in model organisms is that a wide variety of genetic tools in different organisms can be used to validate the mode of action, as well as to

identify off-target effects. To provide a publically accessible resource, we applied our 2-property filter and our Naïve Bayes model to compounds available in 1) the NIH Molecular Screening program (Austin, et al., 2004) and 2) the Zinc catalogue of approximately 14 million purchasable compounds (Irwin and Shoichet, 2005). These results are available on our website, <http://chemogenomics.med.utoronto.ca/supplemental/bioactive/>.

Finally, a primary goal of this work was to encourage compound suppliers to provide libraries directed at model organism screening to the research community. Towards this end, ChemBridge, Inc. has agreed to make our ~7,500 yactive compounds available for purchase as a pre-plated compound set. This library should prove a valuable resource for chemical screening labs working to develop chemical probes using model organisms.

## SIGNIFICANCE

We have presented three approaches, based on yeast growth inhibition, to guide compound selection to reduce the costs associated with model organism screening programs. First, we have demonstrated that compounds that inhibit yeast growth enriches for compounds that induce a variety of phenotypes in diverse model organisms, and these compounds, with further optimization, may yield specific chemical probes. The first approach is then to simply screen compounds for those that inhibit yeast growth. A second approach is to prioritize compounds based on those that pass the two-property filter described here. This approach, depending on the model organism, can decrease costs by ~25%, and is straightforward to implement. The third approach is to purchase compounds based on their likelihood to result in a desired phenotype by applying our Naïve Bayes model. This approach can also dramatically reduce costs. Newly generated screening data can be used to rebuild the model described here in the context of the model organism of interest to increase performance. This iterative approach is key as no model will perform optimally in all applications. Finally, as an experimental resource the yactive compounds are available as a pre-plated collection from ChemBridge, and a list of 14M purchasable compounds scored by our Naïve Bayes model and 2-property filter is available for download from our website (<http://chemogenomics.med.utoronto.ca/supplemental/bioactive/>).

## EXPERIMENTAL PROCEDURES

### Reagents, strains and equipment

The chemical libraries screened were obtained from ChemDiv (Divers, San Diego, CA, USA) and Chembridge (NOVACore and DIVERSet, San Diego, CA, USA) in a 96-well format at 10mM in DMSO. The Spectrum library (Microsource, Gaylordsville, CT, USA) of 2000 compounds was supplied at 2.5mM in DMSO and was a gift from D. Desveaux and D. Guttman (University of Toronto).

*E. coli* strain BW25113 (Datsenko and Wanner, 2000) (*lacI<sup>q</sup> rrnB<sub>T14</sub> lacZ<sub>WJ16</sub> hsdR514 araBAD<sub>AH33</sub> rhaBAD<sub>LD78</sub>*) was a gift from Andrew Emili (University of Toronto), *S. pombe* strain TK1/972 (*h<sup>-</sup>*) was a gift from Charlie Boone (University of Toronto) and *B. subtilis* strain 168 1A700 (*trpC2*) was a gift from Alex ter Beek and Gertien Smits (University of Amsterdam). *C. albicans* heterozygous deletion (Xu, et al., 2007) and



conditional shut-off or GRACE mutants (Roemer, et al., 2003) in the SC5314 background were a gift from Terry Roemer (Merck-Frosst Canada Ltd). *C. neoformans* strain H99 was a gift from Joseph Heitman (Duke University). A549 human lung cancer cells (ATCC number: CCL-185) and HepG2 cells (ATCC number HB-8065) were obtained from American Type Culture Collection (ATCC, Rockville MD, USA). *S. cerevisiae* WPY361 (*MATa upc2-1 ura3-1 his3-11,-15 leu2-3,-112 trp1-1*) was a gift from William Prinz (NIH, Bethesda, MD, USA) (Li and Prinz, 2004). The BX110 *fat-7(wa36);fat-5(tm420)V* double, the BX160 *fat-6(tm331)IV;fat-5(tm420)V* double, and the wild-type (N2) worm strains were obtained from the *C. elegans* Genetics Center (U. Minnesota) maintained at 20°C using standard techniques (Lewis, 1995).

Growth assays were performed in clear, flat bottom 48-, 96- and 384-well microplates (Greiner) sealed with adhesive plate seals (Cat. No. AB-0580, ABgene) using a custom developed platform incorporating microplate readers GENios, Infinite, and Safire<sup>2</sup> (Tecan-US, Durham NC, USA), the Packard Multiprobe II four-probe liquid-handling system (PerkinElmer, Waltham MA, USA). Genome-wide assays were analyzed on Genflex\_Tag\_16K\_dev microarrays (Item No. 511331, Affymetrix, Santa Clara CA, USA) using GeneChip Fluidics Station 450 and GeneChip Scanner 3000 (Affymetrix). For protocol detail see (Pierce, et al., 2007).

### Screening chemical libraries on *S. cerevisiae*

Our wild-type yeast strain, BY4743 (*MATa/a his3 1/his3 1 leu2 0/leu2 0 LYS2/lys2 0 met15 0/MET15 ura3 0/ura3 0*), was grown in YPD medium at 30°C. 20mM HEPES pH 7.0 was added to the YPD medium as indicated in the supplementary files available online. Cells were diluted from a fresh overnight culture to 0.0625 OD<sub>595</sub> in a final volume of 100µL for 96-well and 30µL for 384-well plates. Compounds were added to the culture using a 2µL or 600nL pin tool (V&P Scientific, San Diego, CA, USA) for 96-well or 384-well microplates respectively, to dilute the compounds 50 times to a final DMSO concentration of 2%. The ChemDiv Diverse, NOVACore SAR and NOVACore DIVERSet libraries were screened at 200µM final concentration, and the Spectrum library was screened at 50µM final concentration. Yeast growth was monitored for up to 24 hours by measuring the OD<sub>595</sub> every 15 minutes as described (Giaever, et al., 2004; Lee, et al., 2005). The majority of the compounds screened were soluble at 200µM, with less than 8% compounds having a starting optical density that was significantly different from the DMSO controls indicating either a colored compound or a solubility issue.

