**­Deciphering Complex Traits with High-Order Genetic Analysis**

Albi Celaj1,2,3, Marinella Gebbia1, Louai Musa1, Atina G. Cote2, Jamie Snider1, Victoria Wong1, Minjeong Ko1,2,4, Tiffany Fong1, Paul Bansal1,2, Joseph C. Mellor1, Gireesh Seesankar1, Maria Nguyen1, Shijie Zhou1,3, Nishka Kishore1,2, Igor Stagljar1,3,5, Yo Suzuki6, Nozomu Yachie1,6,7,8,9,10,\*, and Frederick P. Roth1,2,3,6,9,11,12,\*

[Author list and order is not final]

1Donnelly Centre, University of Toronto, Toronto, Ontario, Canada.

2Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada.

3Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada.

4Department of Computer Science, University of Toronto, Toronto, Ontario, Canada.

5Department of Biochemistry, University of Toronto, Toronto, Ontario, Canada.

6Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA, USA.

7Synthetic Biology Division, Research Center for Advanced Science and Technology, the University of Tokyo, Tokyo, Japan.

8Department of Biological Sciences, School of Science, the University of Tokyo, Tokyo, Japan.

9Institute for Advanced Biosciences, Keio University, Tsuruoka, Japan.

10PRESTO, Japan Science and Technology Agency, Tokyo, Japan.

11Canadian Institute for Advanced Research, Toronto, Ontario, Canada.

12Lead contact

\*Correspondence:[fritz.roth@utoronto.ca](mailto:fritz.roth@utoronto.ca) (F.P.R.), [yachie@synbiol.rcast.u-tokyo.ac.jp](mailto:yachie@synbiol.rcast.u-tokyo.ac.jp) (N.Y.)

**Summary**

Complex traits, including most common human genetic diseases, have remained opaque because of daunting experimental and statistical challenges in exhaustively testing variant combinations in multiple genes. Here, we describe an ‘*X*-gene’ genetic analysis (XGA) approach that engineers and profiles combinatorial perturbations in multiple genes. We demonstrate XGA on yeast ABC transporters by engineering 5,352 strains, each deleted for a random subset of 16 transporters, and profiling each strain’s resistance to 16 compounds. XGA yielded 85,632 genotype-to-resistance observations, revealing high-order genetic interactions (involving 3 or more genes) for 13 of the 16 transporters studied. Neural networks yielded intuitive functional models, which were explored further for a fluconazole resistance trait that was influenced non-additively by five genes. Together, results showed that the systematic combinatorial genetic engineering and multiplexed profiling strategy of XGA can provide functional models of complex traits.

**Introduction**

Genes often encode interdependent and functionally-overlapping molecular systems, such that combinations of genetic variants can yield surprising phenotypes (Hartman et al., 2001). This phenomenon defines complex traits and genetic interactions. Observing an interaction when knocking out two genes in pairs, e.g., using synthetic genetic array analysis (SGA), has systematically uncovered functional relationships in yeast (Costanzo et al., 2016) and human cells (Horlbeck et al., 2018; Shen and Ideker, 2018), improving our understanding of gene function (Costanzo et al., 2016) and order-of-action in biological pathways (St Onge et al., 2007).

Genetic interactions with higher complexity, e.g. surprising phenotypes that result from three-gene knockouts, can reveal additional important functions (Haber et al., 2013; Kuzmin et al., 2018). Indeed, three-gene interactions are expected to outnumber two-gene interactions by 100-fold (Kuzmin et al., 2018). Beyond three-gene effects, yet higher-order gene variant combinations have yielded interesting effects – e.g., involving four (Takahashi and Yamanaka, 2006), five (Taylor et al., 2014), seven (Beh et al., 2001), and over twenty genes (Wieczorke et al., 1999). Systematic maps of higher-order interactions between variants at a single locus have been used to understand several diverse processes (Baeza-Centurion et al., 2019; Domingo et al., 2018; Sarkisyan et al., 2016). However, higher-order interactions between variants in different genes have remained poorly characterized, limiting functional understanding of complex multi-gene dependencies.

To systematically investigate complex genetic dependencies beyond 1- and 2-gene genetic analysis, we developed an ‘*X-*gene’ genetic analysis (XGA) strategy that uses many combinations of engineered multi-gene perturbations to profile and interpret high-order genetic interactions. We demonstrate XGA on *S. cerevisiae* ABC transporters, which are involved in cellular efflux of small molecules, and for which several informative multi-knockout phenotypes have been reported (Khakhina et al., 2015; Kolaczkowska et al., 2008; Suzuki et al., 2011). More specifically, we apply XGA systematically to the entire set of 16 yeast ABC transporters that have been implicated in multi-drug resistance. By revealing a multi-knockout genetic landscape for 16 bioactive compounds (‘drugs’), XGA uncovered many drug-dependent high-order genetic interactions involving as many as five genes. A neural network trained on XGA data provided an intuitive genotype-to-phenotype model and functional insights into this system of ABC transporters. Taken together, our results show that XGA can systematically uncover high-order genetic relationships, and use them to functionally model complex molecular systems.

**Results**

**Engineered population profiling: a scheme for generating and characterizing combinatorially complex mutants**

Here we describe our overall XGA strategy (Figure 1), then show the results for XGA as applied to sixteen yeast ABC transporters.

A simple yet powerful way to generate a complex population is to cross two outbred individuals differing at multiple unlinked loci. Progeny (‘segregants’) will inherit a random subset of parental variants and each segregant can then be genotyped and profiled for traits such as gene expression or drug resistance (Bloom et al., 2013). Although applied extensively to pairs of outbred parents, this approach has limitations. First, many genes involved in a process will be missed for lack of functional natural variation in the parental strains (Lee et al., 2014). Second, the fact that parents can differ at >105 positions, coupled with genetic linkage of proximal variants, can complicate identification of the causal variants at each associated locus.

To exploit the power of cross-based approaches while avoiding the limits of natural variation, we designed a population engineering strategy in which targeted polygenic variation is engineered into individual parental strains. Parental strains are then crossed to yield an ‘engineered population’. Thus, a genetic cross yields random segregation only for engineered variants.

As recently reviewed (Kebschull and Zador, 2018), many individual strains can be tracked in a complex heterogenous population using DNA barcodes. We therefore introduced a complex pool of random barcodes into a haploid parental strain (which was wildtype for all ABC transporter genes of interest in this study), as described previously (Díaz-Mejía et al., 2018) (Figure S1, see Methods for details). We crossed this pool *en masse* to a previously-generated “ABC-16 strain” bearing knockouts for all sixteen of the ABC transporters that have been implicated in multidrug resistance (Suzuki et al., 2011). The ABC-16 strain contained all markers necessary to perform mating, sporulation, and selection for haploid cells, while the barcoded wild-type parent provided the marker necessary to select for cells carrying a barcoded *HO* locus (Methods). After mating, sporulation, and selection for barcoded haploid progeny of the cross, we used automated colony-picking to isolate an arrayed collection of ~5,000 MAT**a** and ~5,000 MAT**α** segregants in 384-well plates. This step generated an engineered population in which each individual haploid strain bears a random subset of knockout alleles for the target set of 16 ABC transporters.

For each strain in this arrayed population, we determined the genotype at all sixteen knockout loci and identified the barcode. To genotype, we exploited the fact that each knockout locus in the ABC-16 strain was derived from the YKO yeast deletion collection (Giaever et al., 2002; Suzuki et al., 2011) and is therefore flanked by a deletion-identifying barcode. We adapted the previously-described row-column-plate PCR (RCP-PCR) strategy (Yachie et al., 2016), in which barcodes in each segregant are amplified together with additional PCR-introduced index tags that identify the plate, row, and column of origin for each amplification product (Methods; Figure 1). Thus, a single sequencing experiment revealed all strain-specific tracking barcodes at the *HO* locus and another single experiment revealed identity of every gene deleted in the segregant at each plate location (Methods; Figure 1).

Two independent methods estimated overall per-locus genotyping accuracy to be from 93.2%-93.8% (Figure S2A and S2B, Methods). Based on correlation analysis of the genotyping data, all genes were either unlinked or weakly linked except for *BPT1* and *YBT1* (Figure S2C; r = 0.49), which are separated by 70.1kb on chromosome XII. Considering only those strains with both high-quality genotyping data and at least one unique tracking barcode, our engineered strain population included 6,826 uniquely barcoded and genotyped strains, encompassing 6,087 unique genotypes. These strains were grouped by mating type to yield one pool of 3,231 MAT**a** strains and another pool of 3,595 MAT**α** strains.

[Fritz Stopped Here]

To profile each strain’s resistance or sensitivity to 16 different antifungal or anticancer agents (‘drugs’; Data S3), we grew the strain pools competitively in each drug, and in a solvent (DMSO) condition. We used high-throughput strain barcode sequencing at five points (corresponding to 0, 5, 10, 15, and 20 generations of overall pool growth, Figure 1), to estimate growth rate and resistance for each strain in each drug (Data S5; Methods). We performed additional filtering steps, limiting analysis to strains that were well-represented in the pre-selection pool (≥30 barcode counts at t=0 in the solvent control - 5,790 [85%] of 6,826 strains), and further excluding all 437 strains exhibiting a strong baseline growth defect (i.e., showing <70% of the median baseline growth rate). In total, drug resistance was calculated for each of 2,367 MAT**a** and 2,985 MAT**α** strains, for each of the 16 drugs (Data S5).

**Grouped combinatorial profiles illustrate a complex drug-dependent genetic landscape**

For an initial analysis, we identified and quantitatively modeled associations between individual ABC transporter knockouts and drug resistance phenotypes using a generalized linear model (see Methods). We found 62 associations between knockouts and drug resistance that were reproducible in both MAT**a** and MAT**α** pools (Figure S3A). Most (58/62) of these associations involved five ‘frequently-associated’ ABC transporters—*snq2∆*, *pdr5∆*, *yor1∆*, *ycf1∆*, and *ybt1∆* (Figure S3A). For these five frequently-associated transporters, we detected 16/18 previous associations between drug resistance and individual knockouts, while revealing 40 new associations (Figure S3A; Data S7). For example, we detected 18 drug resistance associations involving the vacuolar ABC transporters *YCF1* and *YBT1*, all of which were novel (Figure S3A, Data S6). We also found 4 associations between knockouts and growth rate in the DMSO control condition (Figure S3B, Data S6). Of these 4 associations, *yor1∆* had a modest effect (7-15% decrease), while the other 3 were weak (<2% decrease; Figure S3B, Data S6).

Considering only the five frequently-associated transporters, we profiled resistance for the set of 32 (25) combinatorial genotypes. We derived a resistance profile for each combinatorial genotype by calculating, for each drug, the average resistance over all strains matching this genotype at all five genes. These grouped resistance profiles showed striking correspondence when calculated separately for MAT**a** and MAT**α** pools (Figure S4). For example, camptothecin and ketoconazole had a very high profile correlation of r ≥ 0.99 (Figure 2A). Indeed, with the exception of colchicine (r = 0.77), all drugs showed high reproducibility (r ≥ 0.94) between independent biological replicate pools (Figure 2B). The five-gene resistance profiles could be used to provide a visual overview of multi-knockout resistance in each drug. We created a ‘radial combinatorial signature’ for each drug, in which the consequences of knocking out increasingly-many ABC transporters are shown as paths leading outward from the central wild-type genotype (Figure 2C-D). As expected, radial combinatorial signatures were visually similar between independent MAT**a** and MAT**α** populations for many drugs (Figure 2D and S5). Given high reproducibility, we merged MAT**a** and MAT**α** data for subsequent analyses, except where noted (Methods).

To analyze the five-gene combinatorial resistance profiles in more detail, we visualized them as fitness landscapes (Figure 3 and S6). For some drugs, we observed a clear sensitivity effect from knocking out only one transporter – e.g. *pdr5∆* for cycloheximide and tamoxifen (Figure 3). In other drugs, we saw increased sensitivity resulting from knocking multiple transporters – e.g. the set {*snq2∆*, *pdr5∆*}under camptothecin, and the set {*snq2∆*, *pdr5∆*, *ybt1∆*, *yor1∆*} under mitoxantrone (Figure 3). These sensitivity patterns are consistent with a scenario in which one or more transporters can efflux a given drug.

