Deciphering complex traits with deep combinatorial genetic analysis

Albi Celaj[1,2,3], Marinella Gebbia[1], Louai Musa[2], Atina Cote[2], Minjeong Ko[2,6], Jamie Snider[1], Victoria Wong[1], Tiffany Fong[5], Paul Bansal[1,2], Joe Mellor[1], Gireesh Seesankar[5], Maria Nguyen[5], Shijie Zhou[1], Igor Stagljar[1], Nozomu Yachie[4,7], and Frederick P. Roth[1,2,3,6,7,8]
[Author list and order is not final]
[1]Donnelly Centre, University of Toronto, Toronto, Ontario, Canada.
[2]Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada.
[3]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada.
[4]Synthetic Biology Division, Research Center for Advanced Science and Technology, the University of Tokyo, Tokyo, Japan.
[5]McMaster University, Hamilton, Ontario, Canada.
[6]Canadian Institute for Advanced Research, Toronto, Ontario, Canada.
[7]Department of Computer Science, University of Toronto, Toronto, Ontario, Canada.
[8]Corresponding authors

Corresponding Author Information:
Frederick P. Roth, Donnelly Centre and Departments of Molecular Genetics and Computer Science, University of Toronto, 160 College St., Toronto, ON M5S 3E1, Canada
Phone: +1-416-946-5130; Email: fritz.roth@utoronto.ca

Nozomu Yachie, Research Center for Advanced Science and Technology Synthetic Biology Division, University of Tokyo, Rm 4-420, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan.
Phone: +81-3-5452-5242 (x55242); Email: yachie@synbiol.rcast.u-tokyo.ac.jp

Abstract

Many traits are controlled by complex biological systems encoded by multiple genes. Understanding complex genotype-to-trait relationships requires perturbing genes in many different combinations and observing the impact. Here we describe a method to efficiently engineer and phenotype many multi-gene variant combinations within a targeted gene set, enabling a deep combinatorial genetic analysis (DCGA). We generated 5,353 strains, each bearing knockouts for a random subset of 16 yeast ABC transporters. For each strain, we determined the genotype and measured resistance to each of 16 bioactive compounds ('drugs'). The resulting genotype-to-resistance landscapes revealed complex drug-dependent genetic interactions for 13 of the 16 transporters studied. For example, a quadruple knockout (*snq2*Δ *yor1*Δ *ybt1*Δ *ycf1*Δ) unexpectedly showed fluconazole resistance which depended on the activity of a fifth transporter, *PDR5*. A neural network model was used to understand the complex genetic landscape and guide further experimental characterization. Thus, DCGA can discover high-order genotype-to-trait relationships and dissect complex biological systems.

Introduction

Extensive functional interdependency and redundancy in many biological systems results in traits which cannot be straightforwardly understood by observing the individual effects of sequence variants[1–4]. Genes encode gene products which often form interdependent pathways and protein complexes, such that combinations of genetic perturbations can yield surprising phenotypes. This phenomenon defines genetic interaction. Observing the phenotypic effects of genes varied in combination, i.e., performing a combinatorial genetic analysis (CGA), can uncover functional dependencies between genes, and can be used to reconstruct large-scale maps of gene co-function[5]. The ability of CGAs to better understand gene function has been amply demonstrated by comprehensive two-gene interaction maps in yeast[5], and by similar ongoing efforts in human cell lines[6,7]. The resulting genetic interactions maps can not only improve understanding of gene function[5,8], but also inform both mechanism and order-of-action in biological pathways[9].

CGA of many biological traits has shown that additional genetic interactions can arise from the simultaneous perturbation of more than two genes. Diverse pathway architectures can yield three-gene interactions (for which a triple mutant phenotype cannot be simply explained by the component single and double mutant phenotypes)[10], and in yeast these alone are likely to vastly outnumber two-gene interactions[11]. Several examples of interactions of even greater complexity have been reported (e.g. five-[12], seven-[13] and over 20-gene interactions[14]), and complex interactions in general may mediate a large majority of genetic background effects affecting growth of yeast knockouts[15]. In the simplest cases, higher-order interactions arise from functional redundancy in gene families, and multiple paralogs must be perturbed simultaneously for phenotypic consequences[16]. Complex interactions may also have medically-relevant phenotypes. For example, CGA of antibiotic resistance genes in *E. coli* has suggested that the abundance of multi-gene interactions can enable many mutational paths towards resistance[17]. In vertebrates, complex multi-gene effects mediate disease, e.g., myeloid malignancies[18,19]. Moreover, discovery of such interactions can be practically useful. For example, the induction of pluripotent stem cells requires a simultaneous increase in the expression of four genes[20].

While two-knockout CGAs have been used extensively, more exhaustive 'deep' combinatorial genetic analysis (DCGA) has been limited. Major experimental challenges exist in generating and characterizing the vast number of mutant combinations required for such studies. Genome-scale DCGA of even three-gene combinations will likely remain out of reach for years to come. Although DCGA can be targeted towards smaller biological subsystems, the large-scale engineering and profiling of many multi-variant strains is a major bottleneck even in yeast. For example, exhaustive DCGA for a set of 10 genes would require construction of 1,024 haploid strains to sample all combinations of two alleles per gene (e.g. a knockout and wild-type), or $\sim 10^6$ strains if diploid genotypes were considered. Although there are methods to generate multi-mutant strains that can circumvent the limited number of usable selection markers, these have focused on construction of single multi-mutant strains[21]. While methods exist to make modifications at multiple loci simultaneously (multiplex automated genome engineering – MAGE)[22,23], major challenges remain in isolating and genotyping the large number of strains required to perform a DCGA. Extensions of MAGE have been developed to allow parallel phenotyping of many strains for DCGA in *E. coli*[24,25], but exhibit high variance across biological replicates, perhaps due to currently-limited accuracy of large-scale genotyping. Methods have been described for parallel

generation and phenotyping of yeast[26] and human cells[27], but the resulting CGA studies have not gone beyond two-gene combinations.