The fitness of BY4743 in compound was expressed as the ratio of the average generation time (AvgG) (AvgG reference/AvgG compound) where the reference condition was grown on the same plate in 2% DMSO. Average generation time is calculated by (time to 5 generations/5)(Lee, et al., 2005). A compound was scored as active when the ratio AvgG was 0.7 or less, corresponding to an IC<sub>30</sub> or greater. Automatic flagging of actives was confirmed by visual inspection of the data.

## Screening chemical libraries on *E. coli*, *B. subtilis*, *S. pombe*, *C. albicans*, *C. neoformans*, *C. elegans* and human cell line A549

Compounds scored as growth inhibitory were transferred to a new 96-well microplate, the hit plate. Hit plates and random naïve plates were screened at 200 $\mu$ M final concentration on *E. coli* strain BW25113 grown in Luria Broth (LB), *B. subtilis* strain 1A700 grown in Nutrient Broth (NB), *S. pombe* strain TK1 grown in YES medium, *S. cerevisiae* strain BY4743, *C. neoformans* strain H99 and *C. albicans* strain HIS3 grown in YPD medium. All media was buffered with 20mM HEPES pH 7 and growth temperature was 37°C for bacteria and 30°C for yeast.

A compound was considered active in *S. pombe*, *C. albicans*, *C. neoformans*, *E. coli* or *B. subtilis* if the area under the growth curve after 20 hours of growth was 50% of that compared to the DMSO control (ratio (compound/control) < 0.5). Compounds that showed a high ratio, defined as greater than 1.2 were excluded. We found the area under the growth curve to be a more robust method for measuring growth rate in the other organisms than the ratio AvgG. A value of 0.5 gives a similar hit-rate as a ratio AvgG of 0.7 in *S. cerevisiae*.

Phenotypic screening of *C. elegans* was performed as reported previously (Kwok, et al., 2006). In brief, molecules were screened in duplicate in 24-well format at 25 $\mu$ M concentration. Two L4 stage N2 animals were deposited per well on agar and the progeny were visually assessed for phenotype, including slow growth, egg laying abnormalities and embryonic lethality, using an MZ12 dissection microscope (Leica Microsystems GmbH, Wetzlar, Germany).

A549 human lung cancer cells were maintained in Dulbecco's Modified Eagle medium (Wisent, St-Bruno, QC, Canada) supplemented with 10% fetal bovine serum (Wisent) and 100 U/ml penicillin/streptomycin (Wisent) in a humidified incubator with 5% CO<sub>2</sub> at 37°C. The cells were seeded in 96-well plates with a density of 2200 cells per well and treated for 48 hours with 50 $\mu$ M compound in 0.5% DMSO. Cell survival was measured using the Sulforhodamine B (SRB) colorimetric assay (Vichai and Kirtikara, 2006) and readout using a SpectraMax Plus<sup>384</sup> (Molecular Devices, Sunnyvale, CA, USA) with the following modification, the cells were stained with 50 $\mu$ L of 0.4% SRB. Actives for the A549 cells were defined as compounds causing 50% viability after 48 hours of growth in the presence of the compound.

### Haploinsufficiency Profiling (HIP) Assay

By serial dilution of 16 hit plates, each containing ~86 active compounds, 79 compounds were found to completely inhibit the growth of wild-type *S. cerevisiae* at a four-fold dilution (50 $\mu$ M). 20 diverse compounds were selected for testing in the HIP assay. The molecular weights of these 20 compounds were verified by liquid chromatography and mass spectrometry to confirm their structure (see Supplementary Methods).

The HIP assay was performed as previously described (Pierce, et al., 2007). Two biological replicates were generated for each compound condition. A significant hit is defined as a gene with a log<sub>2</sub> ratio greater than 1 of the intensity of the DMSO control intensity/drug treatment intensity in both replicates. All raw and ratio data files are available on the

supplementary website. All HIP profiles are shown in Figure S2. In addition, the microarray data is available on ArrayExpress (<http://www.ebi.ac.uk/microarray-as/ae/>).

## Computational Analysis

All chemical analysis and Naïve Bayes model building was performed using the cheminformatic package in Pipeline Pilot version 6.1 (Scitegic Inc. Accelyrs, San Diego, CA USA). Marvin version 5.4.1 (ChemAxon, <http://www.chemaxon.com>) was used for drawing and displaying chemical structures.

Pipeline Pilot was used to standardize the representation of all compounds studied including removing inorganic compounds, salts and duplicates. All data used for the model building is available on our supplementary website. Three methods were tested to represent molecules for the Naïve Bayes model using five-fold cross validation: 1) A vector constructed from physical properties for each compound (LogP, Molecular Weight, Molecular Polar Surface Area, Molecular Solvent Accessible Area, the number of Hydrogen acceptors and donors, the number of rotatable bonds, the number of rings, and the number of aromatic rings). 2) A vector based on MDL Public Keys (Durant, et al., 2002) where the presence or absence of specific substructures is recorded. 3) A vector using the Extended Connectivity Fingerprints (Rogers, 2005) method where the compound is represented by overlapping fragments of a diameter of up to 2/4 bond lengths (ECFP\_2/ECFP\_4). The enrichment factor was used as a measure of accuracy. This is calculated by ranking the library of compounds to be tested by the model score. Next, for different thresholds the number of observed actives was compared with the number of actives expected by random selection.