In other drugs, the fitness landscapes showed more surprising multi-knockout patterns ­conveying both drug resistance and sensitivity. In benomyl, for example, we not only observed sensitivity from knocking out the known primary efflux pump *snq2∆* (20% decreased resistance, *p* = 1.4e-95; Mann-Whitney *U* test)*,* but 13% increased resistance in *pdr5∆* knockouts (p = 1.3e-41), and further resistance in the *pdr5∆ yor1∆* double-mutant (21% increased resistance; p = 1.3e-72). All of these effects had been previously reported (Kolaczkowska et al., 2008; Snider et al., 2013), and were explained by increased *SNQ2*-mediated resistance upon deleting *pdr5∆* and *yor1∆*. Consistent with this explanation, these resistance effects were more modest in a *snq2∆* background (Figure 3). A similar landscape was found in bisantrene, which also showed a strong *snq2∆* effect (Figure S6). Surprisingly, we found that the successive deletion of ABC transporters led to greater resistance for many drugs (Figure 2D and S5). For example, knocking out *pdr5∆*, *snq2∆*, *ybt1∆*, and *ycf1∆* individually or in any combination led to more valinomycin resistance than the wild-type strain (Figure 3).

**XGA reveals many complex drug-dependent genetic interactions**

To identify and model multi-gene knockout effects at all 16 transporters, we extended the above generalized linear model used to capture both single knockout effects and multi-gene interactions (see Methods). All associations and interactions that passed the significance test (*p* < 0.05 after adjusting for multiple testing) are shown in Figure 3A.

This analysis yielded genetic interactions involving two or more genes for fifteen out of sixteen (94%) of the drugs examined (Figure 3A). (The exception was beauvericin for which we only recovered the previously-reported sensitivity of *yor1∆* knockouts (Shekhar-Guturja et al., 2016).) Higher-order genetic interactions (involving three or more genes) were observed for fourteen of sixteen (88%) drugs tested (Figure 3A). Thus, XGA revealed higher-order genetic interactions involving three or more genes for nearly all drug resistance phenotypes studied. In total, genetic interactions were found for 14 of the 16 genes targeted by XGA. Of these 14 genes, 13 were involved in at least one interaction involving three or more genes. Remarkably, 11 of the 16 targeted genes were involved in at least one 5-gene interaction.

Formally identifying complex genetic interactions captured many of the effects that had been readily-apparent by manual examination of the five-gene fitness landscapes, while yielding additional effects. For example, *pdr5∆* was found to have a resistance effect under benomyl, to have a positive genetic interaction with *yor1∆* and, to have a negative genetic interaction with *snq2∆* (Figure 3A, Data S6).

This analysis uncovered strong complex interactions involving genes outside the five-frequently associated transporters. In both cisplatin and mitoxantrone, for example, a five-way positive interaction pointed to the phenomenon that addition of *bpt1∆* confers resistance in a sensitive *pdr5∆ snq2∆ ycf1∆ yor1∆* background (Figure 3B). A five-way positive interaction in bisantrene pointed to a similar, but more modest effect with *vmr1∆* in a *snq2∆ ybt1∆ ycf1∆ yor1∆* background (Figure 3B).

Complex genetic interaction analysis allowed finer parsing of the relationship between genes involved in a higher-order interaction. For example, the striking mitoxantrone sensitivity of the *snq2∆ pdr5∆ ybt1∆ yor1∆* quadruple mutant (Figure 3) was modeled as the combination two single-gene negative effects of *snq2∆* and *pdr5∆* alone, a two-gene negative interaction between *snq2∆* and *pdr5∆*, two three-gene negative interactions (between *snq2∆ pdr5∆* and each of *ybt1∆* and *yor1∆*), and a four-gene {*snq2∆*, *pdr5∆*, *ybt1∆*, *yor1∆*} negative interaction (reflecting the fact that the quadruple mutant is more sensitive than would be expected given the observed resistance of any of the three-deletion subset genotypes; Figure 3B, Data S6). Together, these complex negative genetic interaction patterns suggest that the four genes enable mitoxantrone efflux in parallel. Similar ‘parallel efflux’ genetic interaction patterns were observed, e.g. for {*pdr5∆snq2*} in camptothecin, and {*pdr5∆*, *snq2∆*, *yor1∆*}in cisplatin (Figure 3A-B, Data S6).

**Using XGA to learn intuitive genotype-to-phenotype models of the ABC transporter system**

Above, we used generalized linear models to capture complex genetic effects as a set of single-gene effects and *X*-way genetic interactions. While this model does capture complex genotype-phenotype relationships, does not efficiently convey useful intuition about the system. For example, we reasoned (without benefit of intuitive computational modeling) that a set of transporter genes showing patterns of within-set negative genetic interactions suggests that each transporter is independently capable of drug efflux. Other genetic interaction patterns led us to conclude that the presence of one transporter can positively or negatively influence the activity of another (e.g. influence on Snq2 activityfrom *PDR5* and *YOR1* in benomyl). However, manually derived genetic intuition from a complex system is laborious, and it is difficult to objectively evaluate how well these explanations correspond to observed data. To demonstrate that such intuitions can be derived more systematically from complex genotype-phenotype relationships, we developed a neural network model.

We structured the neural network model (Methods, Figure 4A) to have three layers: 1) an input layer encoding the binary genotype (***G***) for each of the 16 targeted transporters; 2) a middle ‘hidden’ layer with values (***A***; ranging from 0 to 1) that estimate the activity of each of the 16 transporters; and 3) an output layer that quantitatively describes resistance (***R***; ranging from 0 to 1) to each of 16 drugs. To represent regulatory influence relationships between transporters, the links between genotype and activity layers have (initially unknown) ‘influence’ weights (***I***), with positive weights where gene presence increases activity and negative weights where gene presence decreases activity. The links between activity and resistance layers have (initially unknown) non-negative ‘efflux’ weights (***E***) that capture the extent to which each transporter can catalyze the efflux (or otherwise reduce the activity) of each drug. The model also allowed for offset terms for both ***A*** and ***R*** (Methods).

We learned the appropriate network weights with back-propagation and stochastic gradient descent using our complete set of drug resistance phenotypes as training data (see Methods). To favor more parsimonious models, the cost function that was used to optimize network weights contained a penalty which acts to limit the number of non-zero weights (Methods, Figure S7A-B). After the learning procedure, parsimonious models were further favored by setting non-zero weights to zero if they did not consistently depart from zero between repeated runs with different initial parameter settings, or if doing so did not cause a significant difference in model predictions (Methods).

Training this model on our input dataset of 85, 632 genotype-phenotype measurements yielded an interpretable neural network with only 71 non-zero fitted parameters (6 *I*weights, 49 *E* weights, no *A* offset terms and 16 *R*offset terms). Despite its parsimonious nature, the trained neural network model largely recapitulated the observed genotype-phenotype relationships (*r* = 0.96, Figure 4C). To test that this performance generalizes to unseen data, we also trained the model using data from one mating type then tested it using independent data from the other mating type. This yielded similar performance (*r* = 0.95 and *r* = 0.96 when using either mating type **a** or **α** as training, respectively [Figure S7C]), and the resulting independently-trained models also showed strong agreement in parameter values (*r* = 0.98, Figure S7D), suggesting that model parameters were robustly determined.

The first striking result from this model was that all influence (*I*) values were either zero or negative. More specifically, only 6 out of 240 influence values were negative, while all others were zero. Thus, while some ABC transporters exhibited negative influence on other ABC transporters, our study of 16 transporters in 16 drugs found no evidence that any ABC transporter can positively influence any other ABC transporter.

The objectively-trained neural network model provided intuition that was largely in agreement with manual interpretations. For example, the manual genetic interpretation that Pdr5*,* Snq2*,* Yor1*,* and Ybt1 are each independently able to efflux mitoxantrone was also supported by positive links connecting each of these transporters to mitoxantrone (Figure 4B). The model showed Snq2 to have the highest mitoxantrone efflux activity ( = 1.8) followed by Pdr5, Yor1, and Ybt1 ( = 1.5, 0.5, 0.4, respectively; Figure 4B, Data S8). These differences were reflected in the fitness landscape: For example, while mitoxantrone resistance of a *ybt1∆yor1∆* deletion strain (∆*E* = -0.9) was not significantly different than the wild-type (*p* = 0.12), deletion of genes encoding the two transporters with the highest inferred efflux (Snq2 and Pdr5, ∆*E* = -3.3) yieldeda 7% decrease in resistance (*p* = 1.2e-70). The *I* weights also pointed to differential inhibitory effects between transporters: For example, Snq2activity is predicted to be more strongly inhibited by *PDR5* than by *YOR1* ( = -0.69 vs -0.11, Figure 4B, Data S8), which is reflected, for example, by the observation that *pdr5∆* yields greater benomyl resistance than does *yor1∆* (Figure 3).

While the neural network model was accurate overall, we could also evaluate where predictions departed systematically from observation to guide model refinements (Figure S8). For example, while XGA showed that many multi-transporter deletions resulted in *increased* valinomycin resistance (Figure 3), the neural network only captured the decreased resistance resulting from *yor1∆*, yielding poor predictions overall (*r* = 0.5, Figure 4D, left panel). Taking these results together with previous reports of improved valinomycin resistance upon deletion of all 16 transporters (Suzuki et al., 2011), we hypothesized that one or more of the transporters inhibits a valinomycin resistance factor outside of the 16 targeted genes. To formally test whether this explanation better captures the observed data, we added this hypothesized valinomycin resistance factor and its corresponding activity node to the neural network schematic (see Methods). Training this extended neural network using valinomycin data substantially improved correspondence to the observed phenotypes (*r* = 0.86, Figure 4D, right panel) and yielded a model in which *SNQ2*, *PDR5*, *YBT1,* and *YCF1* each negatively influence an unknown valinomycin resistance factor. This model improvement depended on the inclusion of this hypothesized factor (Figure S9A).

**Four genes for which deletion yields synergistic *PDR5*-dependent fluconazole resistance**

One striking phenotype revealed by XGA was a quadruple deletion—*snq2∆ ybt1∆ ycf1∆ yor1∆*—with high resistance to both fluconazole (Figure 6A) and ketoconazole (Figure S6). Interestingly, the quintuple mutant *pdr5∆ snq2∆ ybt1∆ ycf1∆ yor1∆* (adding a *pdr5∆* deletion to the quadruple mutant background) showed fluconazole sensitivity that was comparable to *pdr5∆* alone. Generalized linear regression had modeled the quadruple-knockout resistance phenomenon as the combination of three positive three-gene interactions (all of the three-knockout combinations of {*yor1∆*, *snq2∆, ybt1∆*, *ycf1∆*}except *snq2∆ ybt1∆ ycf1∆*), while its dependence on *PDR5* was modeled by three two-way negative interactions: {*pdr5∆*, *snq2∆*}, {*pdr5∆*, *ycf1∆*}, and {*pdr5∆*, *yor1∆*} (Figure 6A). To confirm the resistance observations in a more uniform genetic background, we generated a single strain for each of the 32 possible combinations of *pdr5∆, snq2∆, yor1∆, ybt1∆,* and *ycf1∆* knockouts, each in a common genetic background (Methods). Fluconazole resistance as estimated by pooled data correlated well with measures of resistance obtained for individual strains — *r* = 0.95 for the fluconazole concentration expected to yield 50% inhibition (IC50; Figure 6B) and *r* = 0.89 for total growth in fluconazole relative to no-drug conditions (Figure S10; Methods). Consistent with pooled results, *snq2∆yor1∆ybt1∆ycf1∆* had the highest resistance.

The neural network had modeled negative influence on Pdr5 from *SNQ2, YBT1, YCF1*, and *YOR1* (Figure 6C), thereby capturing that *snq∆ybt1∆ycf1∆yor1∆* should be more resistant to fluconazole than strains carrying any subset of these knockouts. Given the known protein-protein interaction between Pdr5 and Snq2 (Snider et al., 2013), and previous reports of improved Pdr5-dependent drug resistance from knocking out *snq2∆* or *yor1∆* (Kolaczkowska et al., 2008), one might hypothesize that repression of *PDR5* from these two genes may be mediated by direct interactions. This ‘direct repression’ is also supported by the observed homodimeric interactions of Pdr5 and Snq2 (Snider et al., 2010; Tarassov et al., 2008), such that heterodimerization of Pdr5 and Snq2 transporters can draw subunits away from the homodimeric Pdr5 complex, thereby reducing Pdr5 efflux activity. However, this model also predicts a similar heterodimeric interaction between Pdr5 and Yor1. Because all known protein interaction testing methods miss the majority of real interactions (Braun et al., 2009), we used two distinct assays: MYTH (Snider et al., 2013) and PCA (Tarassov et al., 2008) to test the Pdr5-Yor1 interaction. All previously-known MYTH and PCA interactions amongst Pdr5, Snq2, and Yor1 were recovered (Figure 6D, S11, S12). Although PCA (Figure S11) did not detect the predicted Pdr5-Yor1 interaction, MYTH could detect this interaction (Figure 5E, S12), thus confirming a key prediction of the direct repression model for the Pdr5-dependent decrease in fluconazole resistance provided by *YOR1*. Given a much-higher baseline abundance of Pdr5 than Snq2 (Wang et al., 2015), a heterodimeric repression model is also consistent with the modeled effect that negative influence of Snq2 by Pdr5 will be greater than negative influence of Pdr5by Snq2 (*I =* -0.69vs-0.11*,* Figure 6D). This is because a greater proportion of Snq2 would be affected by each heterodimeric interaction than would Pdr5 (Celaj et al., 2017)*.*

However, it has been reported that knocking out multiple transporters can also lead to a transcriptional response. For example, a previous study found that while *pdr5∆* and *yor1∆* each resulted in increased benomyl resistance, combining *pdr5∆yor1∆* additionally resulted in *SNQ2* mRNA induction (Snider et al., 2013). Similarly, *yor1∆snq2∆* has shown evidence for *PDR5* induction, whereas the transcriptional effects of each single knockout alone were less clear (Kolaczkowska et al., 2008). These multi-knockout transcriptional responses can result in complex influence effects that would not be well-captured by our original neural network, which can learn only additive influences. Indeed, while the additive influence model largely captured one- and two- knockout effects, it did not predict the extent of fluconazole resistance of the three- and four-deletion strains that showed complex positive interactions (Figure 6E). Therefore, we extended this model by adding an extra hidden neuron to mediate influence on Pdr5 from *SNQ2, YBT1, YCF1*, and *YOR1* (see Methods). Adding this ‘influence mediator’ to the network captured the observed fluconazole resistance for the surprisingly-resistant three- and four-knockout strains (Figure 6F-G). We could not capture these effects by similarly re-training the original neural network on only fluconazole resistance data (Figure S9 B), indicating that the performance improvement depended on modeling non-additive influence effects.