Here we describe an 'engineered population profiling' strategy enabling DCGA in yeast. We apply this strategy to a target set of all 16 yeast ABC transporters implicated in multi-drug resistance, carrying out high-order DCGA for each of 16 drug resistance phenotypes. ABC transporters were chosen as the pilot gene set for several reasons: First, ABC transporters are an important and conserved gene family which mediates functions such as multidrug resistance, disease progression, and basic cellular homeostasis[28,29]. Indeed, ABC transporters are one of the largest and oldest gene families with over 10,000 members across all three domains of life[30]. Second, although many ABC transporters are generally thought of as imparting drug resistance as one might expect for an efflux pump (and indeed the ABC-16 strain is generally more drug sensitive[21]), ABC transporter knockouts can mediate either drug sensitivity or resistance, and some two-gene ABC transporter knockouts have exhibited synergistic drug resistance[21,31,32]. Complex dependence between mammalian ABC transporters has also been observed, e.g. increased expression of ABCC3 upon disruption of ABCC2 in rats[33] and in humans in the context of Dubin-Johnson syndrome[34]. In another example, mouse ABCG5 and ABCG8 show increased expression in response to disruption of ABCG2 (a protein that confers breast cancer xenobiotic resistance)[35]. Finally, a DCGA study of these 16 transporters is made simpler by the fact that the ABC-16 strain does not show major fitness defects in the absence of drugs[21]. Therefore, we expect progeny bearing a subset of these 16 knockouts will generally be viable, enabling study of the full range of genotypes across a range of different drug exposures.

We show that the resulting multi-knockout phenotype data can be used to model a system of functional relationships amongst ABC transporters. For example, we discovered a quadruple knockout combination (*snq2*Δ *yor1*Δ *ybt1*Δ *ycf1*Δ) that conferred unexpectedly high resistance to fluconazole and ketoconazole that depended on a fifth gene, *PDR5*. We used a non-linear neural network model of the system to guide further mechanistic exploration of this phenomenon. Together, our results show that engineered population profiling can yield many unexpected high-order genetic relationships that shed light on complex molecular systems.

Results

Here we briefly describe the overall strategy for engineered population profiling and its component parts, then show results of the strategy as applied to a set of yeast ABC transporters.

Engineered population profiling: a scheme for generating combinatorial mutants

A simple yet powerful way to generate a complex population is to cross two outbred individuals. In this way, each offspring inherits a random variant at each position of unlinked variation that differs between the parents[36], and can then be genotyped and profiled for traits such as gene expression[37] or small molecule resistance[38]. However, such approaches are traditionally used with natural isolates, presenting several limitations. For example, many genes are undetected in such studies due to limited natural variation in parental strains[39]. Furthermore, diverse parents differing at hundreds of thousands of positions are typically used, which makes it difficult to pinpoint causal variants, and brings multiple testing issues such that a prohibitive number of individuals would be required for a DCGA. To extend cross-based approaches beyond natural strains, we therefore

designed a population engineering strategy in which all variation of interest is engineered into one or a few individuals, and these individuals are then crossed to yield a population of random segregants.

Generating a large pool of barcoded parental cells

A straightforward way to enable tracking of individual strains in a complex population is through the use of DNA barcodes[40]. We therefore designed the process so that one of the haploid parental strains is transformed with a complex pool of random barcodes, such that each cell of one parental strain bears a single specific random barcode. For this, we adopted previously-described methods to create a large pool of uniquely-identifiable clones for one of the parental strains[26,41] (Fig S1, see Methods for details). Because each haploid progeny cell resulting from a cross with this pool will also be uniquely barcoded, the 'parental barcoder pool' is a generally useful reagent that can be employed in different crosses for DCGA.

As described below, the unique tracking identifier facilitates large-scale genotyping and phenotyping of progeny. Isolating a strain, sequencing its identifier barcode, and performing PCR-based genotyping, for example, associates the identifier barcode with a genotype, thereafter allowing for a 'barcode-to-genotype lookup'. An individual barcode identifier also allows for straightforward growth-based phenotyping, in that relative strain abundance measured over time in a competitive pool using high-throughput barcode sequencing can be interpreted as a phenotype[42]. Thus, this trackable genotyped population can be stored as a pool and aliquots of the pool can be interrogated for various phenotypes by tracking competitive growth of each strain in parallel under multiple conditions.

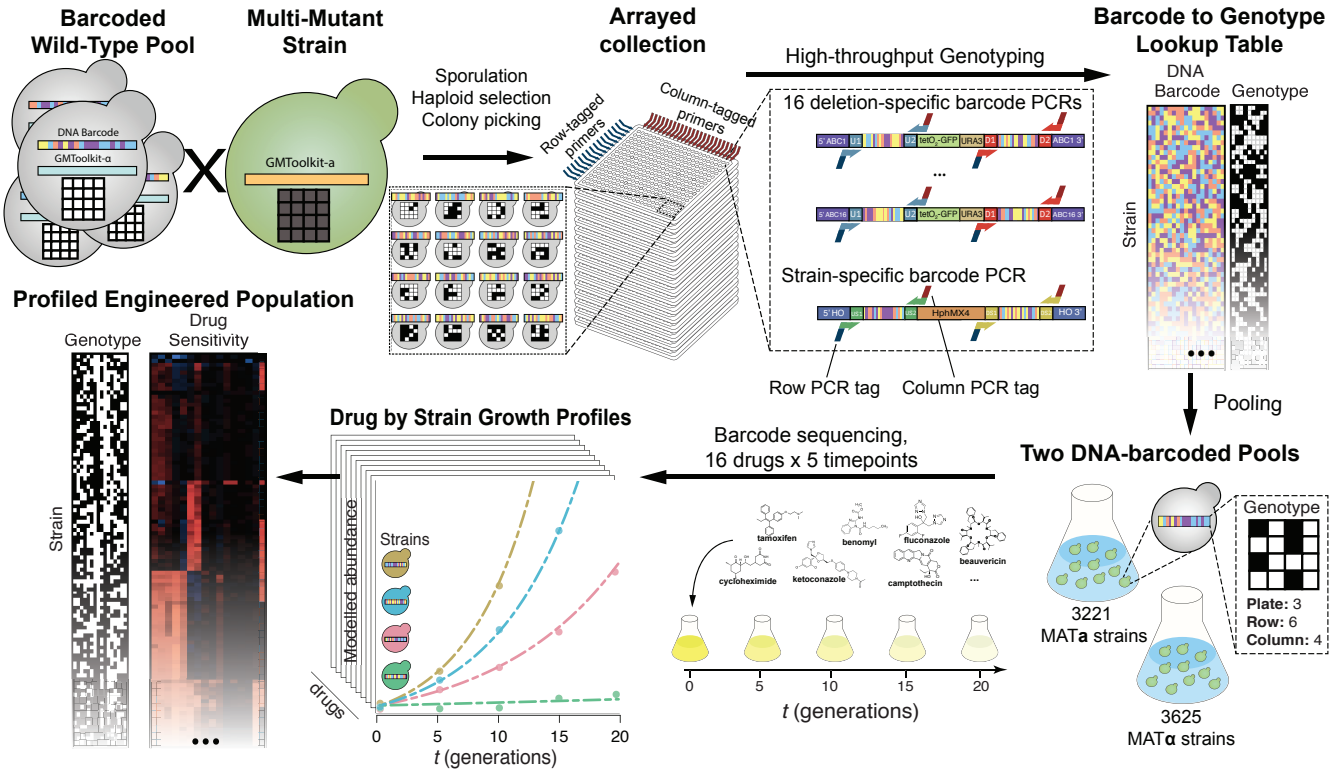
Engineering a population of high-order combinatorial ABC transporter knockouts