Tanimoto coefficient, also known as the Jaccard Coefficient (Rogers and Tanimoto, 1960), was used to calculate the similarity between two compounds and calculated based on the number of features in common between the compounds divided by the total number of features present.

LogP values were calculated using the Ghose and Crippen algorithm (Viswanadhan, et al., 1989).

## Statistical Analysis

The significance of the effect of the 2-property and Lipinski filters was calculated using the hypergeometric test, termed “phyper” function, in R. To test if the distribution of LogP and the number of hydrogen acceptors for active and inactive compounds was significantly different, a two-sample Kolmogorov-Smirnov test (Durbin, 1973) was implemented using R.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank Kahlin Cheung-Ong for performing the SRB assay on cultured human cells, Sally Cheung for help with screening, Karen Robinson (Dalton Pharma Services, Toronto, Canada) for verifying the molecular weight of 60 compounds and William Lee for helpful comments on the manuscript. This work was supported by grants from the

NHGRI to CN and GG, the CIHR to CN (MOPS-84305) and to GG (MOPS-81340), a grant from the Canadian Cancer Society to GG and CN (#020380), NSERC support to GDB and IMW and a Marie Curie fellowship to IMW. GG is a Canada Research Chair in Chemical Genetics.

## REFERENCES

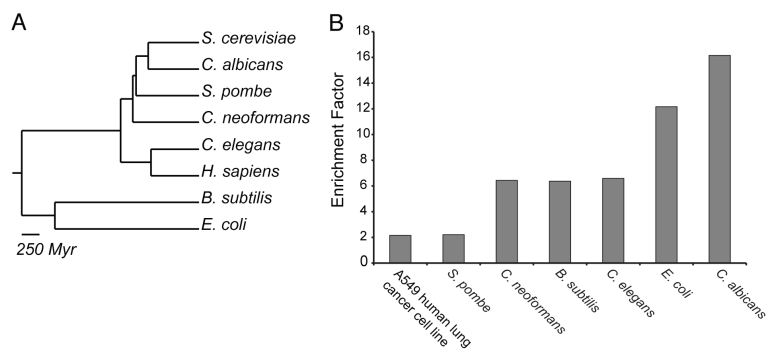
- Agresti JJ, Antipov E, Abate AR, Ahn K, Rowat AC, Baret J-C, Marquez M, Klibanov AM, Griffiths AD, Weitz DA. Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proc Natl Acad Sci USA*. 2010; 107:4004–4009. [PubMed: 20142500]
- Al-Awqati Q. One hundred years of membrane permeability: does Overton still rule. *Nat Cell Biol*. 1999; 1:E201–202. [PubMed: 10587658]
- Austin CP, Brady LS, Insel TR, Collins FS. NIH Molecular Libraries Initiative. *Science*. 2004; 306:1138–1139. [PubMed: 15542455]
- Baetz K, McHardy L, Gable K, Tarling T, Reberieux D, Bryan J, Andersen RJ, Dunn T, Hieter P, Roberge M. Yeast genome-wide drug-induced haploinsufficiency screen to determine drug mode of action. *Proc. Natl. Acad. Sci. U. S. A.* 2004; 101:4525–4530. [PubMed: 15070751]
- Blackburn AS, Avery SV. Genome-wide screening of *Saccharomyces cerevisiae* to identify genes required for antibiotic insusceptibility of eukaryotes. *Antimicrob. Agents Chemother*. 2003; 47:676–681. [PubMed: 12543677]
- Botet J, Mateos L, Revuelta JL, Santos MA. A chemogenomic screening of sulfanilamide-hypersensitive *Saccharomyces cerevisiae* mutants uncovers ABZ2, the gene encoding a fungal aminodeoxychorismate lyase. *Eukaryot Cell*. 2007; 6:2102–2111. [PubMed: 17873082]
- Botstein D, Fink GR. Yeast: an experimental organism for modern biology. *Science*. 1988; 240:1439–1443. [PubMed: 3287619]
- Brock TJ, Browse J, Watts JL. Genetic regulation of unsaturated fatty acid composition in *C. elegans*. *PLoS Genet*. 2006; 2:e108. [PubMed: 16839188]
- Burns AR, Wallace IM, Wildenhain J, Tyers M, Giaever G, Bader GD, Nislow C, Cutler SR, Roy PJ. A predictive model for drug bioaccumulation and bioactivity in *Caenorhabditis elegans*. *Nat Chem Biol*. 2010; 6:549–557. [PubMed: 20512140]
- Buurman ET, Andrews B, Blodgett AE, Chavda JS, Schnell NF. Utilization of target-specific, hypersensitive strains of *Saccharomyces cerevisiae* to determine the mode of action of antifungal compounds. *Antimicrob. Agents Chemother*. 2005; 49:2558–2560. [PubMed: 15917573]
- Charlton-Menys V, Durrington PN. Squalene synthase inhibitors : clinical pharmacology and cholesterol-lowering potential. *Drugs*. 2007; 67:11–16. [PubMed: 17209661]
- Crowley JH, Leak FW, Shianna KV, Tove S, Parks LW. A mutation in a purported regulatory gene affects control of sterol uptake in *Saccharomyces cerevisiae*. *J. Bacteriol*. 1998; 180:4177–4183. [PubMed: 9696767]
- Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci USA*. 2000; 97:6640–6645. [PubMed: 10829079]
- Dias PJ, Teixeira MC, Telo JP, Sa-Correia I. Insights into the mechanisms of toxicity and tolerance to the agricultural fungicide mancozeb in yeast, as suggested by a chemogenomic approach. *Omics*. 2010; 14:211–227. [PubMed: 20337531]
- Dielbandhoosing SK, Zhang H, Caro LH, van der Vaart JM, Klis FM, Verrips CT, Brul S. Specific cell wall proteins confer resistance to nisin upon yeast cells. *Appl. Environ. Microbiol*. 1998; 64:4047–4052. [PubMed: 9758839]
- Dorer RK, Zhong S, Tallarico JA, Wong WH, Mitchison TJ, Murray AW. A small-molecule inhibitor of Mps1 blocks the spindle-checkpoint response to a lack of tension on mitotic chromosomes. *Curr. Biol*. 2005; 15:1070–1076. [PubMed: 15936280]
- Durant JL, Leland BA, Henry DR, Nourse JG. Reoptimization of MDL keys for use in drug discovery. *J. Chem. Inf. Comput. Sci*. 2002; 42:7.
- Durbin, J. Distribution theory for tests based on the sample distribution function; 1973. p. 64
- Endo A. Chemistry, biochemistry, and pharmacology of HMG-CoA reductase inhibitors. *Klin. Wochenschr*. 1988; 66:421–427. [PubMed: 3294493]