Hypothesizing that the modeled complex influence effects reflect (at least in part) an underlying non-additive transcriptional response, we used qRT-PCR to measure *PDR5* mRNA levels in two double-knockout strains – *snq2∆yor1∆*, two transporters localized in the plasma membrane; and *ybt1∆ycf1∆*, two transporters localized in the vacuole – as well as the hyper-resistant quadruple knockout (*snq2∆yor1∆ybt1∆ycf1∆*). Based on modeling results, we expected *snq2∆yor1∆* and *ybt1∆ycf1∆* to have only modest increases in Pdr5 activity relative to the wild-type (1.27× and 1.20×, respectively), and *snq2∆yor1∆ybt1∆ycf1∆* to have a strong increase (2.69×) (Figure 6H). Similar effects were expected when considering only ‘indirect’ influences from the hidden mediating factor (1.18×, 1.06×, 2.23× for *snq2∆yor1∆*, *ybt1∆ycf1∆*, and *snq2∆yor1∆ybt1∆ycf1∆,* respectively) (Figure 6H). Using qRT-PCR, we found *snq2∆yor1∆* to have a ~1.3× increased *PDR5* mRNA level relative to wild-type. Although this was numerically consistent with the previously-reported ~1.5× increase for *snq2∆yor1∆* (Kolaczkowska et al., 2008), the change did not reach statistical significance in our hands (p = 0.27, Figure 6H) and the previous report did not contain a statistical test. No evidence of increased *PDR5* mRNA levels in the *ybt1∆ycf1∆* strainwas observed(0.9 fold expression, p = 0.69, Figure 6H). More persuasively, we found that *PDR5* mRNA levels were significantly higher in *snq2∆yor1∆ybt1∆ycf1∆* than in wild type (2.1× increase; *p* = 0.032; Figure 6H). Overall, these relative expression patterns were consistent with the relative activity expected from the neural network model.

Taken together, these experiments provide support for two different mechanisms whereby gene deletions can relieve Pdr5 inhibition, one in which deletion of *snq2∆* and *yor1∆* can relieve direct physical inhibition of Pdr5, and another in which deletion of four genes relieves *PDR5* expression.

**Discussion**

Here we presented an implementation of XGA, an extension of pairwise genetic interaction studies using approaches like SGA, to demonstrate in-depth profiling of complex multi-gene variant effects. XGA of 16 ABC transporters uncovered complex genetic phenomena that were not evident from one- and two- gene knockout effects, and the resulting data was used to learn functional system models of ABC transporters. Broadly, we illustrate the potential for carrying out XGA in other sets of functionally-related genes.

The demonstrated cross-based XGA approach can be adapted with multi-variant strains in many model organisms. An XGA of GPCR signaling, for example, could be performed using an existing yeast mutant with 16 GCPR pathway-related genes deleted (Shaw et al., 2018). With CRISPR, variation in yeast can be engineered at up to five loci with a single transformation (Jakočiūnas et al., 2015), facilitating construction of multi-variant strains. Simultaneous variant engineering at 3 – 6 loci has also been described in other model organisms: e.g. mouse (Wang et al., 2013), zebrafish (Jao et al., 2013), *C. elegans* (Xu et al., 2016), and *Arabidopsis* (Zhang et al., 2016).

XGA may also be performed by directly engineering trackable multi-allele diversity into a population, as technical advances in this area continue to be made (Wong et al., 2016; Zeitoun et al., 2017). For example, pooled approaches to profile two-gene combinatorial mutants in both yeast (Díaz-Mejía et al., 2018) and human cells (Horlbeck et al., 2018; Najm et al., 2017; Shen et al., 2017; Wong et al., 2016) may be expanded to allow for higher combinatorial complexity at large scale. In human cells, methods developed for single-cell sequencing may also be adapted for directly-engineered XGA to permit richer phenotyping. Indeed, it is already possible to profile single-cell RNA levels (Dixit et al., 2016) and chromatin state (Rubin et al., 2018) following combinatorial gene disruptions, but this has not yet been applied to large numbers of multi-gene variants. Single-cell barcoding methods may also permit more in-depth sampling of the vast combinatorial space – here, we sampled ~8% of 65,536 knockout combinations at 16 genes, which grows to ~107 combinations if heterozygotes were also considered.

Many future expansions can be envisioned even for the demonstrated XGA study of yeast ABC transporters. For example, the constructed engineered population provides a resource that readily allows XGA of ABC transporters in other compounds. Additional genes may also be deleted in each pool, thereby expanding XGA targets (e.g. to study a Pdr1-dependent transcriptional response). Higher-content phenotyping could also provide a richer profile of the cellular response to ABC transporter perturbation (Khakhina et al., 2015). For example, while no evidence of a transcriptional response was found when deleting *YBT1* and *YCF1*, both genes have been shown to catalyze the movement of substrates from the cytosol to the vacuole (Sousa et al., 2015), and might act to competitively sequester Pdr5 efflux products. The use of fluorescence-conjugated drugs (Benhamou et al., 2017), coupled with the ability to image and genotype multi-knockout strains (Emanuel et al., 2017) could be used to test for this at scale.

Interestingly, there is also evidence for complex influence between ABC transporters in mammals. For example, ABCC3 increases in expression when ABCC2 is disrupted in Dubin-Johnson Syndrome (Donner and Keppler, 2001; König et al., 1999), and ABCG5/ABCG8 both increase in expression when ABCG2 (a protein that confers breast cancer xenobiotic resistance in humans) is knocked out in mice (Huls et al., 2008). An analogous XGA of human ABC transporters could yield better understanding of their roles not only in the drug response and chemotherapeutic resistance, but in numerous diseases (Huls et al., 2008).

We showed that computational modeling of the complex genetic landscape can be used to construct and objectively evaluate genotype-to-phenotype models. This is in contrast to insight extracted from automated epistasis analysis, which often aims to mine genetic interactions for evidence of biologically-ordered pairs rather than construct explanatory models of all observed interactions (Angeles-Albores et al., 2018; Boettcher et al., 2018; St Onge et al., 2007). In addition to ordered pairs, many genetic interactions can also arise from the effects of genes on unobserved ‘hidden’ variables that mediate the measured phenotype (Otwinowski et al., 2018). For example, our neural network could model complex negative interactions in terms of the influence of ABC transporters on an indirectly-observed drug efflux activity. In the case of valinomycin, we could greatly improve models by hypothesizing the influence of multiple ABC transporters via an unknown valinomycin resistance factor. Neural networks can learn directed non-linear relationships amongst potentially vast numbers of biologically-motivated hidden variables (Ma et al., 2018). While even biologically-motivated neural network models can easily grow complex and challenging to interpret, here we showed that they can be simply constructed using directly-interpretable parameters (e.g. ‘activity’, ‘influence’, and ‘efflux’), and then iteratively expanded as needed.

In summary, we showed that measuring and modeling high-order genotype-to-phenotype relationships via XGA can help to functionally dissect and understand a complex system.

**Star Methods**

**Key Resource Table**

|  |  |  |
| --- | --- | --- |
| **Reagent or Resource** | **Source** | **Identifier** |
| **Experimental Models: Organisms/Strains** | | |
| RY0622 | Suzuki et al., 2011 | N/A |
| RY0146 | Suzuki et al., 2011 | N/A |
| RY0566 | Suzuki et al., 2011 | N/A |
| RY0148 | Suzuki et al., 2011 | N/A |
| Barcoded RY0148 pool | This paper | N/A |
|  | | |
| **Chemicals, Peptides, and Recombinant Proteins** | | |
| fluconazole | Sigma-Aldrich | F8929 |
| ketoconazole | Sigma-Aldrich | K1003 |
| miconazole | Sigma-Aldrich | 1443409 |
| itraconazole | Sigma-Aldrich | I6657 |
| beauvericin | Sigma-Aldrich | B7510 |
| tamoxifen | Sigma-Aldrich | T5648 |
| benomyl | Sigma-Aldrich | 45339 |
| cycloheximide | Sigma-Aldrich | C1988 |
| methotrexate | Sigma-Aldrich | M9929 |
| camptothecin | Sigma-Aldrich | C9911 |
| cisplatin | Sigma-Aldrich | P4394 |
| bisantrene | Sigma-Aldrich | B4563 |
| mitoxantrone | Sigma-Aldrich | 6545 |
| colchicine | Sigma-Aldrich | 9754 |
| imatinib | Sigma-Aldrich | 270784 |
| valinomycin | Sigma-Aldrich | V3639 |
|  | | |
| **Oligonuclides** | | |
| All DNA primers used, see Data S1 | This paper | N/A |
|  | | |
| **Recombinant DNA** | | |
| Plasmid: pSH47 | Euroscarf | P30119 |
| Plasmid: pIS420 | Euroscarf | P30575 |
|  | | |
| **Software and Algorithms** | | |
| Analysis pipeline (written in R) | This paper | https://github.com/a3cel2/xga |

**Contact For Reagent and Resource Sharing**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Frederick P. Roth(fritz.roth@utoronto.ca).

**Experimental Model and Subject Details**

***Saccharomyces cerevisiae* strains**

RY0622 (ABC-16/‘Green Monster’ MAT**a**):

*MAT***a** *adp1Δ snq2Δ ycf1Δ pdr15Δ yor1Δ vmr1Δ pdr11Δ nft1Δ bpt1Δ ybt1Δ pdr18Δ yol075cΔ aus1Δ pdr5Δ pdr10Δ pdr12Δ can1Δ::GMToolkit*-**a** (*CMVpr-rtTA KanMX4 STE2pr-Sp-his5*) *his3Δ1 leu2Δ0 ura3Δ0 met15Δ0*

RY0146 (“Toolkit-a” strain):

*MAT****a*** *lyp1Δ his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 can1Δ::GMToolkit-a (CMVpr-rtTA KanMX4 STE2pr-Sp-his5)*

RY0566 (“Toolkit-a” strain with tet-inducible GFP-URA3):

*MAT****a*** *lyp1Δ his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 can1Δ::GMToolkit-a (CMVpr-rtTA KanMX4 STE2pr-Sp-his5) hoΔ::tetO2-GFP-URA3*

RY0148 (“Toolkit-**α”** strain):

*MAT****α*** *lyp1Δ his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 can1Δ::GMToolkit-α (CMVpr-rtTA NatMX4 STE3pr-LEU2)*

Barcoded RY0148 pool:

*MAT****α*** *lyp1Δ his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 can1Δ::GMToolkit-α (CMVpr-rtTA NatMX4 STE3pr-LEU2) ho*∆*::loxP UP-tag HphMX4 DN-tag lox2272*

**Methods Details**

**Creating the barcoder plasmid**

We added a barcoder locus flanked by *loxP* and *lox2272* into a pSH47 plasmid backbone expressing GAL1pr-CRE. This barcoder locus consisted of a random 25bp DNA sequence (‘UP tag’) in between two common primer regions (‘US1’ and ‘US2’), followed by a HphMX4 cassette, and another random 25bp DNA sequence (‘DN tag’) in between two common primer regions (‘DS1’ and ‘DS2’).