After creating a universal barcoded pool with a wild-type parental strain background, we crossed this pool *en masse* with a previously-generated "ABC-16 strain" bearing knockouts of all 16 ABC transporters implicated in multidrug resistance[21]. The ABC-16 strain had previously been engineered to contain 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 selection for barcoded haploid progeny of the cross, automated colony-picking enabled isolation of 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.

Identifying genotypes and unique associated barcodes within the engineered population

After creating a large collection of barcoded multi-knockout progeny, we genotyped each strain and linked this genotype to an individual DNA barcode identifier *en masse*. For this purpose, we exploited the fact that the ABC-16 strain was derived from crosses between barcoded YKO yeast deletion collection strains[21,43], so that each knockout carries at least one identifying barcode that flanks and uniquely identifies the deleted gene. We adapted our previously-described row-column-plate PCR (RCP-PCR) strategy[44], which allows amplification of barcodes in each segregant while introducing additional index tags that identify the plate, row, and column of origin for each amplification product (Methods; Fig 1). Thus, a single next generation sequencing

Figure 1


Barcoded Wild-Type Pool
Multi-Mutant Strain
DNA Barcode
GMToolkit-α
GMToolkit-a
Sporulation
Haploid selection
Colony picking
Arrayed collection
Row-tagged primers
Column-tagged primers
High-throughput Genotyping
16 deletion-specific barcode PCRs
Strain-specific barcode PCR
Row PCR tag
Column PCR tag
Barcode to Genotype Lookup Table
DNA Barcode
Genotype
Strain
Pooling
Two DNA-barcoded Pools
3221 MATa strains
3625 MATα strains
Genotype
Plate: 3
Row: 6
Column: 4
Barcode sequencing, 16 drugs x 5 timepoints
tamoxifen
benomyl
fluconazole
beauvericin
cycloheximide
ketoconazole
camptothecin
0
5
10
15
20
t (generations)
Drug by Strain Growth Profiles
Modelled abundance
Strains
drugs
0
5
10
15
20
t (generations)
Profiled Engineered Population
Genotype
Drug Sensitivity
Strain

reaction can reveal both the strain-specific tracking barcode at the HO locus and the identity of every gene deleted in the segregant at each plate location (Methods; Fig 1).

To validate and calibrate the genotypes determined by high-throughput-sequencing, multiple replicates of 40 individually genotyped 'gold standard' strains, as well as two additional control strains with known genotypes, were added to the collection at defined positions (Methods; Data S2). Using data from calibration strains, we estimated an overall genotyping accuracy of 93.2% (Fig S2A, Methods). An independent method relying on the distribution of knockouts in the pool estimated a similar overall accuracy of 93.8% (Fig S2B, Methods). Based on the genotyping data, all genes were either unlinked or weakly linked except for *BPT1* and *YBT1* (Fig S2C; r = 0.49), which are separated by 70.1kb on chromosome XII. Surprisingly, three gene pairs – *YOR1-YCF1*, *YOR1-BPT1*, and *SNQ2-PDR5* – exhibited weak but significant negative correlation in the appearance of KO genotypes ($-0.04 \geq r \geq -0.08$) (Fig S2C). This effect may have arisen via a negative genetic interaction conferring lower growth for the corresponding double-knockout genotypes during the sporulation, haploid selection, or automated colony picking steps.

Considering only those strains with both high-quality genotyping data and at least one unique tracking barcode, this yielded 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.

Phenotyping the engineered population for diverse drug resistance traits

Knowledge of the tracking barcode for each segregant enabled us to profile each strain's resistance or sensitivity to particular drugs[45]. Strain pools were grown competitively in each of 16 different anticancer and antifungal drugs (Data S3), as well as a solvent control. Using high-throughput strain barcode sequencing[42], strain frequency was measured at five time points (corresponding to 0, 5, 10, 15, and 20 generations of overall pool growth, Fig 1), allowing us to compute a growth rate for each strain (Data S5; Methods).

Inferring genotype-phenotype relationships in an engineered population

By combining genotypes with barcode abundance time-course measurements, we sought both to infer phenotypes for each segregant and associate genotypes with particular phenotypes.