- Ericson E, Gebbia M, Heisler LE, Wildenhain J, Tyers M, Giaever G, Nislow C. Off-target effects of psychoactive drugs revealed by genome-wide assays in yeast. *PLoS Genet.* 2008; 4:e1000151. [PubMed: 18688276]
- Ernst R, Kueppers P, Stindt J, Kuchler K, Schmitt L. Multidrug efflux pumps: substrate selection in ATP-binding cassette multidrug efflux pumps--first come, first served? *The FEBS journal.* 2010; 277:540–549. [PubMed: 19961541]
- Flanagan CA, Schnieders EA, Emerick AW, Kunisawa R, Admon A, Thorner J. Phosphatidylinositol 4-kinase: gene structure and requirement for yeast cell viability. *Science.* 1993; 262:1444–1448. [PubMed: 8248783]
- Fouchet MH, Donche F, Martin C, Bouillot A, Junot C, Boullay AB, Potvain F, Magny SD, Coste H, Walker M, et al. Design and evaluation of a novel series of 2,3-oxidosqualene cyclase inhibitors with low systemic exposure, relationship between pharmacokinetic properties and ocular toxicity. *Bioorg. Med. Chem.* 2008; 16:6218–6232. [PubMed: 18467104]
- Frearson J, Collie I. HTS and hit finding in academia - from chemical genomics to drug discovery. *Drug Discovery Today.* 2009; 14:1150–1158. [PubMed: 19793546]
- Gamo F-J, Sanz LM, Vidal J, de Cozar C, Alvarez E, Lavandera J-L, Vanderwall DE, Green DVS, Kumar V, Hasan S, et al. Thousands of chemical starting points for antimalarial lead identification. *Nature.* 2010; 465:305–310. [PubMed: 20485427]
- Giaever G, Flaherty P, Kumm J, Proctor M, Nislow C, Jaramillo DF, Chu AM, Jordan MI, Arkin AP, Davis RW. Chemogenomic profiling: identifying the functional interactions of small molecules in yeast. *Proc Natl Acad Sci USA.* 2004; 101:793–798. [PubMed: 14718668]
- Giaever G, Shoemaker DD, Jones TW, Liang H, Winzeler EA, Astromoff A, Davis RW. Genomic profiling of drug sensitivities via induced haploinsufficiency. *Nat. Genet.* 1999; 21:278–283. [PubMed: 10080179]
- Gleeson MP, Hersey A, Montanari D, Overington J. Probing the links between in vitro potency, ADMET and physicochemical parameters. *Nat Rev Drug Discov.* 2011; 10:197–208. [PubMed: 21358739]
- Hacker, MP.; Messer, WS.; Bachmann, KA. *Pharmacology: Principles and Practice.* 2009.
- Hillenmeyer ME, Fung E, Wildenhain J, Pierce SE, Hoon S, Lee W, Proctor M, St Onge RP, Tyers M, Koller D, et al. The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science.* 2008; 320:362–365. [PubMed: 18420932]
- Hu W, Sillaots S, Lemieux S, Davison J, Kauffman S, Breton A, Linteau A, Xin C, Bowman J, Becker J, et al. Essential gene identification and drug target prioritization in *Aspergillus fumigatus*. *PLoS Pathog.* 2007; 3:e24. [PubMed: 17352532]
- Irwin JJ, Shoichet BK. ZINC--a free database of commercially available compounds for virtual screening. *Journal of chemical information and modeling.* 2005; 45:177–182. [PubMed: 15667143]
- Jolidon S, Polak AM, Guerry P, Hartman PG. Inhibitors of 2,3-oxidosqualene lanosterol-cyclase as potential antifungal agents. *Biochem. Soc. Trans.* 1990; 18:47–48. [PubMed: 2185085]
- Khozoe C, Pleass RJ, Avery SV. The antimalarial drug quinine disrupts Tat2p-mediated tryptophan transport and causes tryptophan starvation. *J. Biol. Chem.* 2009; 284:17968–17974. [PubMed: 19416971]
- Klekota J, Roth FP. Chemical substructures that enrich for biological activity. *Bioinformatics.* 2008; 24:2518–2525. [PubMed: 18784118]
- Kolaczowski M, Kolaczowska A, Luczynski J, Witek S, Goffeau A. 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.* 1998; 4:143–158. [PubMed: 9818966]
- Kwak YS, Han S, Thomashow LS, Rice JT, Paulitz TC, Kim D, Weller DM. *Saccharomyces cerevisiae* genome-wide mutant screen for sensitivity to 2,4-diacetylphloroglucinol, an antibiotic produced by *Pseudomonas fluorescens*. *Appl. Environ. Microbiol.* 2011; 77:1770–1776. [PubMed: 21193664]