First, a barcoded HphMX4 construct was created. HphMX4 was amplified from a pIS420 plasmid using the STEP1F and STEP1R primers containing HphMX4 homology and US2/DS1 overhangs (Data S1). The PCR program used for this step was 98°C for 30sec; 25 cycles of 98°C for 10sec, 59°C for 10sec, 72°C for 60sec; 72°C for 5min; 4°C forever. These PCR products were purified using a Qiagen Qiaspin kit and confirmed using 2% gel electrophoresis. To the resulting products, the STEP2F and STEP2R primers were used to add the random barcodes and US1/DS2 regions with the following PCR program: 98°C for 30sec; 25 cycles of 98°C for 10sec, 68°C for 10sec, 72°C for 60sec; 72°C for 5min; 4°C forever. These resulting products were again purified using a Qiagen Qiaspin kit and ~1.5-1.6kb products were confirmed using 2% gel electrophoresis. To add *loxP*/*lox2272* sites, PCR was performed with the STEP2 products using the SacI-loxP-HphMX4-Barcode-F / SacI-lox2272-HphMX4-Barcode-R primers. The PCR program used for this step was: 98°C for 30sec; 26 cycles of 98°C for 15sec, 64°C for 20sec, 72°C for 65sec; 72°C for 5min; 4°C forever. The resulting PCR products were purified using a Qiagen Qiaspin Kit, and ~1950bp products were confirmed using 2% gel electrophoresis.

To confirm correct synthesis of the barcoded HphMX4 construct, two PCR reactions were performed on the resulting products. The first PCR reaction was performed with the SacI Reamp F/US2 primer pairs, and the second was performed using DS1/SacI Reamp R primer pairs. The PCR program used for both of these reactions was: 98°C for 30sec; 25 cycles of 98°C for 10sec, 59°C for 15sec, 72°C for 30sec; 72°C for 5min; 4°C forever. Expected sizes (~132bp, 137bp) were confirmed using 4% gel electrophoresis. All above PCR reactions were performed using High Fidelity Phusion Master Mix (NEB).

To prepare for cloning of the barcoder locus, pSH47 was digested with SacI using 100μl of 250ng/μl pSH47, 100μl NEB Buffer 4, 10μl BSA, 10μl SacI-HF in 1ml sterile water. 100μl of this mixture was incubated at 37°C for two hours, and inactivated by incubation at 65°C for 20min. Digest products were purified using a Qiagen Qiaspin kit, and confirmed using 0.8% gel electrophoresis.

**Generating a barcoder strain**

A linear URA3 cassette flanked by *loxP* and *lox2272* sites and homology to the *HO* gene was amplified from purified pIS418 with the 5'HO-loxP-URA and URA-lox2272-3'HO primers using the following PCR program: 98°C for 30sec; 25 cycles of 98°C for 10sec, 60°C for 10sec, 72°C for 70sec; 72°C for 5min; 4°C forever. This PCR reaction was performed using High Fidelity Phusion Master Mix (NEB)and was purified using a Qiagen Qiaspin kit. This cassette was integrated into the *HO* locus of the RY0148 strain through transformation to serve as the ‘landing pad’ for barcode integration using an EZ transformation kit. Transformants selected for growth in SC –Ura plates, and were later verified to exhibit no growth in 5-FOA. A transformant was selected to confirm *HO* locus integration using three PCR reactions with the following primer pairs: 5'HO-URAreamp + midURA-5'; 5'HO-URAreamp + midURA-3'; 5'HO-URAreamp + 3'HO-URAreamp. All PCR reactions were performed using High Fidelity Phusion Master Mix (NEB) with the following program: 98°C for 30sec; 25 cycles of 98°C for 10sec, 50°C for 10sec, 72°C for 70sec; 72°C for 5min; 4°C forever. Expected PCR product size was confirmed using 2% gel electrophoresis.

The *HO*::*loxP*-*URA3*-*lox2272* integrant strain was then transformed with a mixture of digested pSH47 and purified PCR products to enable in-yeast-assembly (Gibson et al., 2009). Transformation was carried out using a previously established protocol (Gietz and Schiestl, 2007), with a ~1:6 mixture of digested pSH47:HphMX4 barcode cassette (~12μg digested pSH47 and 15μg cassette). Transformants were grown at 30°C in YPG +HygroB plates for 3 days, allowing both selection of successful transformants and Gal1p-Cre induction. These cells were then scraped and grown overnight in 5-FOA plates to select against non-recombinant strains and strains containing the barcoder plasmids.

Twenty colonies were confirmed to have barcode integration using PCR and Sanger sequencing. Lysates were made by mixing a sample of each colony with 2μl Sterile DNA Free Water, 2μl 0.2M pH 7.4 Sodium Phosphate Buffer, 0.5 μl 5U/μl Zymoresearch zymolyase and incubated at 37°C for 25min and 95°C for 10 min, and stopped by adding 125μl of sterile DNA-free water. To each lysed colony, two sets of primer pairs to verify the strain barcode-specific UP and DN tag - US2 and a sequence complementary to 5’ of the *HO* gene (5’HO); DS1 and a sequence complementary to the 3’ of the *HO* gene (3’HO), using the following program: 98°C for 30sec; 25 cycles of 98°C for 10sec, 59°C for 15sec, 72°C for 30sec; 72°C for 5min; 4°C forever. PCR reactions were performed using High Fidelity Phusion Master Mix (NEB) and analyzed using gel electrophoresis. EXOSAP-IT purification (Thermo Fisher Scientific, Cat.78201) was performed on the PCR products, and they were Sanger sequenced with the 5’HO seq and 3’HO seq primers to confirm the correct barcode construct.

**Creating a ‘gold standard’ genotyped set**

To create a ‘Gold Standard’ genotyped set, 40 progeny strains (19 MAT**a** and 21 MAT**α**) were subject to individual strain genotyping. For these 40 strains, and for an RY0148 isolate, the strain-specific UP and DN tags were also PCR amplified using two sets of primers and subject to Sanger sequencing as above.

To genotype each strain at the 16 ABC transporter loci, two PCR reactions were performed for each locus - one to determine the presence of a GFP integration cassette, and another to determine the presence of the wild type gene, as previously described (Suzuki et al., 2011). For the cassette confirmation reactions, locus–specific PCR primers from the 5′ flanking sequences of each gene were paired with a common primer complementary to the *GFP* cassette (Data S2). Gene presence confirmation primers were designed individually for each gene (Data S2). PCR reactions were performed with a Platinum HiFi mix using the following program: 94°C for 2min; 34 cycles of 94°C for 30sec, 55°C for 30sec, 68°C for 60sec; 68°C for 10min; 4°C forever. PCR products were analyzed using gel electrophoresis.

**Generating barcoded random knockout progeny**

Mating, sporulation, and haploid selection was performed between the RY0622 ‘Green Monster’ strain (MAT**a**) and the barcoded RY0148 pool (MAT**α**) as previously described (Suzuki et al., 2011), selecting for MAT**a** and MAT**α** progeny separately. The two pools were then grown in YPD +HygroB to select for barcoded haploids. The SC–Leu pool was further grown in SC–Ura to select against barcoder strain parents that may have escaped diploid selection. Using a QPix colony picker, 5,461 MAT**a**  and 5,461 MAT**α** colonies were picked onto 384 well plates. In addition, 299 known positions in both the MAT**a** and MAT**α** arrayed collections consisted of known strains – either one of 40 ‘Gold Standard’ genotyped strains, RY0148, or RY0622 – to act as genotyping controls (Data S2).

**Pooled strain genotyping**

A previously-developed Row-Column-Plate (RCP)-PCR protocol (Yachie et al., 2016) was adapted in order to perform *en-masse* genotyping of the random knockout progeny using high throughput sequencing. This protocol first uniquely tags PCR products originating from the same well on a given plate, by the use of a 5’ tag encoding the well row (R) in forward primers, a 5’ tag encoding the well column (C) in the reverse primers (Yachie et al., 2016). Additionally, these primers contain a linker sequence (PS1 or PS2) which primes a second reaction encoding the plate of origin (Data S2).

For each well in the collection, lysates were made on a new set of plates. 4 μl of overnight yeast culture was mixed with 8 μL 0.2 M sodium phosphate buffer (pH 7.4), 4 μl DNA free dH2O, 0.05 μl 5 U/μl Zymoresearch zymolyase and incubated at 37 °C for 35 minutes. 64 μl DNA free dH2O was added to each well to prepare PCR template.

Four ‘Row-Column’ PCR reactions were performed on the lysates with the following primer pairs: PS1+R+U1 and PS2+C+U2 to amplify DNA barcodes encoding the UP tags for each gene deletion; PS1+R+D1 and PS2+C+D2 to amplify the deletion-specific DN tags; PS1+R+US1 and PS2+C+US2 to amplify the strain-specific UP tag; PS1+R+DS1 and PS2+C+DS2 to amplify the strain-specific DN tag (Data S2). PCR reactions were performed with 2 μl of lysed colonies using a Hydrocycler with the following program: 95 °C for 5 min; 23 cycles of 95 °C for 60 sec, 57 °C for 35 sec, 72 °C for 45 sec; 72 °C for 2 min; 4 °C forever. Row-Column PCR products from each plate were pooled and size was verified on a 4% agarose gel. PCR products from each plate were combined, and Illumina adapters containing plate tags were added using an additional PCR reaction as previously described (Yachie et al., 2016). A pair of PXX\_PE1.0 and PYY\_PE2.0 primers (Data S2) were added to 3-6 μl pooled products (calibrated to ~150 ng) from each plate to encode the plate of origin, and were amplified using the following PCR program: 98 °C for 30 sec; 15 cycles of 98 °C for 10 sec, 59 °C for 15 sec, 72 °C for 40 sec; 72 °C for 2 min; 4 °C forever. All PCR reactions above were performed using High Fidelity Phusion Master Mix (NEB).

Expected product size from the plate tags was confirmed on 4% agarose gel. PCR products were purified using a Qiagen MinElute Gel Extraction kit (Cat. No.28604), and qPCR was performed on all plate tag PCR products using a light cycler and KAPA Illumina sequencing quantification kit. qPCR results were used to pool approximately equal amounts of all samples, and 100 μl of this multiplexed sample were run on a 4% gel. Products of the desired size (260-290 bp) were isolated from each lane, purified using a Qiagen MinElute Gel Extraction kit, and another qPCR was run on the purified sample.

**Analysis of pooled strain genotyping data**

Pooled strain genotyping PCR products were sequenced using an Illumina HiSeq 2000, and the reads were demultiplexed into individual samples corresponding to a plate and well of origin using a Perl script.

For each sample, a genotype calling pipeline determined the strain-specific tag sequences and genotype from the reads. The parameters of this pipeline were trained based on known reference strains. Cross-validated accuracy for each gene is reported in Figure S2A.

UP or DN tag identity and a corresponding genotype was successfully determined for 7,195 samples. For 7,030 samples, the UP or DN tag was unique, and for 165 samples, both the UP and DN tag sequences were redundant with another sample where the called genotype was isogenic or highly similar (≤ 2 differences), indicating the presence of a single strain in multiple wells. If a single strain existed in multiple wells, genotyping data from a single well was assigned to the barcode at random.

**Refining Genotype for Putative Wild-Type Strains**

For 131 MAT**α** and 73 MAT**a** strains, pooled sequencing analysis had called the genotype as wild-type. Many of these strains were isolated and tested for the presence of one or more gene knockout cassettes by growth in SC–Ura. Out of 96 MAT**α** strains, 74 exhibited no detectable growth in SC–Ura (indicating the absence of any knockout cassettes), and likely arose from remaining barcoder parents which had escaped a previous SC–Ura selection step. The genotypes for the these 74 strains were kept as is, while the other 23 strains, as well as 46 untested strains were discarded from the analysis. Out of 45 MAT**a** strains, all exhibited growth in SC–Ura (indicating a knockout cassette at one or more loci). Individual genotyping was performed for these MAT**a** strains, and was successful for 40 of 45 strains, confirming the lack of true wild types. These strains had their genotype corrected (Data S2). The 5 unsuccessfully genotyped strains, as well as 28 untested strains were discarded from analysis. When calculating linkage and distribution of gene knockouts (Figure S2C), the wild-type MAT**α** strains were excluded from analysis as they were likely parental strains rather than progeny arising from mating.

**Estimating genotyping accuracy by knockout distribution**

To lend independent support to the genotyping accuracy determined by gold standard strains, an alternate method based on the distribution of knockouts in the population was used. Since *en masse* genotyping associates barcode sequences with ABC transporter knockouts, the absence of a given barcode implies either a wild-type genotype at that locus or a failure in amplification, sequencing, or calling. Conversely, cases where a wild-type is called as a mutant are expected to be comparably rare. Excess wild-type calls lead to a reduction in the average number of knockouts in the pool, and can be used to estimate genotyping accuracy. The average number of knockouts in the pool was 7.0, lower than the 8 expected with perfect genotyping. If wild-type to mutant miscalls are negligible, this number is most likely with an ‘asymmetric’ genotyping accuracy of 93.8%, compared to the 93.2% estimated by comparison to gold standards (Figure S2B).