Strains that were well represented in the pre-selection pool offered the best opportunity to detect changes in subsequent time points. Therefore, all further analyses included only the 5,790 (85%) of 6,826 strains that were initially well-represented (≥30 barcode counts at t=0 in the solvent control). To identify gene deletions which have a drug-independent effect, we used the time-course of barcode abundance for each strain to estimate its growth rate in the solvent control and applied a generalized linear model to test association between each gene knockout and growth rate (see Methods). In both the MAT**a** and MAT**α** pools, *yor1*Δ, *snq2*Δ, *ybt1*Δ, and *bpt1*Δ were found to have a statistically significant impact on drug-independent growth rate (Data S6, Fig S3). However, the impacts of *snq2*Δ, *ybt1*Δ, and *bpt1*Δ on drug-independent growth were each small (<2% decrease in the modeled growth rate), while *yor1*Δ had a stronger, but still modest effect (7-15% decrease). We further excluded all 437 strains exhibiting a strong drug-independent growth defect (showing <70% of the median drug-independent growth rate), and drug resistance (growth rate in drug relative to that in solvent control) was measured for each remaining strain (Methods).

In total, drug resistance was calculated for each of 2,367 MAT**a** and 2,986 MAT**α** strains, for each of 16 drugs (Fig 1, Data S5).

For an initial analysis, we sought to limit the complexity of the genetic landscape to the subset of ABC transporters most relevant for resistance or sensitivity for this set of drugs. We applied the above-described generalized linear model to identify and quantitatively model associations between individual knockouts and drug resistance (see Methods). Strong drug-knockout associations were defined by a >10% change in modeled resistance, while other significant associations were defined to be weak. In total we found 62 drug-knockout associations, of which 19 were strong (Data S6).

Because 87% of the single-gene associations (81% of weak associations and 100% of strong associations) involved only five ABC transporters—*snq2*Δ, *pdr5*Δ, *yor1*Δ, *ycf1*Δ, and *ybt1*Δ—we initially restricted our attention to these transporters. For these five 'frequently-associated' transporters, we recovered 14 of 18 previously-reported single-knockout phenotypes, including 6 out of the 7 which had been reported in at least two publications (Fig S4; Data S7). There were 40 novel drug-knockout associations involving one of the five transporters, 33 of which were weak and 7 which were strong. For the vacuolar ABC transporters *YCF1* and *YBT1*, 18 drug-knockout associations were found, all of which were novel (Fig S4, Data S6). Taken together, we detected 79% of 18 previous associations between drugs and individual knockouts of the five targeted transporters, while revealing 40 new associations.

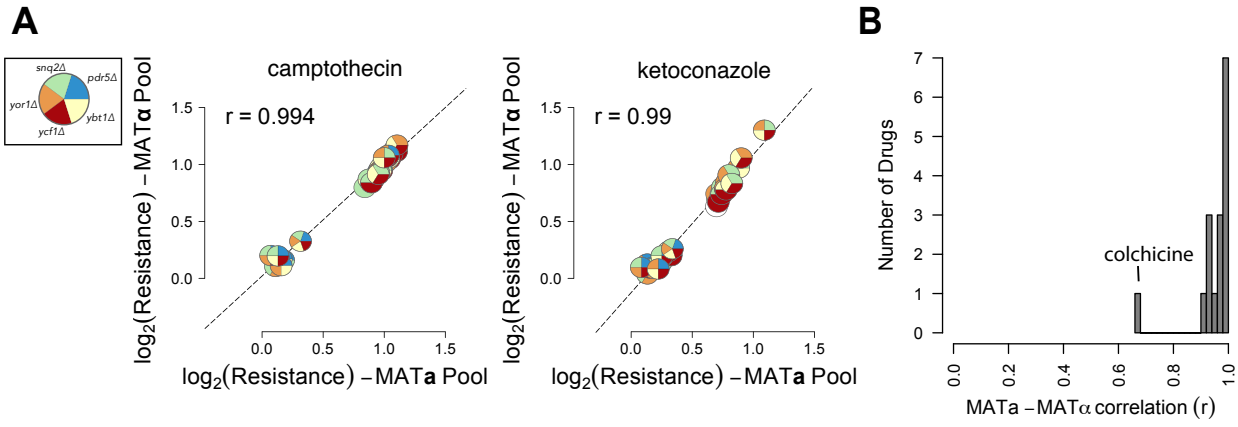
To model the impact of more complex genotypes on drug resistance, we again initially considered only the frequently-associated five transporters (*snq2*Δ, *pdr5*Δ, *yor1*Δ, *ycf1*Δ, and *ybt1*Δ). For each of the 32 (2^5) possible five-gene genotypes, we derived a phenotypic profile by calculating, for each drug, the average resistance over all strains matching this genotype at all five genes. These profiles were initially calculated separately for MAT**a** and MAT**α** strains (Fig S5). Extremely high reproducibility ($r \geq .99$) was observed for camptothecin and tamoxifen between MAT**a** and MAT**α** populations (Fig 2A), and high reproducibility ($r \geq 0.95$) between these independent biological replicate pools was observed for 13 of 16 drugs (Fig 2B). Thus, focusing on the five genes for which drug resistance phenotypes appeared to be most prevalent, we derived robust phenotypic profiles for all possible knockout combinations.