- Kwok T, Ricker N, Fraser R, Chan AW, Burns A, Stanley EF, Mccourt P, Cutler S, Roy P. A small-molecule screen in *C. elegans* yields a new calcium channel antagonist. *Nature*. 2006; 441:91–95. [PubMed: 16672971]
- Lain S, Hollick JJ, Campbell J, Staples OD, Higgins M, Aoubala M, McCarthy A, Appleyard V, Murray KE, Baker L, et al. Discovery, in vivo activity, and mechanism of action of a small-molecule p53 activator. *Cancer cell*. 2008; 13:454–463. [PubMed: 18455128]
- Lee W, St Onge RP, Proctor M, Flaherty P, Jordan MI, Arkin AP, Davis RW, Nislow C, Giaever G. Genome-wide requirements for resistance to functionally distinct DNA-damaging agents. *PLoS Genet*. 2005; 1:e24. [PubMed: 16121259]
- Lees ND, Skaggs B, Kirsch DR, Bard M. Cloning of the late genes in the ergosterol biosynthetic pathway of *Saccharomyces cerevisiae*--a review. *Lipids*. 1995; 30:221–226. [PubMed: 7791529]
- Lenhard JM. Lipogenic Enzymes as Therapeutic Targets for Obesity and Diabetes. *Curr. Pharm. Des.* 2011; 17:325–331. [PubMed: 21375498]
- Lewis, JA.; Fleming, JT. *Caenorhabditis elegans: Modern Biological Analysis of an Organism*. Academic Press; San Diego, CA: 1995.
- Li X, Zolli-Juran M, Cechetto JD, Daigle DM, Wright GD, Brown ED. Multicopy suppressors for novel antibacterial compounds reveal targets and drug efflux susceptibility. *Chem. Biol.* 2004; 11:1423–1430. [PubMed: 15489169]
- Li Y, Prinz WA. ATP-binding cassette (ABC) transporters mediate nonvesicular, raft-modulated sterol movement from the plasma membrane to the endoplasmic reticulum. *J. Biol. Chem.* 2004; 279:45226–45234. [PubMed: 15316012]
- Lipinski C, Hopkins A. Navigating chemical space for biology and medicine. *Nature*. 2004; 432:855–861. [PubMed: 15602551]
- Lipinski C, Lombardo F, Dominy BW. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev.* 1997; 23:3–25.
- Lum PY, Armour CD, Stepaniants SB, Cavet G, Wolf MK, Butler JS, Hinshaw JC, Garnier P, Prestwich GD, Leonardson A, et al. Discovering modes of action for therapeutic compounds using a genome-wide screen of yeast heterozygotes. *Cell*. 2004; 116:121–137. [PubMed: 14718172]
- Morgan DJ, Weisenhaus M, Shum S, Su T, Zheng R, Zhang C, Shokat KM, Hille B, Babcock DF, McKnight GS. Tissue-specific PKA inhibition using a chemical genetic approach and its application to studies on sperm capacitation. *Proc. Natl. Acad. Sci. U. S. A.* 2008; 105:20740–20745. [PubMed: 19074277]
- Muegge I. Selection criteria for drug-like compounds. *Med. Res. Rev.* 2003; 23:302–321. [PubMed: 12647312]
- Muren E, Oyen M, Barmark G, Ronne H. Identification of yeast deletion strains that are hypersensitive to brefeldin A or monensin, two drugs that affect intracellular transport. *Yeast*. 2001; 18:163–172. [PubMed: 11169758]
- Ntambi JM, Miyazaki M, Stoehr JP, Lan H, Kendziorski CM, Yandell BS, Song Y, Cohen P, Friedman JM, Attie AD. Loss of stearoyl-CoA desaturase-1 function protects mice against adiposity. *Proc Natl Acad Sci USA*. 2002; 99:11482–11486. [PubMed: 12177411]
- Oprea TI, Bologa CG, Boyer S, Curpan RF, Glen RC, Hopkins AL, Lipinski CA, Marshall GR, Martin YC, Ostopovici-Halip L, et al. A crowdsourcing evaluation of the NIH chemical probes. *Nat Chem Biol*. 2009; 5:441–447. [PubMed: 19536101]
- Paixao SM, Silva L, Fernandes A, O'Rourke K, Mendonca E, Picado A. Performance of a miniaturized algal bioassay in phytotoxicity screening. *Ecotoxicology*. 2008; 17:165–171. [PubMed: 17978872]
- Paulusma CC, Oude Elferink RPJ. The type 4 subfamily of P-type ATPases, putative aminophospholipid translocases with a role in human disease. *Biochim. Biophys. Acta*. 2005; 1741:11–24. [PubMed: 15919184]
- Pierce SE, Davis RW, Nislow C, Giaever G. Genome-wide analysis of barcoded *Saccharomyces cerevisiae* gene-deletion mutants in pooled cultures. *Nature protocols*. 2007; 2:2958–2974.
- Proctor, M.; Urbanus, ML.; Fung, EL.; Jaramillo, DF.; Davis, RW.; Nislow, C.; Giaever, G. The Automated Cell, Compound and Environment Screening System (ACCESS) for Chemogenomic