**Individual liquid growth profiling**

To measure individual strain growth, 100μl of starting culture at 0.0625 OD600nm was grown in a 96 well-plate in a temperature-controlled shaking spectrophotometer (Tecan GENios microplate reader). Growing cultures were shaken at 800 rpm at 30°C and OD600nm of each well was measured every 15 min.

**Population growth profiling by high-throughput sequencing**

Progeny with at least one mapped strain-specific barcode (Data S2) were combined into two separate liquid YPD + 15% glycerol pools separated by mating type, and kept at −80°C. Samples from the original YPD + glycerol pool were thawn and added to the appropriate drug or solvent containing medium at a final concentration of 0.0625 OD600 nm in 10ml. In addition, a ‘0 generation’ sample was immediately harvested from the YPD + glycerol pool and processed for DNA extraction and sequencing. After growth to approximately 2 OD600 nm (~5 generations), cells were collected and processed for sequencing, and a small aliquot was diluted in fresh media (at a final concentration of 0.0625 OD600 nm in 10ml) in presence of drug or solvent to be grown for an additional 5 generations. This process was repeated for a total of 20 generations of growth, with collected samples corresponding to approximately 5, 10, 15, and 20 generations.

Harvested samples were subject to genomic DNA extraction using a YeaStar™ Genomic DNA Kit, quantified using the Invitrogen Quant-IT dsDNA BR Assay kit (Cat No. Q32853), and diluted to a final concentration of 25 ng/μl. 350ng of DNA from each sample was indexed with the following PCR mixture: 20 μl of 2x Platinum PCR SuperMix High Fidelity, 1 μL of 10 μM F primer, and 1 μl of 10 μM R primer. F and R primer pairs were PXX+US1/PYY+US2 and PXX+DS1/PYY+DS2 for the strain-specific UP and DN tag, respectively. PXX and PYY correspond to sequences containing plate-specific Illumina sequencing adapters, as well as tags which were used to demultiplex the samples (See Data S2). PCR products were amplified using the following program: 98 °C for 30 sec; 24 cycles of 98 °C for 10 sec, 60 °C for 10 sec, 72 °C for 1 min; 72 °C for 5 min; 4 °C forever. After indexing, equal volumes of UP-tag and DN-tag PCR products from each pool were run on a 3% agarose gel. The expected 210bp bands were isolated and purified using a Qiagen MinElute Gel Extraction kit. DNA size and purity was confirmed by using the Agilent Bioanalyzer High Sensitity DNA kit. DNA yield was quantified in triplicate using a KAPA SYBR FAST Universal qPCR kit (Cat. No. KK4824). Approximately equal amounts of each sample were combined and sequenced using an Illumina NextSeq 500 High Output v2 kit.

**Sequence data processing**

Paired-end Illumina sequencing data were first de-multiplexed using a custom Python script which searches for an exact match to the tag regions of the PXX and PYY primers within each pair of reads. For each strain in each de-multiplexed sample (corresponding to a combination of mating type, timepoint, and drug), strain identification is attempted. To perform this identification, a search is performed for all barcodes matching the sample mating type. If an exact match is not found, up to two ungapped mismatches are permitted to assign a putative strain identity, which is then accepted if there are at least 2 additional mismatches separating this identity with the next closest match (e.g. if 2 mismatches are present with the closest match, then the next closest match must have 4 or more mismatches). This process was performed for both the forward and reverse reads (corresponding to the UP and DN tags) for each strain, and potential cases where the putative strain identity differed between tags were discarded. All samples with less than 200,000 reads in either the UP or DN tag were discarded from the analysis.

**Defining a resistance metric**

Following processing of the sequence data, a count was assigned for each strain in a pool under drug sequenced at time (. The counts in each sample were then converted to a frequency by division with the total count for all strains in that sample:

If both an UP and DN tag for a given strain were successfully linked to a genotype, is estimated from the UP and DN counts and averaged, otherwise the available tag is used.

We modeled each strain as growing constantly from an initial abundance () by an exponential growth rate in each drug over time , such that:

Here, is proportional to a desired ‘absolute’ starting abundance metric for each strain (e.g. number of cells), , such that:

Similarly, we also obtain a proportional estimate for by using the frequency at each time point, multiplied by expected relative pool growth at that time compared to time 0 (, since is defined by the number of generations since t = 0)

Therefore, integrating over time can be used to compute a metric that is proportional to the absolute ‘area under the growth curve’ () from timepoints to (the total number of pool generations measured). Here, frequencies between measured timepoints were linearly interpolated:

In the constant exponential growth model, integrating over all time points results in the following relationship with growth rate:

Given the above relationships of the frequency-based metrics with absolute abundance-based metrics, we can substitute with and with . Therefore, the unknown constant relating frequency-based abundance estimates to absolute abundance estimates for each strain ( is not needed to compute growth rate (:

Because and are both known, we then numerically solve for the which satisfies this relationship using the optimize() function in R, setting a minimum of -10 and maximum of 10 for the interval (there is no simple closed-form solution for this equation). is an estimate of the growth rate relative to the pool as a whole (i.e. a strain with perfectly ‘keeps up’ with the pool). In practice, estimates the average relative exponential growth rate from 0 to (e.g. a prolonged lag phase effectively lowers the average exponential growth rate). For linear regression and neural network training, the minimum is set to 1e-10 to avoid numerical errors in the respective algorithms. To obtain the resistance for each strain in each drug (, is divided by growth in the DMSO control (:

We note that experimental uncertainty in the collected generation times *t* can introduce range uncertainty in , such that resistance estimates from the MAT**a** and MAT**α** pool may be highly correlated, but may differ in range in some drugs (Figure S5). To correct for any potential pool-of-origin effects in arising from merging the MAT**a** and MAT**α** populations, we use the line of best fit derived in Figure S5 to rescale estimates from the MAT**a** pool to match the MAT**α** pool before merging. Any small remaining effects are handled by plate-of-origin terms in linear modeling (below).

**Finding complex genetic interactions using a general linear model**

The multiplicative model of genetic interactions (Mani et al., 2008) was applied to the metric. In this model, the expected resistance of a double knockout strain in a given drug () is the product of the resistances of the corresponding single knockout strains:

To express this model in an additive form, we can state this relationship as an exponentiated sum of the log-resistances of the single knockouts - , so that:



We defined a two-gene interaction term as the log-ratio of the observed fitness to the fitness expected by single-gene effects, rather than the traditional linear difference from a multiplicative estimate (a difference-based metric is not readily generalizable to high-order effects).



This interaction term can be added to 2) to express the observed rather predicted double mutant fitness:

When modeling the expected triple mutant fitness, all relevant two-gene interaction terms are added as such:

Similarly, a three gene interaction term is the deviation from the one- and two- gene expectation:

This definition can be extended analogously for interactions of arbitrary complexity, with terms denoting interactions between the corresponding knockouts. Specifically, in each drug we fit a general linear model which aims to predict the fitness of each given its knockout genotype , which consists of a subset of ABC transporter knockouts :

To train this model, is encoded as a set of 16 binary variables, where 0 represents a wild-type and 1 represents a knockout at a given gene. Therefore, to predict phenotype from , the relevant coefficients are added only if the corresponding gene is knocked out, and the coefficients are added only if all the genes in subset are knocked out. For each drug, we fit this model using the glm() function in R, with terms to a chosen level of complexity. This model consists of Equation 7) with an added ‘plate of origin’ term for each strain to correct for any potential batch effects.

To perform the marginal association in Figure S4, we fit a model with only terms (and plate-of-origin effects), and performed stepwise feature elimination (eliminating the gene with the highest p-value at each step) until all included terms had a significance level of . Linear model term significance was tested using the Type III Sums of Squares ANOVA implementation given in the car package in R. The same method was used to perform the marginal association in Figure S3, substituting for .

We expanded this approach to train models containing terms of up to -way complexity using a “stepwise expansion” method. First, we use the marginal association procedure above to initialize the model at . Then, *n* is incremented by 1, and all possible -way interactions between the genes contained in the existing (i.e. ) model are added as additional features (all plate-of-origin effects are re-added at each step). Each term in this proposed “-way” model is tested for significance using Type III Sums of Squares ANOVA, those with *p* > 0.05 are discarded, and the model is updated. This stepwise expansion procedure is repeated until either reaches 5, or the number of unique genes in the model is less than (i.e. there are no more possible interaction terms to search for). After the stepwise expansion procedure is finished, the remaining terms are more rigorously tested for statistical significance by performing stepwise feature elimination (as in the marginal association procedure) until all included terms have a significance level of , where *k* is the number of all possible 1-5 gene combinations amongst the marginally associated genes. Plate-of-origin effects are re-added to the model before stepwise elimination, and are subject to the same significance criteria.

**Defining a neural network system model**

We will define an ‘efflux and influence’ schematic of ABC transporter function which we will later fit as a neural network. First, we normalize resistance data in each drug by dividing with the maximum observed resistance in that drug:

We then model a sigmoidal relationship between drug concentration and normalized resistance:

Here is the concentration of the given drug, and , are an unknown constants which define the dose-response curve (such that yields the expected IC50). In addition, we model each transporter as encoding a resistance factor which acts to additively lower the effective concentration of a drug (for example, by efflux out of the cell):

Here, is the set of ABC transporters present in a genotype: , and is the clearance coefficient of a given ABC transporter for a given drug (i.e. ). Importantly, a dose response curve in this form can be expressed as the activation of a sigmoid neuron, where is collapsed into a single bias term , and are the weights learned as inputs to this neuron from the ABC transporters. As each transporter must act to lower effective drug concentration in this model, we constrain to be non-negative.

We then model influence between ABC transporters. To do this, we first decompose the clearance coefficient of each ABC transporter . That is, each ABC transporter is also given a degree of activity (a value between 0 and 1) which depends on the genotype - . We initially set a common set of influence relationships across drugs, so that the activity variable is modeled as being dependent on genotype , but not the drug . In this extension, is the product of and , a ‘maximal’ efflux/clearance capacity of a given transporter for a given drug ():

We then allow to capture influence. That is, can be influenced by other ABC transporters:

Where are the ‘influences’ from other ABC transporters. While the form of is itself unknown, here we also modeled it as having a sigmoidal form for simplicity:

**Learning a non-linear system model as a neural network**

To create the above model and learn the , , and parameters from our data, we used the keras library in R to construct a neural network of the appropriate form.

We first provide the genotype of each strain as the input layer to the neural network by encoding in binary form. That is, we create an input layer of length 16, where each input value will be either 1 for ABC transporter presence, or 0 for a knockout for each of .

We then provide a second layer of length 16 to keras, where the weights from the input layer to the second layer encode the influence weights from and to each transporter-transporter pair , , and the second layer acts to compute the activity state for each transporter. Specifically, we create a second sigmoid layer of length 16, and connect each transporter in the first layer to each transporter in the second layer, except where , as a transporter cannot inhibit itself in this model. The activity state for each transporter is then computed by the neurons in the second layer in this network from their inbound influence connection , and a learned bias term . Notably, the neural network multiplies each outgoing inhibitory connection by its corresponding genotype value in , such that all outgoing inhibitory weights from transporter are set to 0 if it is knocked out. To analogously set the activation state of each transporter in the second layer to if it is knocked out, each neuron in the second layer is then multiplied element-wise by its corresponding value in using the layer\_multiply() function.

To encode the efflux weights for each transporter-drug pair , we then added another sigmoidal layer of length 16, which was fully connected to the genotype-multiplied second layer. The kernel\_constraint argument was used with this layer to ensure that non-negative parameters are learned. Each neuron in this third layer predicts the normalized resistance to each compound by multiplying the activation state of each transporter with the learned efflux weights to compute the clearance coefficients for each compound-transporter pair, and furthermore learns a bias term which defines the shape of the dose-response curve.

In addition to the above schematic, L1 regularization with coefficient was added to both the weights and the bias term which defines for each transporter, using the kernel\_regularizer and bias\_regularizer parameters in the second layer. Regularization on achieves sparsity in their weights, as it is otherwise possible, for example, to add to a transporter which has no weights, thus learning parameters which are not supported by any phenotypes. Because the clearance coefficient of each gene for each drug is defined by a product , regularization of the bias term acts to keep close to , effectively setting a prior on . This prior on avoids parameterizing by setting a large bias such that baseline , which then allows weights to be added to transporters without affecting phenotype predictions. Thus, regularization of indirectly enforces sparsity in the parameters, as each directly impacts resistance predictions when is not close to . While more complex regularization schemes can potentially impose three separate regularization weights for the terms, the bias on , and the terms, here we found that using a single weight for regularizing both and the bias for without any further regularization to the terms was sufficient for learning a sparse predictive model.