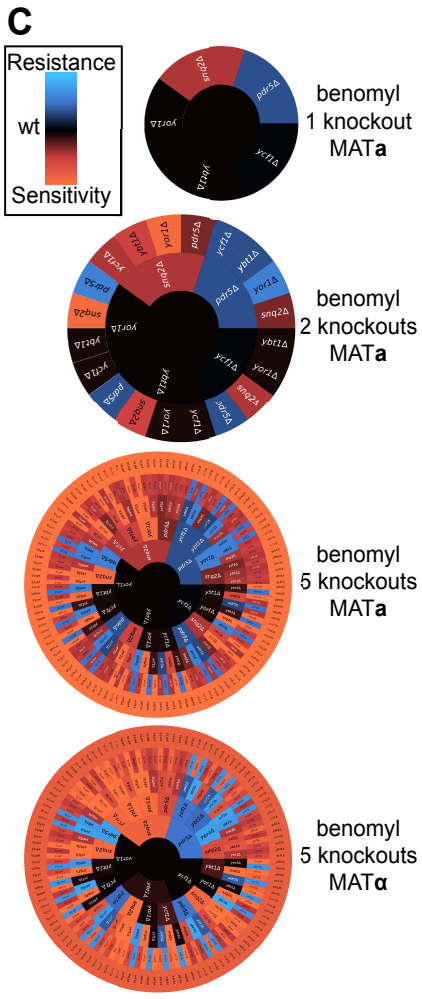
To visualize the complex phenotypic landscape of this exhaustive set of knockout combinations, we developed a radial representation in which the drug resistance consequences of knocking out increasingly-many ABC transporters in a specific order can be explored by tracing different paths leading outward from the central wild-type genotype (Fig 2C). Reflecting the quantitative reproducibility of our profiles, graphs were visually similar between independent biological replicate MAT**a** and MAT**α** populations for many drugs, while showing large differences only for colchicine (Fig 2D, S6). Given the high reproducibility, we merged the MAT**a** and MAT**α** data for all subsequent analyses, except where noted (Methods).

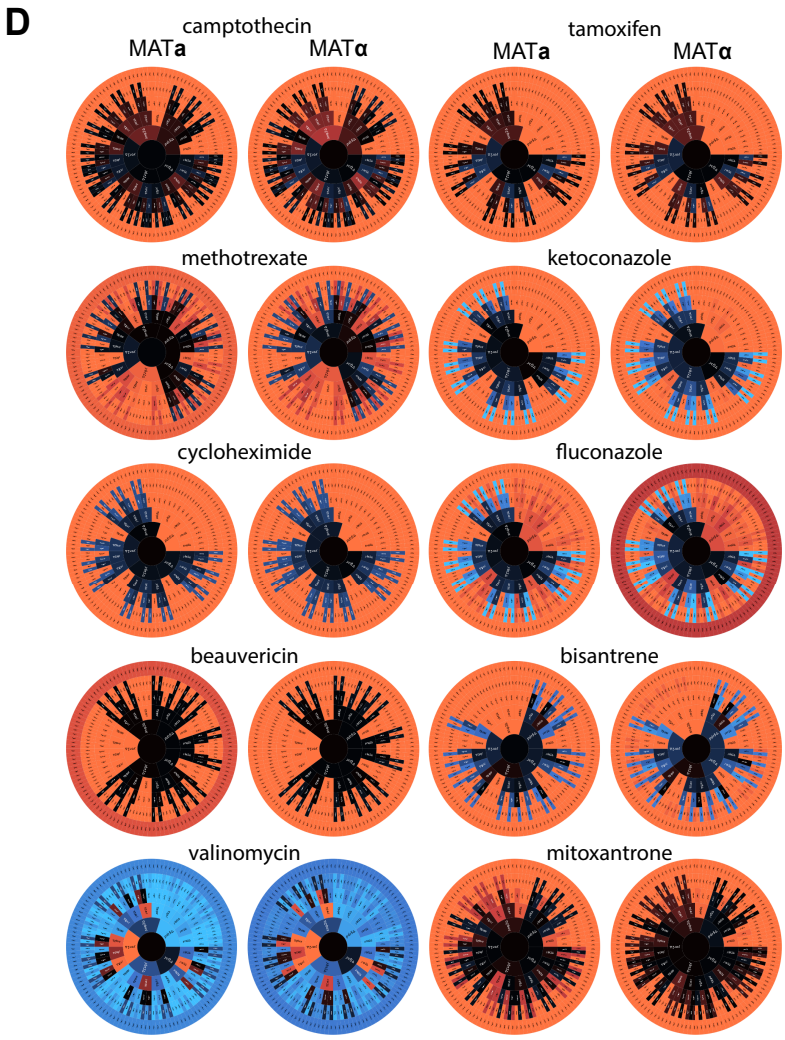
DCGA reveals high-order combinatorial drug resistance effects

Given the well-established role of ABC transporters in drug efflux, one might naïvely expect that the wild-type ABC transporter genes primarily confer drug resistance, such that deleting these genes would lead to increased drug sensitivity. Therefore, it was striking to observe many combinations of ABC-transporter deletions with increased drug *resistance* (Fig 2D).

Figure 2


A
snq2Δ
pdr5Δ
yor1Δ
ybt1Δ
ycf1Δ
camptothecin
r = 0.994
log2(Resistance) – MATα Pool
log2(Resistance) – MATa Pool
ketoconazole
r = 0.99
B
Number of Drugs
colchicine
MATa – MATα correlation (r)


C
Resistance
wt
Sensitivity
benomyl
1 knockout
MATa
benomyl
2 knockouts
MATa
benomyl
5 knockouts
MATa
benomyl
5 knockouts
MATα


D
camptothecin
MATa
MATα
tamoxifen
MATa
MATα
methotrexate
ketoconazole
cycloheximide
fluconazole
beauvericin
bisantrene
valinomycin
mitoxantrone

Visualizing the knockout profiles for each genotype in a fitness landscape representation (Fig 3A), we first verified that these fitness landscapes could capture previously-reported relationships between ABC transporters and benomyl resistance. Our knockout profiles clearly captured the sensitivity of *snq2*Δ deletions to benomyl (Fig 3A left panel; 20% decreased resistance, p = 5.8e-80; Wilcoxon rank sum test), which was expected given that Snq2 is known to be the primary efflux pump for benomyl[46]. We also observed several previously-reported phenomena, including increased benomyl resistance in *pdr5*Δ knockouts (13% increased resistance; p = 1.5e-96) and a further increased benomyl resistance of the *pdr5*Δ *yor1*Δ double-mutant (21% increased resistance; p = 1.3e-72). These increases were dependent on the presence of *SNQ2*[31,32], as *pdr5*Δ*snq2*Δ resulted in only a 5% increase in resistance relative to *snq2*Δ (and a 14% decrease relative to the wild-type), and a comparable 6% relative increase was observed with *pdr5*Δ*yor1*Δ*snq2*Δ (p = 1.4e-45 and 1.2e-38, respectively, when compared to expected knockout effects in a wild-type background, Fig 3A left panel). We did not observe *yor1*Δ to confer benomyl resistance (p = 0.09), a phenomenon that was previously reported to be relatively weak[32]. Thus, engineered population profiling could largely recapitulate previously-reported ABC transporter knockout relationships to benomyl resistance, including the effects of two- and three-gene combinatorial deletion.