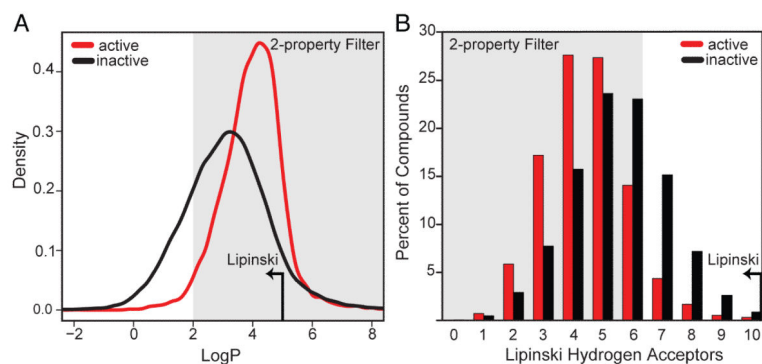


- Screening. In: I, CJ.; G, OS., editors. *Yeast Systems Biology. Methods and Protocols*. Humana Press. Springer; New York: 2011. p. xxx-xxxx.
- Roemer T, Jiang B, Davison J, Ketela T, Veillette K, Breton A, Tandia F, Linteau A, Sillaots S, Marta C, et al. Large-scale essential gene identification in *Candida albicans* and applications to antifungal drug discovery. *Mol. Microbiol.* 2003; 50:167–181. [PubMed: 14507372]
- Rogers B, Decottignies A, Kolaczowski M, Carvajal E, Balzi E, Goffeau A. The pleiotropic drug ABC transporters from *Saccharomyces cerevisiae*. *J. Mol. Microbiol. Biotechnol.* 2001; 3:207–214. [PubMed: 11321575]
- Rogers D. Using Extended-Connectivity Fingerprints with Laplacian-Modified Bayesian Analysis in High-Throughput Screening Follow-Up. *Journal of Biomolecular Screening.* 2005; 10:682–686. [PubMed: 16170046]
- Rogers DJ, Tanimoto TT. A Computer Program for Classifying Plants. *Science.* 1960; 132:1115–1118. [PubMed: 17790723]
- Schuldiner M, Collins SR, Thompson NJ, Denic V, Bhamidipati A, Punna T, Ihmels J, Andrews B, Boone C, Greenblatt JF, et al. Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell.* 2005; 123:507–519. [PubMed: 16269340]
- Specht KM, Shokat KM. The emerging power of chemical genetics. *Curr. Opin. Cell Biol.* 2002; 14:155–159. [PubMed: 11891113]
- Vichai V, Kirtikara K. Sulforhodamine B colorimetric assay for cytotoxicity screening. *Nature protocols.* 2006; 1:1112–1116.
- Viswanadhan VN, Ghose AK, Revankar GR, Robins RK. Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics. *J. Chem. Inf. Comput. Sci.* 1989; 29:163–172.
- Voyron S, Rocco F, Ceruti M, Forni P, Pla AF, Sarpietro MG, Varese GC, Marchisio VF. Antifungal activity of bis-azasqualenes, inhibitors of oxidosqualene cyclase. *Mycoses.* 2010; 53:481–487. [PubMed: 19549106]
- Wang Y, Bolton E, Dracheva S, Karapetyan K, Shoemaker BA, Suzek TO, Wang J, Xiao J, Zhang J, Bryant SH. An overview of the PubChem BioAssay resource. *Nucleic Acids Res.* 2010; 38:D255–266. [PubMed: 19933261]
- Wheeler GN, Brändli AW. Simple vertebrate models for chemical genetics and drug discovery screens: lessons from zebrafish and *Xenopus*. *Dev. Dyn.* 2009; 238:1287–1308. [PubMed: 19441060]
- Workman P, Collins I. Probing the probes: fitness factors for small molecule tools. *Chem. Biol.* 2010; 17:561–577. [PubMed: 20609406]
- Xu D, Jiang B, Ketela T, Lemieux S, Veillette K, Martel N, Davison J, Sillaots S, Trosok S, Bachewich C, et al. Genome-wide fitness test and mechanism-of-action studies of inhibitory compounds in *Candida albicans*. *PLoS Pathog.* 2007; 3:e92. [PubMed: 17604452]
- Xu D, Sillaots S, Davison J, Hu W, Jiang B, Kauffman S, Martel N, Ocampo P, Oh C, Trosok S, et al. Chemical genetic profiling and characterization of small-molecule compounds that affect the biosynthesis of unsaturated fatty acids in *Candida albicans*. *J. Biol. Chem.* 2009; 284:19754–19764. [PubMed: 19487691]
- Yan Z, Costanzo M, Heisler LE, Paw J, Kaper F, Andrews BJ, Boone C, Giaever G, Nislow C. Yeast Barcoders: a chemogenomic application of a universal donor-strain collection carrying bar-code identifiers. *Nat Methods.* 2008; 5:719–725. [PubMed: 18622398]