The neural network model was compiled with the mean-squared error (‘mse’) loss function, using the adam optimizer with a learning rate of 0.05. Training was performed for 10,000 epochs, using a batch size of 1,000 and 10% split between training and validation (validation\_split = 0.1). Model initialization and training was repeated 10 times, and the weights to the final model were set to the mean weights learned from these 10 iterations. In addition, standard deviation was calculated between these 10 iterations, and an absolute Z score was computed for each parameter:

Given the non-deterministic nature of the algorithm, we wanted to confidently ensure that non-zero parameters are not a result of stochastic noise, and therefore non-zero weights with < 4 were set to 0.

We searched for an appropriate regularization rate by performing the above training and averaging procedure using a range of rates from to . We first searched 13 intervals between to (Figure S8A-B). After observing high mean-squared error (MSE) and a lack of reproducible parameters at regularization rates below (Figure S8B) and a smaller ‘jump’ in MSE around (Figure S8 A), we searched another 11 intervals between to (Figure S8 A-B). We then chose a regularization rate of , as any rate higher than this resulted in a jump in mean-squared error in both the MAT**a** and MAT**α** pools (Figure S8B), while lowering this rate did not have a clear impact on mean-squared error but increased the number of non-zero parameters (Figure S8A).

After using the training and averaging procedure to learn model weights, we tested each non-zero weight for predictive support. First, we compute the vector of squared residuals in the initial learned model over strains and drugs, given the set of initial non-zero weights :

Then, for each , we set , and compute the squared residuals in the proposed reduced model:

Considering only data where settings to 0 made a predictive difference ( at a numerical tolerance of ), we then compute the paired Mann-Whitney U statistic between and to derive a p-value for , and keep all features with in the final model.

**Extensions to the Neural Network**

For all training in single drugs, learning rate was decreased to 0.01, as we found this to lower variance in parameter values between different initializations.

To extend the neural network for valinomycin (Figure 4D), we added an extra ‘X’ variable to the neural network input to the and set its value to 1 for each strain.

To extend the neural network for fluconazole (Figure 5B), we performed several additional steps. First, for each transporter in the layer, we added a single extra ‘always-present’ neuron () which provides an additional indirect connection between the and layers. Thus, while each transporter retains original influence connections between and , additional to influence connections were added. Each node is then connected only to its corresponding transporter in (i.e. each indirect node can only influence a single transporter). As in the original network, we omitted both direct and indirect connections between a transporter and itself. These additional influence connections, as well as the bias on , were also subject to L1 regularization with rate (more complex models can again set a separate regularization rate here).

As this three-layer network presented additional training challenges compared to the original two-layer model (e.g. more than twice as many potential parameters), we made additional *a priori* modeling decisions for simplification*.* First, as we aimed to simply extend the two-layer results, we restricted to encode only the presence of *PDR5*, *SNQ2*, *YBT1*, *YCF1*,and *YOR1*, and restricted the *A* layer to encode only the presence of *PDR5.* This reduced the number of potential parameters from 545 to 13. Second, while parameters were highly reproducible between most initializations, we noticed that a small number of random initializations resulted in convergence to a different set of parameters with a higher mean-squared error, even at high regularization rates (data not shown). To better ensure robust parameterization we used the median between 10 runs rather than the average to assign weights to the final neural network, and did not employ the filter described above. Here, training with resulted in a similar model as the two-layer network (data not shown). Therefore, we performed a separate ‘three-layer ’ search for this network, searching 13 intervals between to (Figure S10C). For three-layer training, we found that > negatively impacts MSE (Figure S10C), and therefore used a less-restrictive to train the three-layer network in Figure 5B. Learned weights were subject to the same statistical significance test as for the two-layer network.

**Targeted mating and selection to obtain 32 knockouts**

The TWAS21230902 strain (genotyped as *pdr10∆ pdr18∆ pdr5∆ snq2∆ ybt1∆ ycf1∆ yor1∆* by RCP-PCR; Data S2) was subject to individual strain genotyping (Suzuki et al., 2011), which confirmed the expected wild-type and knockout PCR products at each locus. This strain (MAT**α**) was mated with RY0566 (MAT**a**), and was subject to sporulation and MAT**a** haploid selection (Suzuki et al., 2011). Individuals from this cross were arrayed onto 96 well plates, and individually genotyped at *PDR10* and *PDR18*. Strains with no deletions at these genes were further genotyped at *PDR5, SNQ2, YBT1, YCF1,* and *YOR1.* PCR reactions for individual genotyping of these progeny used the Qiagen Multiplex PCR Plus Kit with the following program: 95°C for 5min; 34 cycles of 95°C for 30sec, 57°C for 30sec, 72°C for 30sec; 68°C for 10min; 4°C forever. After analysis of genotyping results, one strain of each genotype combination was chosen to create the 32-strain collection. These chosen 32 strains were again individually genotyped at these 5 loci for validation.

**Analysis of Liquid Growth Data**

Individual strains with 32 knockout combinations at *PDR5, SNQ2, YBT1, YCF1,* and *YOR1* were each grown in fluconazole at concentrations of 1.9, 3.9, 7.8, 15.6, 23.4, 31.2, 35 and 40μM. Each genotype was grown an average of 2.7 times in each concentration (Data SXX). For each growth experiment, a culture was started at 2% DMSO at the same time to act as a solvent control. Each culture was started at an initial cell concentration of 0.0625 OD600. OD600 was measured every 10 minutes using a Tecan plate reader for a minimum of 20 hours. To calculate resistance, we divided the OD measured in the drug by the OD measured in the solvent at the time which the culture first saturated in the solvent. To automatically determine a saturation timepoint, we took the second derivative of the growth curve (using a window size of 4 tecan measurements to calculate the first derivative) and determined the time which it is maximized. Automatically determined saturation times were checked visually. Multiple replicates were averaged to yield the values in Figure S11. To determine the fitted IC50 values in Figure 4D, averaged resistance values were linearly interpolated between measured concentrations.

**MYTH testing of protein-protein interactions**

*PDR5*, *YOR1*, and *SNQ2* were cloned into the L2 AMBV MYTH bait vector to add a Cub-LexA-VP16 MYTH tag as previously described (Snider et al., 2010). A previously-cloned artificial MYTH-tagged bait plasmid was retrieved, and acted as a negative interaction control. NubG-PDR5 (PDR5 prey) and NubI-PDR5 (PDR5 positive interaction control) strains were retrieved from a previously-constructed genomic prey library (Snider et al., 2010). Previously-constructed Ost1p-NubG (negative interaction control) and Ost1p-NubI (positive interaction control) strains were also retrieved. All prey-bait combinations were obtained using individual transformations and selected for growth in SD –Trp (SD –W)(Snider et al., 2010). Colonies of transformed strains were grown in solid medium for 5 days in SD –W, SD –Trp–Ade–His (SD –WAH), SD –WAH +25μM fluconazole + 2% DMSO, SD –WAH +50μM fluconazole + 2% DMSO, and SD –WAH + 2% DMSO.

**PCA testing of protein-protein interactions**

*PDR5*, *YOR1*, and *SNQ2* MAT**a** (mDHFR-F[1,2]-NatMX fusions) and MAT**α**  (mDHFR-F[3]-HphMX fusions) PCA strains were obtained from a previous genome-wide screen(Tarassov et al., 2008). Additional strains acting required to recreated positive and negative controls were also obtained from this screen (Figure S11). Strains were individually mated and diploids were selected on solid YPD supplemented with Hygromycin B and Nourseothricin (YPD +Hyg +Nat). Diploid strains were spotted on solid YPD +Hyg +Nat supplemented with either 2% DMSO, 2% DMSO + 200 μg/mL methotrexate, or 2% DMSO + 200 μg/mL methotrexate + 46.8μM fluconazole. Strains were grown for 72 hours at 30°C.

**Quantitative RT-PCR**

RNA was extracted from cultures growing exponentially in 23.43μM fluconazole using the QIAGEN RNeasy® kit. 1μg of isolate was treated with DNAse and analyzed using an Agilent Bioanalyzer to quantify nucleic acid concentration and verify purity. cDNA synthesis was performed using a combination of oligo-DT and random hexamer primers using the Thermo Scientific™ Maxima™ H Minus First Strand cDNA Synthesis Kit. qPCR on these samples was then performed using a Bioline SensiFAST™ SYBR No-ROX qPCR kit and Ct values were quantified using a CFX machine. cDNA synthesis and qPCR was performed for *PDR5* and *UBC6* (which acted as loading control). Relative expression of *PDR5* in all strains was calculated as . For each strain, values for the cDNA samples were quantified multiple times to assess technical variability ( was measured in triplicate, and was measured in triplicate), and these multiple measurements were averaged before calculating relative expression. qRT-PCR was performed for three individual cultures of each strain in each genetic background. RY0566 was used as the wildtype.

**Data and Software Availability**

R scripts used to perform computational analyses are available at: https://github.com/a3cel2/xga

**Competing Interests**

The authors declare that they have no competing interests.

**Acknowledgements**

We are grateful for helpful comments from [XX] at the outset of this project, to B. Andrews and C. Boone for providing reagents. We gratefully acknowledge support by the Canadian Excellence Research Chairs (CERC) Program and by National Human Genome Research Institute of the National Institutes of Health (NIH/NHGRI) grants CA130266 and HG004756. FPR was also supported by a Canadian Institutes of Health Research Foundation Grant, an NIH/NHGRI Center of Excellence in Genomic Science award (HG004233), by NIH/NHGRI Grant HG001715, and by the One Brave Idea Foundation

**Author Contributions**

F.P.R., Y.S., N.Y., & A.C. conceived of this study. N.Y., M.G., L.M., T.F., P.B., G.S., S.Z., M.N., and V.W. performed experiments, with advice and assistance from A.G.C., J.S., I.S. and J.C.M.. A.C., M.K., and N.Y. analyzed the data, with advice from F.P.R.. A.C., F.P.R, & N.Y. wrote the paper.

**Additional Data Files**

**Additional Data S1.** List of primers used in this study. Includes the primers used to construct the barcoder locus, perform genotyping, carry out RCP-PCR, and for adding index tags to sequencing libraries.

**Additional Data S2.** Genotyping data for the engineered population. Includes a list of control strains used in high-throughput genotyping, initial genotyping results, re-genotyping of putatively wild-type strains, and the final set of genotyping data used.

**Additional Data S3.** Chemicalcompounds used in this study and their concentration in the pooled growth experiments.

**Additional Data S4.** Growth and drug resistance measurements for all strains as inferred from barcode-sequencing readout of pooled growth experiments.

**Additional Data S5.** Summary of single-gene effects and genetic interactions as obtained by generalized linear modeling.

**Additional Data S6.** Previously-known associationsfor the five “frequently-associated” ABC transporters with resistance or sensitivity to the 16 drugs studied.

**Additional Data S7.** Neural network model parameters.

**References**

Angeles-Albores, D., Puckett Robinson, C., Williams, B.A., Wold, B.J., and Sternberg, P.W. (2018). Reconstructing a metazoan genetic pathway with transcriptome-wide epistasis measurements. Proc. Natl. Acad. Sci. U. S. A. *115*, E2930–E2939.

Baeza-Centurion, P., Miñana, B., Schmiedel, J.M., Valcárcel, J., and Lehner, B. (2019). Combinatorial Genetics Reveals a Scaling Law for the Effects of Mutations on Splicing. Cell *176*, 549–563.e23.

Beh, C.T., Cool, L., Phillips, J., and Rine, J. (2001). Overlapping functions of the yeast oxysterol-binding protein homologues. Genetics *157*, 1117–1140.

Benhamou, R.I., Bibi, M., Steinbuch, K.B., Engel, H., Levin, M., Roichman, Y., Berman, J., and Fridman, M. (2017). Real-Time Imaging of the Azole Class of Antifungal Drugs in Live Candida Cells. ACS Chem. Biol. *12*, 1769–1777.

Bloom, J.S., Ehrenreich, I.M., Loo, W.T., Lite, T.-L.V., and Kruglyak, L. (2013). Finding the sources of missing heritability in a yeast cross. Nature *494*, 234–237.

Boettcher, M., Tian, R., Blau, J.A., Markegard, E., Wagner, R.T., Wu, D., Mo, X., Biton, A., Zaitlen, N., Fu, H., et al. (2018). Dual gene activation and knockout screen reveals directional dependencies in genetic networks. Nat. Biotechnol. *36*, 170–178.

Braun, P., Tasan, M., Dreze, M., Barrios-Rodiles, M., Lemmens, I., Yu, H., Sahalie, J.M., Murray, R.R., Roncari, L., de Smet, A.-S., et al. (2009). An experimentally derived confidence score for binary protein-protein interactions. Nat. Methods *6*, 91–97.