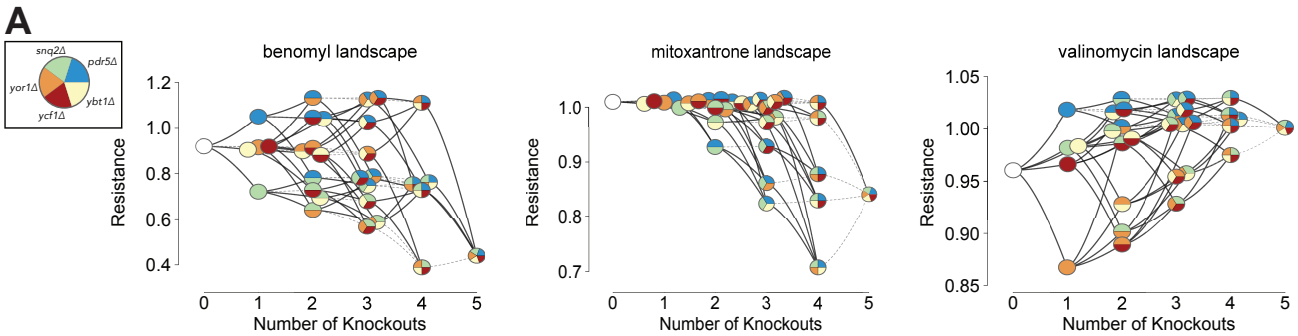
Many of the complex interactions we observed suggested the expected phenomenon of multiple partially-redundant efflux pumps acting in parallel. Specifically, we saw gene sets where each individual knockout shows sensitivity to a drug, and each higher-order knockout combination exhibits drug sensitivity that is higher than any of the individual component knockouts. Examples of this include the set {*snq2*Δ, *pdr5*Δ} under camptothecin (Fig S7), and the set {*snq2*Δ, *pdr5*Δ, *ybt1*Δ, *yor1*Δ} under mitoxantrone (Fig 3A middle panel, S7). These sensitivity patterns are consistent with a simple scenario in which each transporter can efflux a given drug.

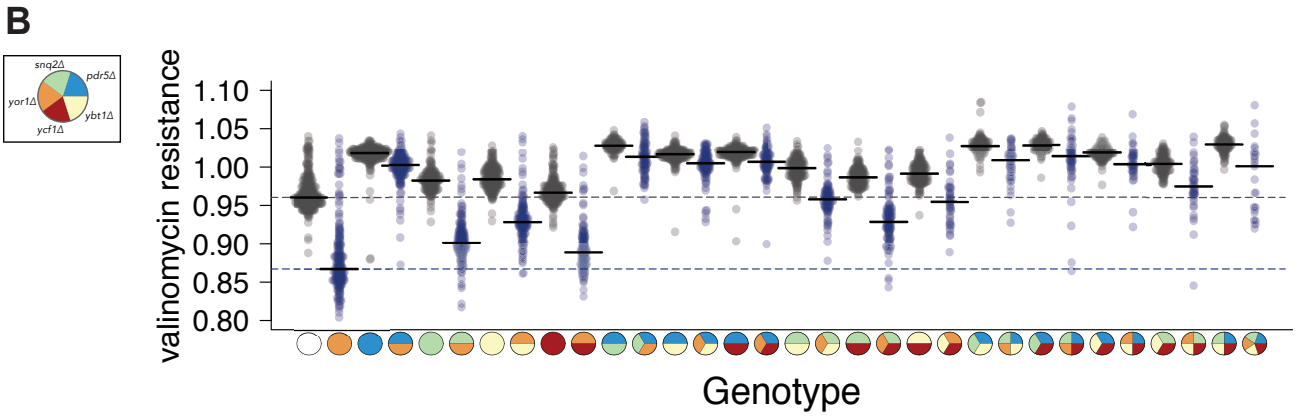
In other cases, the fitness landscapes showed more surprising multi-knockout patterns conveying both drug resistance and sensitivity. For many compounds, multiple paths of successive introduction of deletions led to greater resistance, yielding multi-knockout strains that were considerably more resistant than a wild-type cell (Fig S7). For example, knocking out *pdr5*Δ, *snq2*Δ, and *ybt1*Δ individually or in any combination led to more valinomycin resistance than the wild-type strain (Fig 3A right panel).