**Figure 1. Yactives Enrich for Phenotypes in Diverse Organisms**

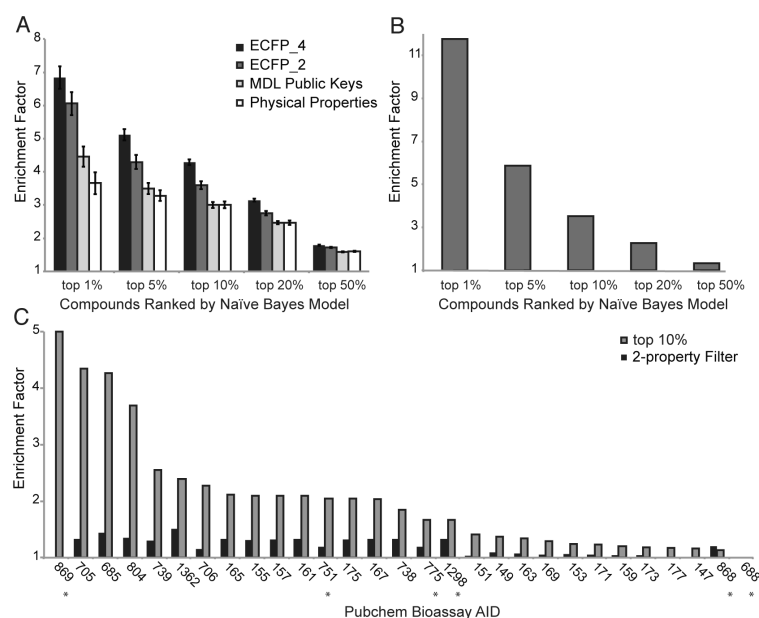
(A) Phylogenetic tree of the eight organisms screened in this study. Yeast generally requires higher screening concentrations than mammalian screens for similar inhibition levels, as shown in Figure S1 (B) Enrichment factor; ratio of hit-rate for yactive compounds compared to randomly selected compounds empirically determined by screening seven organisms.



**Figure 2. A Two-Property Filter Enriches for Yactive Compounds**

(A) Distribution plot of LogP for active compounds vs. inactive compounds. The shaded region indicates compounds that pass the 2-property filter of  $> 2$  LogP, significantly increasing the percentage of phenotypic-inducing compounds. For comparison, the Lipinski limit of  $< 5$  is indicated by an arrow.

(B) Histogram showing the percentage of compounds of active compounds vs. inactive compounds. The shaded region indicates compounds that pass the 2-property filter of  $> 6$  hydrogen acceptors, significantly increasing the fraction of phenotypic-inducing compounds. For comparison, the Lipinski limit of  $< 10$  is indicated by an arrow.

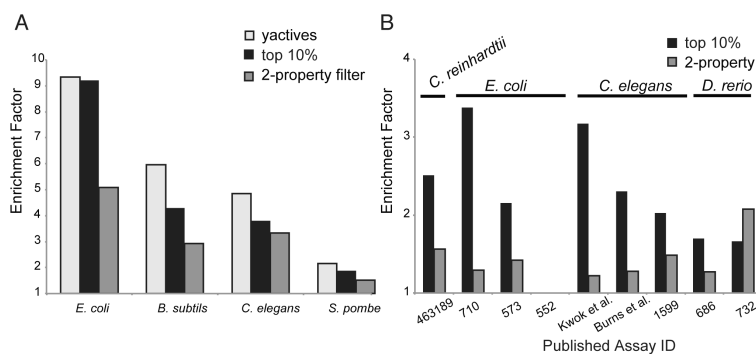


### Figure 3. Application of a Naïve Bayes Model Enriches for Yactive Compounds

(A) Cross-validation results for four different structure encoding schemes. The error bars represent the standard deviation obtained by repeating the procedure five times.

(B) Performance of the Naïve Bayes model on the Spectrum Library, a library independent of that used for training the model.

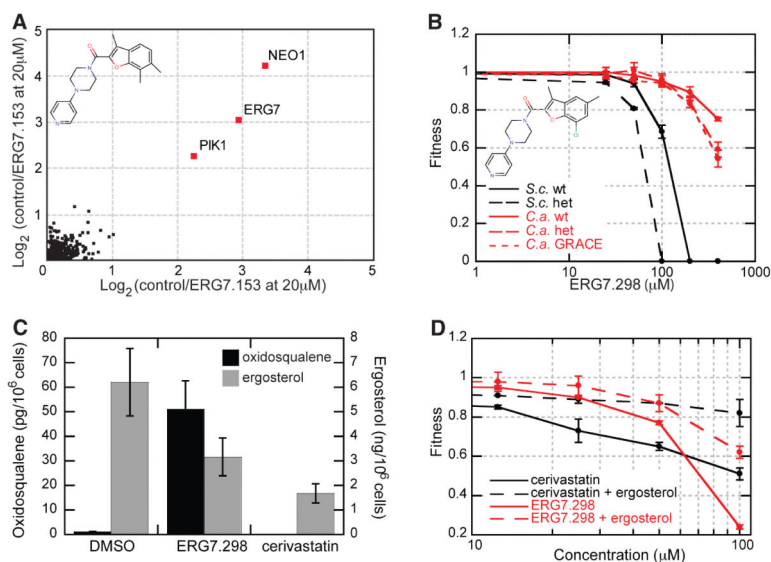
(C) The median enrichment of the Naïve Bayes model and the 2-property testing using 29 publicly available yeast BioAssays available from PubChem. Assay details are available in Table S4, results for the 2-property filter in Table S5, and the Naïve Bayes model results in Table S6. Absent bars indicate an enrichment factor <1. Assays marked with an asterisk (\*) used readouts that were not based on growth inhibition.