Celaj, A., Schlecht, U., Smith, J.D., Xu, W., Suresh, S., Miranda, M., Aparicio, A.M., Proctor, M., Davis, R.W., Roth, F.P., et al. (2017). Quantitative analysis of protein interaction network dynamics in yeast. Mol. Syst. Biol. *13*, 934.

Costanzo, M., VanderSluis, B., Koch, E.N., Baryshnikova, A., Pons, C., Tan, G., Wang, W., Usaj, M.M.M., Hanchard, J., Lee, S.D., et al. (2016). A global genetic interaction network maps a wiring diagram of cellular function. Science (80-. ). *353*.

Díaz-Mejía, J.J., Celaj, A., Mellor, J.C., Coté, A., Balint, A., Ho, B., Bansal, P., Shaeri, F., Gebbia, M., Weile, J., et al. (2018). Mapping DNA damage-dependent genetic interactions in yeast via party mating and barcode fusion genetics. Mol. Syst. Biol. *14*, e7985.

Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C.P., Jerby-Arnon, L., Marjanovic, N.D., Dionne, D., Burks, T., Raychowdhury, R., et al. (2016). Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. Cell *167*, 1853–1866.e17.

Domingo, J., Diss, G., and Lehner, B. (2018). Pairwise and higher-order genetic interactions during the evolution of a tRNA. Nature *558*, 117–121.

Donner, M., and Keppler, D. (2001). Up-regulation of basolateral multidrug resistance protein 3 (Mrp3) in cholestatic rat liver. Hepatology *34*, 351–359.

Emanuel, G., Moffitt, J.R., and Zhuang, X. (2017). High-throughput, image-based screening of pooled genetic-variant libraries. Nat. Methods *14*, 1159–1162.

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

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods *6*, 343–345.

Gietz, R.D., and Schiestl, R.H. (2007). High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat. Protoc. *2*, 31–34.

Haber, J.E., Braberg, H., Wu, Q., Alexander, R., Haase, J., Ryan, C., Lipkin-Moore, Z., Franks-Skiba, K.E., Johnson, T., Shales, M., et al. (2013). Systematic triple-mutant analysis uncovers functional connectivity between pathways involved in chromosome regulation. Cell Rep. *3*, 2168–2178.

Hartman, J.L., Garvik, B., and Hartwell, L. (2001). Principles for the Buffering of Genetic Variation. Science (80-. ). *291*.

Horlbeck, M.A., Xu, A., Wang, M., Bennett, N.K., Park, C.Y., Bogdanoff, D., Adamson, B., Chow, E.D., Kampmann, M., Peterson, T.R., et al. (2018). Mapping the Genetic Landscape of Human Cells. Cell *174*, 953–967.e22.

Huls, M., Brown, C.D.A., Windass, A.S., Sayer, R., van den Heuvel, J.J.M.W., Heemskerk, S., Russel, F.G.M., and Masereeuw, R. (2008). The breast cancer resistance protein transporter ABCG2 is expressed in the human kidney proximal tubule apical membrane. Kidney Int. *73*, 220–225.

Jakočiūnas, T., Bonde, I., Herrgård, M., Harrison, S.J., Kristensen, M., Pedersen, L.E., Jensen, M.K., and Keasling, J.D. (2015). Multiplex metabolic pathway engineering using CRISPR/Cas9 in Saccharomyces cerevisiae. Metab. Eng. *28*, 213–222.

Jao, L.-E., Wente, S.R., and Chen, W. (2013). Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. Proc. Natl. Acad. Sci. *110*, 13904–13909.

Kebschull, J.M., and Zador, A.M. (2018). Cellular barcoding: lineage tracing, screening and beyond. Nat. Methods *15*, 871–879.

Khakhina, S., Johnson, S.S., Manoharlal, R., Russo, S.B., Blugeon, C., Lemoine, S., Sunshine, A.B., Dunham, M.J., Cowart, L.A., Devaux, F., et al. (2015). Control of Plasma Membrane Permeability by ABC Transporters. Eukaryot. Cell *14*, 442–453.

Kolaczkowska, A., Kolaczkowski, M., Goffeau, A., and Moye-Rowley, W.S. (2008). Compensatory activation of the multidrug transporters Pdr5p, Snq2p, and Yor1p by Pdr1p in Saccharomyces cerevisiae. FEBS Lett. *582*, 977–983.

König, J., Rost, D., Cui, Y., and Keppler, D. (1999). Characterization of the human multidrug resistance protein isoform MRP3 localized to the basolateral hepatocyte membrane. Hepatology *29*, 1156–1163.

Kuzmin, E., VanderSluis, B., Wang, W., Tan, G., Deshpande, R., Chen, Y., Usaj, M., Balint, A., Mattiazzi Usaj, M., van Leeuwen, J., et al. (2018). Systematic analysis of complex genetic interactions. Science (80-. ). *360*, eaao1729.

Lee, A.Y., St Onge, R.P., Proctor, M.J., Wallace, I.M., Nile, A.H., Spagnuolo, P.A., Jitkova, Y., Gronda, M., Wu, Y., Kim, M.K., et al. (2014). Mapping the cellular response to small molecules using chemogenomic fitness signatures. Science *344*, 208–211.

Ma, J., Yu, M.K., Fong, S., Ono, K., Sage, E., Demchak, B., Sharan, R., and Ideker, T. (2018). Using deep learning to model the hierarchical structure and function of a cell. Nat. Methods *15*, 290–298.

Mani, R., St Onge, R.P., Hartman, J.L., Giaever, G., and Roth, F.P. (2008). Defining genetic interaction. Proc. Natl. Acad. Sci. U. S. A. *105*, 3461–3466.

Najm, F.J., Strand, C., Donovan, K.F., Hegde, M., Sanson, K.R., Vaimberg, E.W., Sullender, M.E., Hartenian, E., Kalani, Z., Fusi, N., et al. (2017). Orthologous CRISPR–Cas9 enzymes for combinatorial genetic screens. Nat. Biotechnol. *36*, 179–189.

Otwinowski, J., McCandlish, D.M., and Plotkin, J.B. (2018). Inferring the shape of global epistasis. Proc. Natl. Acad. Sci. U. S. A. *115*, E7550–E7558.

Rubin, A.J., Parker, K.R., Satpathy, A.T., Qi, Y., Wu, B., Ong, A.J., Mumbach, M.R., Ji, A.L., Kim, D.S., Cho, S.W., et al. (2018). Coupled Single-Cell CRISPR Screening and Epigenomic Profiling Reveals Causal Gene Regulatory Networks. Cell.

Sarkisyan, K.S., Bolotin, D.A., Meer, M. V., Usmanova, D.R., Mishin, A.S., Sharonov, G. V., Ivankov, D.N., Bozhanova, N.G., Baranov, M.S., Soylemez, O., et al. (2016). Local fitness landscape of the green fluorescent protein. Nature *533*, 397–401.

Shaw, W.M., Yamauchi, H., Mead, J., Gowers, G.F., Oling, D., Larsson, N., Wigglesworth, M., Ladds, G., and Ellis, T. (2018). Engineering a model cell for rational tuning of GPCR signaling. BioRxiv 390559.

Shekhar-Guturja, T., Tebung, W.A., Mount, H., Liu, N., Köhler, J.R., Whiteway, M., and Cowen, L.E. (2016). Beauvericin Potentiates Azole Activity via Inhibition of Multidrug Efflux, Blocks *C. albicans* Morphogenesis, and is Effluxed via Yor1 and Circuitry Controlled by Zcf29. Antimicrob. Agents Chemother. *60*, AAC.01959-16.

Shen, J.P., and Ideker, T. (2018). Synthetic Lethal Networks for Precision Oncology: Promises and Pitfalls. J. Mol. Biol. *430*, 2900–2912.

Shen, J.P., Zhao, D., Sasik, R., Luebeck, J., Birmingham, A., Bojorquez-Gomez, A., Licon, K., Klepper, K., Pekin, D., Beckett, A.N., et al. (2017). Combinatorial CRISPR–Cas9 screens for de novo mapping of genetic interactions. Nat. Methods *14*, 573–576.

Snider, J., Kittanakom, S., Damjanovic, D., Curak, J., Wong, V., and Stagljar, I. (2010). Detecting interactions with membrane proteins using a membrane two-hybrid assay in yeast. Nat. Protoc. *5*, 1281–1293.

Snider, J., Hanif, A., Lee, M.E., Jin, K., Yu, A.R., Graham, C., Chuk, M., Damjanovic, D., Wierzbicka, M., Tang, P., et al. (2013). Mapping the functional yeast ABC transporter interactome. Nat. Chem. Biol. *9*, 565–572.

Sousa, C.A., Hanselaer, S., and Soares, E. V. (2015). ABCC Subfamily Vacuolar Transporters are Involved in Pb (Lead) Detoxification in Saccharomyces cerevisiae. Appl. Biochem. Biotechnol. *175*, 65–74.

St Onge, R.P., Mani, R., Oh, J., Proctor, M., Fung, E., Davis, R.W., Nislow, C., Roth, F.P., and Giaever, G. (2007). Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. Nat. Genet. *39*, 199–206.

Suzuki, Y., St Onge, R.P., Mani, R., King, O.D., Heilbut, A., Labunskyy, V.M., Chen, W., Pham, L., Zhang, L. V, Tong, A.H.Y., et al. (2011). Knocking out multigene redundancies via cycles of sexual assortment and fluorescence selection. Nat. Methods *8*, 159–164.

Takahashi, K., and Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell *126*, 663–676.

Tarassov, K., Messier, V., Landry, C.R., Radinovic, S., Serna Molina, M.M., Shames, I., Malitskaya, Y., Vogel, J., Bussey, H., and Michnick, S.W. (2008). An in vivo map of the yeast protein interactome. Science *320*, 1465–1470.

Taylor, M.B., Ehrenreich, I.M., Rothstein, R., Hu, T., and Mast, J. (2014). Genetic Interactions Involving Five or More Genes Contribute to a Complex Trait in Yeast. PLoS Genet. *10*, e1004324.

Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F., and Jaenisch, R. (2013). One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. Cell *153*, 910–918.

Wang, M., Herrmann, C.J., Simonovic, M., Szklarczyk, D., and von Mering, C. (2015). Version 4.0 of PaxDb: Protein abundance data, integrated across model organisms, tissues, and cell-lines. Proteomics *15*, 3163–3168.

Wieczorke, R., Krampe, S., Weierstall, T., Freidel, K., Hollenberg, C.P., and Boles, E. (1999). Concurrent knock-out of at least 20 transporter genes is required to block uptake of hexoses in Saccharomyces cerevisiae. FEBS Lett. *464*, 123–128.

Wong, A.S.L., Choi, G.C.G., Cui, C.H., Pregernig, G., Milani, P., Adam, M., Perli, S.D., Kazer, S.W., Gaillard, A., Hermann, M., et al. (2016). Multiplexed barcoded CRISPR-Cas9 screening enabled by CombiGEM. Proc. Natl. Acad. Sci. U. S. A. *113*, 2544–2549.

Xu, S., Wang, Z., Kim, K.W., Jin, Y., and Chisholm, A.D. (2016). Targeted Mutagenesis of Duplicated Genes in Caenorhabditis elegans Using CRISPR-Cas9. J. Genet. Genomics *43*, 103–106.

Yachie, N., Petsalaki, E., Mellor, J.C., Weile, J., Jacob, Y., Verby, M., Ozturk, S.B., Li, S., Cote, A.G., Mosca, R., et al. (2016). Pooled-matrix protein interaction screens using Barcode Fusion Genetics. Mol. Syst. Biol. *12*, 863.

Zeitoun, R.I., Pines, G., Grau, W.C., and Gill, R.T. (2017). Quantitative Tracking of Combinatorially Engineered Populations with Multiplexed Binary Assemblies. ACS Synth. Biol. *6*, 619–627.

Zhang, Z., Mao, Y., Ha, S., Liu, W., Botella, J.R., and Zhu, J.-K. (2016). A multiplex CRISPR/Cas9 platform for fast and efficient editing of multiple genes in Arabidopsis. Plant Cell Rep. *35*, 1519–1533.

**Figures**

**Figure 1. Overview of the Engineered Population Profiling Process**

A population is engineered by mating the barcoded wild-type pool with a multi-mutant strain (here, the ABC-16 strain). Barcoded haploid progeny inherit a random combination of knockout (black) and wild-type (white) alleles at multiple loci (16 in this study), and are arrayed from single colonies into 384-well plates. The genotype of each strain in this collection was associated with a DNA barcode *en masse* using a tag-based PCR indexing strategy. Strains with a successfully determined barcode and genotype are pooled together by mating type, and grown under each of 16 drugs, as well as a solvent control (DMSO). High throughput sequencing of strain-specific DNA barcodes at t = 0, 5, 10, 15, and 20 generations of pool growth reconstructs the resistance of each strain to each drug. See also Figures S1 and S2.