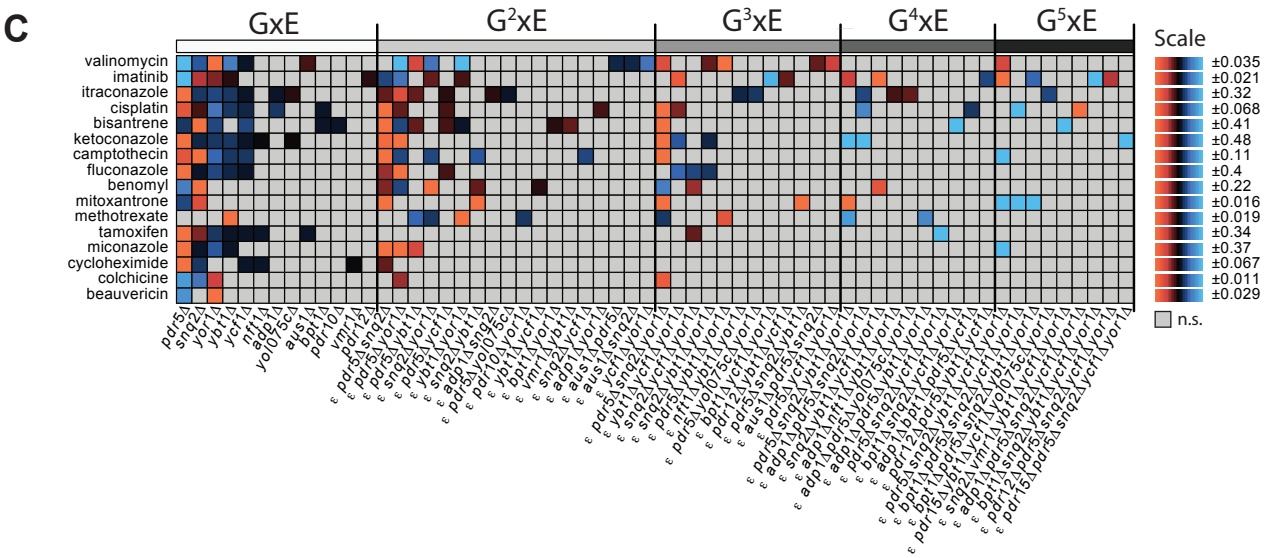
When considering only the five 'frequently-associated' genes, the set of strains matching a specific genotype may in fact have heterogeneous genotypes owing to the variable presence of additional knockouts at the other 11 targeted transporter loci. We therefore visualized the distribution of valinomycin resistance for each of the 5-gene genotypes (grouping the results to show the effects of deleting *YOR1* in each genetic background [Fig 3B]). There was clearly high phenotypic variability within strains matching many of the five-gene genotypes. We therefore systematically expanded our search for multi-gene effects to include all 16 genes, using an extension (see Methods) of the linear model described above in the context of single-gene effects. All single and multi-gene interactions that passed the significance test ($p < 0.05$ after adjusting for multiple testing) are shown in Figure 3C.

Our analysis yielded genetic interactions involving two or more genes for fifteen out of sixteen (94%) of the drugs examined (Fig 3C), with the exception of beauvericin for which we only found

Figure 3


A
snq2Δ
pdr5Δ
yor1Δ
ybt1Δ
ycf1Δ
benomyl landscape
mitoxantrone landscape
valinomycin landscape
Resistance
Number of Knockouts


B
valinomycin resistance
Genotype


C
GxE
G²xE
G³xE
G⁴xE
G⁵xE
Scale
valinomycin
imatinib
itraconazole
cisplatin
bisantrene
ketoconazole
camptothecin
fluconazole
benomyl
mitoxantrone
methotrexate
tamoxifen
miconazole
cycloheximide
colchicine
beauvericin
n.s.

the previously-reported sensitivity of *yor1*Δ knockouts[47]. Higher-order genetic interactions (involving three or more genes) were observed for fourteen of sixteen (88%) of drugs tested (Fig 3C). Here the exception (beyond beauvericin) was cycloheximide. For cycloheximide, we observed the previously-known strong single-gene *pdr5*Δ effect[32,46,48,49], many weak single-knockout effects, and only one weak two-gene interaction between *pdr5*Δ and *snq2*Δ (Fig 3C). Thus, engineered population profiling revealed higher-order genetic interaction involving three or more genes for nearly all drug resistance phenotypes studied.

In total, genetic interactions were found for 14 of the 16 genes that we targeted in our engineered population. Of these 14 genes, 13 were involved in at least one complex interaction involving three or more genes. Remarkably, 11 of the 16 targeted genes were involved in at least one 5-gene interaction. Examples of strong complex interactions involving genes that were excluded from our initial manual exploration of the complex landscape included complex positive interactions involving *pdr15*Δ, *bpt1*Δ, *adp1*Δ, and *vmr1*Δ. In each of these examples, a knockout of one of these genes conferred some resistance in a highly-sensitive multi-knockout background (Fig 3C).

Formalizing the identification of complex genetic interactions captured many of the effects that had been readily-apparent by manual examination of the fitness landscapes, while yielding additional effects. For example, *yor1*Δ was found to have no main effect under benomyl, to have a positive genetic interaction with *pdr5*Δ and, surprisingly, to have a negative genetic interaction with *snq2*Δ (Fig 3C, Data S6). In camptothecin, *pdr5*Δ and *snq2*Δ each had a minor individual negative effect on resistance, and a strong negative interaction was observed between them (Fig 3C, Data S6).

Formal 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*Δ *pdr5*Δ *ybt1*Δ *yor1*Δ quadruple mutant was modelled as the combination of small marginal effects of *snq2*Δ and *pdr5*Δ alone, a two-gene negative interaction between *snq2*Δ and *pdr5*Δ, two three-gene negative interactions (between *snq2*Δ *pdr5*Δ and each of *ybt1*Δ and *yor1*Δ), and a four-gene {*snq2*Δ, *pdr5*Δ, *ybt1*Δ, *yor1*Δ} 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; Fig 3C, Data S6). Together, these complex negative genetic interaction patterns suggest that the four genes enable mitoxantrone efflux in parallel. A similar 'parallel action' genetic interaction pattern was observed for {*pdr5*Δ, *snq2*Δ, *yor1*Δ} in cisplatin (Fig 3C, Data S6).