**Figure 4. Yactives, Application of the Naïve Bayes Model and a Two-Property Filter Enrich for Compounds that Induce Phenotypes in Model Organisms**

(A) Comparison of the Performance of the Yactives/Naïve Bayes model/2-property filter. Each bar represent the average performance over two test sets, one containing all of the compounds screened (see Table S2) and a subset containing only random compounds (see Table S7).

(B) The median performance of the Naïve Bayes model and the 2-property filter applied to nine publicly available assays, grouped by model organism. Assay details are available in Table S4, results for the 2-property filter are in Table S8 and the Naïve Bayes model results are in Table S9.



**Figure 5. Identification of a Novel Inhibitor of Lanosterol Synthase (Erg7)**

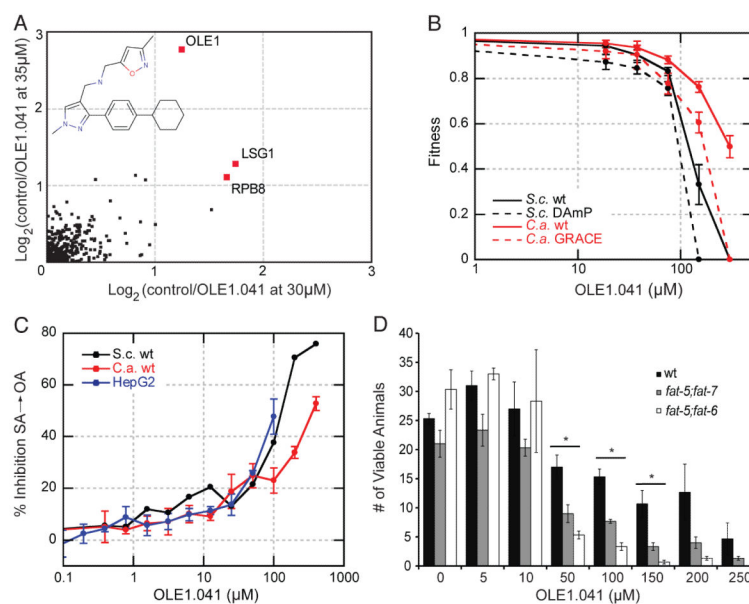
(A) Chemical structure of ERG7.153 and the HIP results for this compound. Sensitivity of a deletion strain to compound is expressed as the  $\log_2$  ratio of control intensity/treatment intensity. The graph shows the results from two replicate experiments performed using 20 $\mu$ M ERG7.153. Deletion strains are labeled with their gene name if the  $\log_2$  ratio is above one in both replicates.

(B) Dose response curve and chemical structure for ERG7.298 (a close analog of ERG7.153, original compound no longer available), a putative Erg7 inhibitor, in wild-type strain of *S. cerevisiae* (*S.c.*) and the *erg7* heterozygous deletion strain (*het*) as well as for *C. albicans* (*C.a.*) and the tet-repressable *erg7* GRACE mutant (Roemer, et al., 2003). The fitness of a strain in compound was calculated from high-resolution growth curves as the ratio (control/treatment) of the average doubling time in the first five generations of growth. Experiments were performed in quadruplicate and error bars represent the standard deviation.

(C) Bar graph showing the amount of 2,3-oxidosqualene (black) and ergosterol (grey) in DMSO, ERG7.298 and cerivastatin treated cells. Experiments were performed in triplicate, the error bars represent the standard deviation.

(D) Dose response curve of cerivastatin and ERG7.298 in the *upc2-1* strain grown with and without 20 $\mu$ g/ml ergosterol. The fitness was calculated as in (B), the experiment was performed in triplicate, the error bars represent the standard deviation. Additional information can be found in Figure S3.





**Figure 6. Identification of a Novel Inhibitor of Delta(9) Fatty Acid Desaturase (Ole1)**

(A) Chemical structure of OLE1.041 and the HIP results for this compound performed using 30 and 35 $\mu$ M OLE1.041, details as in Figure 5(A). All 20 HIP profiles generated in this study are shown in Figure S2.

(B) Dose response curve for OLE1.041, in wild-type strains of *S. cerevisiae* (*S.c.*) compared to the *ole1* DAmP strain and *C. albicans* (*C.a.*) wild-type strain compared to the *ole1* GRACE strain, details as in Figure 5(B).

(C) % inhibition of radiolabelled stearic acid to oleic acid conversion by OLE1.041 *in vivo* in *S. cerevisiae*, *C. albicans* and human HepG2 cells. The experiment was performed in duplicate for *S. cerevisiae* and quadruplicate for *C. albicans* and HepG2, the error bars represent the standard deviation.

(D) Phenotypic analysis of OLE1.041 in *C. elegans* wild-type, *fat-5;fat-7* and *fat-5;fat-6* mutant strains. The bar graph shows the number of viable animals remaining following treatment with a DMSO control and across a range of OLE1.041 concentrations. Significant differences ( $p$ -value  $<0.05$ ) between the wild-type and mutant strains are indicated with an asterisk (\*). The error bars represent the standard deviation in repeating the experiment in triplicate.