**Figure 2. Illustrating a Reproducible 5-gene Resistance Landscape**

**(A)** Comparison of grouped resistance profiles in camptothecin and ketoconazole in MAT**a** and MAT**α** pools. Individual strains were grouped by their genotype at *pdr5∆, snq2∆, ybt1∆, ycf1∆,* and *yor1∆*, as indicated by the legend. See also Figures S3 and S4.

**(B)** Distribution of correlations of grouped resistance profiles between biological replicates (MAT**a** and MAT**α** pools) for all drugs tested.

**(C)** A radial combinatorial signature of benomyl resistance. The graph center represents strains that are wild-type at all 5 loci, while radial segments extending outwards represent strains carrying successively more knockout alleles. Each section is colored by the mean resistance of the corresponding strain group relative to the 5-gene wild-type. Extensions to 1, 2, and 5 total knockouts are illustrated. Color scale extends by +/- 1 standard deviation of all drug resistance values observed in a pool. See also Figure S5.

**(D)**  5-knockout radial signatures are shown for 8 drugs.

**Figure 3. 5-gene Resistance Landscapes Illustrate Complex Drug Resistance Phenomena**

A resistance landscape of 5-gene groups is shown for six drugs. Groups differing by a single additional knockout are connected by lines. Solid lines indicate significant differences in resistance (Bonferroni-adjusted *p* < 0.05, Mann-Whitney *U* test), otherwise dashed lines are used. See also Figure S6.

**Figure 4. Environment-Dependent Complex Genetic Interactions amongst 16 ABC Transporters**

**(A)** All single-gene knockout effects and *X*-way genetic interactions (‘ε terms’) mediating resistance to each compound that passed significance criteria (Bonferroni adjusted *p* < 0.05). Magnitude of genetic effects were determined by generalized linear model, then rescaled for each drug.

**(B)** Illustration of modeled genetic effects in cisplatin, mitoxantrone, and bisantrene. For three selected 5-way genetic interactions, top panels show the distribution of drug resistance phenotypes for grouped strains sharing the 5-gene knockout genotype, and for strains grouped by other relevant genotypes. (In each panel, genotypes are defined only by the five genes of interest, and may vary at other loci.) Differences in median resistance (black lines) between the indicated 4- and 5-gene groups are evaluated with a Mann-Whitney *U* test. The bottom row shows genetic effects extracted from Figure 4A.

**Figure 5. An Intuitive Neural Network Model of Complex Genotype-Phenotype Relationships**

**(A)** Schematic of the neural network trained to infer transporter-drug and transporter-transporter influence relationships from complex genotype-to-phenotype relationships. The 16-transporter genotype (***G***) is given as binary input (1 = presence, 0 = absence for each transporter). The activity of each transporter (***A***) is computed using the set of learned transporter-transporter influence weights (***I***). ***A***is set to zero for transporters that are absent in ***G***. Resistance to each of the 16 tested compounds (***R***) is then computed by the learned transporter-drug efflux weights (***E***), each multiplied by the ***A*** values of the corresponding transporters.

**(B) *I*** and **E** weights learned by the neural network model after training. See also Figures S7A-B and S8.

**(C)** Agreement between measured and neural-network-modeled drug resistance. See also Figures S7C-D and S8.

**(D)** Extending the model of valinomycin resistance improves agreement with measurement. For simplicity, data is shown only for the five ‘frequently-associated’ transporters. The trained neural network weights (top) are shown for the original model (top-left) and one with an extra node in the activity layer to model potential influence on a hypothesized resistance factor (top right). See also Figure S8.

**Figure 6. Deciphering a Complex Fluconazole Resistance Trait**

**(A)** Illustratingthe complex genetics of fluconazole resistance. Top panel - A landscape of normalized fluconazole resistance for five ABC transporter knockouts (as in Figure 3).Bottom panel - Significant single-knockout effects and *x-*gene genetic interactions mediating fluconazole resistance (as in Figure 4).

**(B)** Extending the model of fluconazole resistance, by adding an extra node between the ***G*** and ***A*** layer to model non-additive influence of four transporters on Pdr5 activity, improves agreement with measurement. Limitation to frequently-associated five transporters and the method for calculating correlation values are as described in Figure 5D.

**(C)** Close agreement between measures of fluconazole resistance derived by *en masse* analysisof pooled strains with individually-measured resistance (IC50) values in independently-constructed strains.

**(D)** The synergistic effect of deleting four ABC transporters on Pdr5 activity is primarily explained by an indirect influence on *PDR5* transcript levels. Gray bars represent *PDR5* expression in the specified genotype, relative to that of wild-type, with error bars indicating standard error (n=3). Significance was assessed by *t*-test. Colored bars show model-inferred Pdr5 activity (see Figure 5B, top-right panel) for each genotype, relative to that of wild-type, considering all (red) or only the indirect (orange) influences.

**(E)** Interactionevidence supporting the plausibility of direct repression of Pdr5 by Snq2 and Yor1. This study confirmed all previously known-interactions shown (Snider et al., 2013; Tarassov et al., 2008), and revealed a novel Pdr5-Yor1 protein interaction (Figure S11, S12). Direct influence of Yor1 on Pdr5 activity was predicted by both original and extended neural network models for fluconazole (Figure 5B). Influences from the neural network model (Figure 4B) are shown here with red edges. Whole-organism protein levels (Wang et al., 2015) are indicated by node size.

**Figure S1 Creating a Pool of Uniquely Barcoded Strains (‘Barcoder Pool’), Related to Figure 1**

**A** Engineering a barcoder pool cassette. An HphMX4 cassette was amplified from pIS420, with overhangs adding the US2 and DS1 sites. A second PCR reaction was performed to add 25 random base pairs for use as UP and DN tags, as well as two constant US1 and DS2 regions. A third PCR reaction adds *loxP*/*lox2272* sites, and homology to the pSH47 SacI site.

**B** Transforming a pool of barcoder parents. RY0148 was modified to add a *loxP-URA3-lox2272* site and was co-transformed with the barcoder pool cassette and SacI-digested pSH47 to enable reconstitution of a pSH47-based barcoder plasmid construct through in-yeast assembly. Transformants were selected by growth in YPG +Hyg for 3 days to allow for both selection of successful in-yeast assembly products, as well as induction of Cre to enable recombination and replacement of URA3 with the barcoder pool cassette. Loss of URA3 through Cre-enabled recombination is selected by subsequent growth in 5-FOA.

**Figure S2 Analysis of Pool Genotyping Quality, Related to Figure 1**

**A** Expected genotyping accuracy at the 16 ABC transporters surveyed. Accuracy was estimated by evaluating the performance of the RCP-PCR genotyping protocol on a set of ‘gold standard’ strains (Methods, Data S2).

**B** Distribution of knockouts in the combined MAT**a** and MAT**α** pools. The observed number of strains with a given number of knockouts are indicated in gray. The expected number of strains with a given number of knockouts at 93.8% genotyping accuracy under a random assortment model are indicated in black.

**C** Tests of gene linkage within the MAT**a** pools (upper triangle) and MAT**α** pools (lower triangle). The Pearson correlation coefficient of the corresponding genotype pairs are indicated on the right. Pairs without significant correlation (Bonferroni-corrected *p* value ≥ 0.05) are shaded in gray. Three pairs of unlinked genes– *YOR1*-*YCF1*, *YOR1*-*BPT1*, and *SNQ2*-*PDR5* – exhibited weak but significant negative correlation in the appearance of KO genotypes (-0.04 ≥ r ≥ -0.08). This effect may have arisen via negative genetic interactions conferring lower growth for the corresponding double-knockout genotypes during the sporulation, haploid selection, or automated colony picking steps.

**Figure S3 Marginal Single-Gene Knockout Effects in the Pool, Related to Figure 2**

**A** A linear model was used to formally identify significant gene knockout effects mediating resistance (growth in drug relative to growth in drug) for each of the tested drugs. Linear model terms which were significant (Bonferroni adjusted *p* < 0.05) in both MAT**a** and MAT**α** pools for their given drug are colored according the legend on the left. Other terms are colored in gray.

**B** As in A, showing knockout effects that mediate growth

**Figure S4 Reproducibility of Grouped Genotype Resistance, Related to Figure 2**

As in Figure 2A, showing all 16 drugs

**Figure S5 Radial Combinatorial Signatures in Additional Drugs, Related to Figure 2**

As in Figure 2C, showing 5 additional drugs

**Figure S6 Resistance Landscapes for all Drugs, Related to Figure 2**

As in Figure 2D, showing all 16 drugs

**Figure S7 Neural Network Construction and Evaluation, Related to Figure 4**

**A** Number of reproducible network parameters (*Z* > 4 estimated from 10 iterations, Methods) as a function of the regularization rate ***λ***. 13 intervals are plotted from 10-6 to 100 (left), and 11 intervals are plotted from 10-4 to 10-3 (right). Values between intervals are linearly interpolated.

**B** As in S7A, showing the overall mean-squared error of the neural network.

**C** Comparing the normalized resistance of each strain measured by engineered population profiling to resistances modeled by the neural network. Results are shown when the network is trained on either the MAT**a** or MAT**α** population, and then tested on either the MAT**a** or MAT**α** population.

**D** Comparing the learned network weights when the network is trained on either the MAT**a** or MAT**α** population separately.

**Figure S8 Neural Network Performance for Single Drugs, Related to Figure 4**

As in Figure 4D, showing neural network performance for each drug (considering 5-gene groups)

**Figure S9 Neural Network Training with Single-Drug Data, Related to Figures 4 and 5**

**A** As in Figure 4D and Figure S8. Performance is shown for the original network trained only on valinomycin data. No substantial improvement over the original neural network is evident.

**B** As in A, showing the original neural network trained only on fluconazole data. No substantial predictive improvement over the original neural network is evident for *PDR5+* groups.

**C** As in Figure S7B, showing the mean-squared error of the ‘indirect influence’ fluconazole neural network trained in Figure 5C (right panel) as a function of the regularization rate ***λ*.** 13 intervals are plotted from 10-6 to 100.

**Figure S10 Comparing Single-Strain Fluconazole Resistance to 5-gene Groups, Related to Figure 5**

Fluconazole resistance of individual strains containing 32 knockout combinations at *pdr5∆*, *snq2∆*, *ybt1∆*, *ycf1∆*, and *yor1∆* are compared to the normalized resistance of the grouped pool data (Figure 5A). Growth of individual strains was measured at 1.9, 3.9, 7.8, 15.6, 23.4, 31.2, 35, and 40μM of fluconazole. The pool was grown at 23.4μM.

**Figure S11 Measuring Protein-Protein interactions between Pdr5, Snq2, and Yor1 using mDHFR PCA, Related to Figure 5**

*PDR5*, *YOR1*, and *SNQ2* MAT**a** (mDHFR-F[1,2]-NatMX fusions) and MAT**α** (mDHFR-F[3]-HphMX fusions) PCA strains were obtained from a previous genome-wide screen (Tarassov et al., 2008). Strains were individually mated to obtain the indicated diploids. Diploid strains were spotted on YPD containing either DMSO, DMSO + methotrexate (MTX), or DMSO + MTX + 46.8μM fluconazole. MTX selects for successful reconstruction of mDHFR from the F[1,2] and F[3] fragments via a protein-protein interaction. Link-F[1,2]/ Link-F[3] is a diploid strain which tests against interaction of the universal linker regions when fused to the mDHFR fragments. Zip-F[1,2]/ Zip-F[3] is a diploid strain which tests for interaction between two leucine Zipper sequences fused to the mDHFR fragments. Strains were grown for 3 days at 30°C.

**Figure S12** **Measuring Protein-Protein Interactions of Pdr5 with Snq2 and Yor1 using MYTH, Related to Figure 5**

NubG-PDR5, NubI-PDR5, Ost1-NubG, and Ost1-NubI strains were retrieved from a previously constructed genomic prey library25 and were each transformed with plasmids containing clones of *PDR5*, *YOR1*, *SNQ2*, or an artificial bait fused to Cub. NubI fusions are expected to spontaneously reconstitute ubiquitin with Cub, while NubG fusions are expected to require a protein-protein interaction for reconstitution. Ost1 is a component of the oligosaccharyltransferase complex localized to the endoplasmic reticulum membrane and is not expected to interact with any baits tested. Colonies of transformed strains were spotted on SD –Trp (SD –W), SD –Trp–Ade–His (SD –WAH), SD –WAH +25μM fluconazole + 2% DMSO, SD –WAH +50μM fluconazole + 2% DMSO, and SD –WAH + 2% DMSO. SD –WAH conditions select for reconstitution of ubiquitin.