Objectively modeling the ABC transporter system

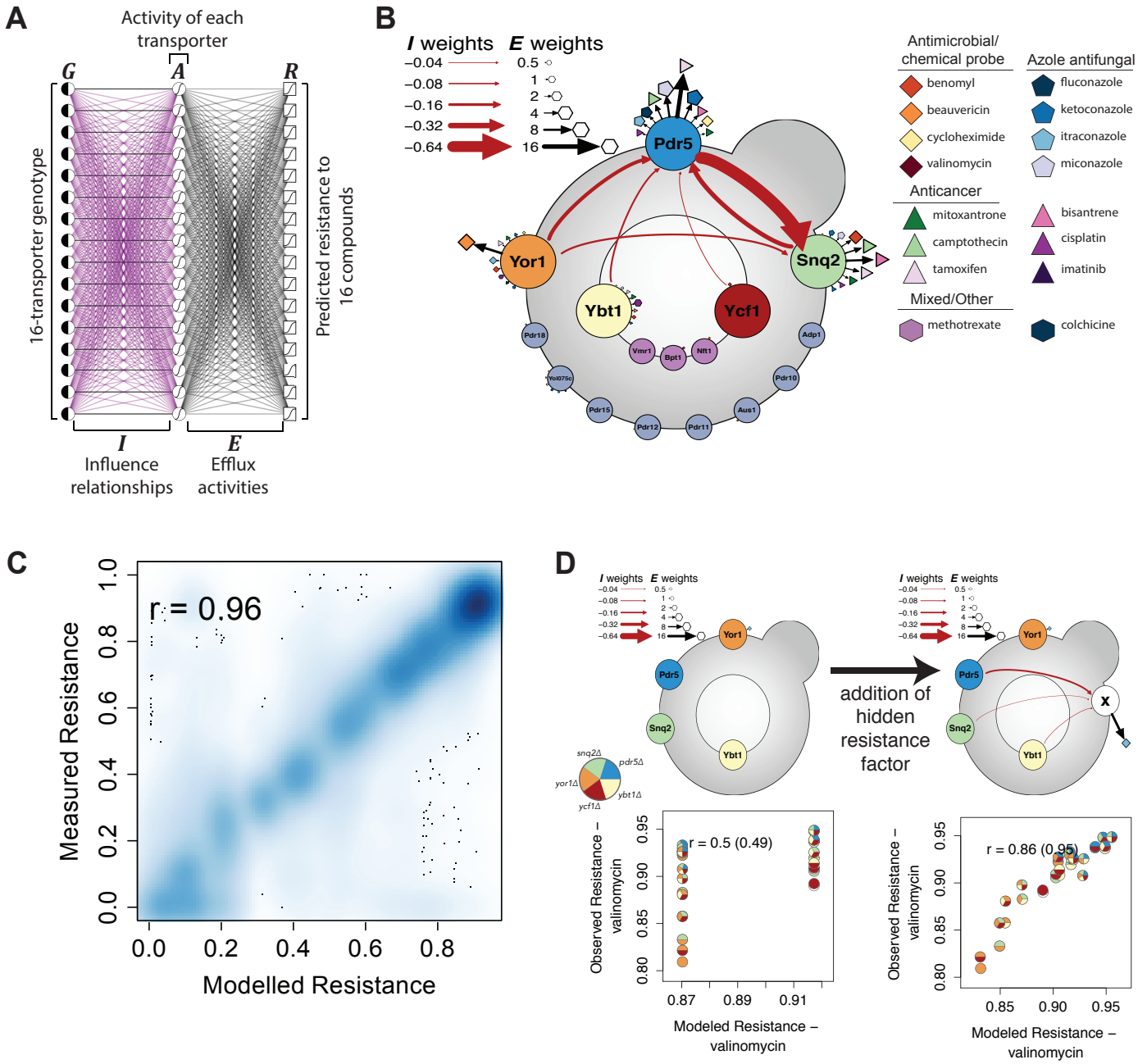
The generalized linear models that were trained for each drug resistance phenotype do achieve the important goal of capturing a complex genotype-phenotype relationship, while also describing single-gene effects and genetic interactions. However, these models do not efficiently convey useful intuition about the system. Above, we manually reasoned that the observation of negative genetic interactions amongst a set of transporter genes suggests that each transporter is independently capable of drug efflux. Alternatively, manual application of classical epistasis analysis might lead us to conclude that the presence of one transporter can activate or repress another (either directly or indirectly). However, manually derived intuition from a complex system is laborious, error-prone, and potentially subjective.

To more systematically derive intuitive models of the system from complex genotype-phenotype relationships, we developed a neural network model. We structured the neural network model (Methods, Fig 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)*** that capture the activity of each of the 16 transporters; and 3) an output layer that quantitatively describes resistance (***R***) to each of 16 drugs. The links between input and hidden layers have (initially unknown) weights (***I***) that represent possible pairwise regulatory influence relationships between transporters (with positive weights for increased activity and negative weights for decreased activity). The links between the hidden and output layers have (also initially unknown) weights (***E***) that capture the extent to which each transporter can catalyze the efflux (or otherwise reduce the intracellular activity) of each drug. Using our complete set of drug resistance phenotypes for each genotype as training data, we learned the network weights using back-propagation with stochastic gradient descent (Methods). The cost function that was used to optimize network weights contained a penalty which acts to limit the number of non-zero weights, and has the effect of favoring more parsimonious models (Methods, Fig S8A-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 an input dataset of 97,392 training examples (6,087 unique genotypes × 16 drugs), we learned an interpretable neural network with only 74 non-zero fitted parameters.

Despite its relatively parsimonious nature, the resulting neural network model largely recapitulated the input data (r = 0.96, Fig 4C). Over-fitted models may exaggerate performance when tested using data that was also used in training. Therefore, we also assessed the model on data from one mating type and testing it on the other. We found similar performance when the model was tested with data that had not been used in training (r = 0.95 and r = 0.96 when using either mating type **a** or **α** as training, respectively [Fig S8C]). Training using each of these two independent biological replicate datasets also yielded strong agreement in the parameter values (r = 0.98, Fig S8D), suggesting that model parameters were robustly determined.

The objectively-trained model provided intuition that was largely in agreement with manual interpretations. For example, in keeping with the observation that *snq2*Δ, *yor1*Δ, *ybt1*Δ, and *ycf1*Δ increased activity of *PDR5*, the model found *SNQ2*, *YOR1*, *YBT1*, and *YCF1* to each have a negative influence on *PDR5* activity (Fig 4B). The manual genetic interpretation that Pdr5, Snq2, Yor1, and Ybt1 are each independently able to efflux mitoxantrone was also supported by positive E links connecting each of these transporters to mitoxantrone (Fig 4B). The model showed Snq2 to have the highest mitoxantrone efflux activity (E = 2.3) followed by Pdr5, Yor1, and Ybt1 (E = 1.9, 0.6, 0.6, respectively; Fig 4B, Data S8). These differences were reflected in the fitness landscape: For example, resistance of *pdr5*Δ*ybt1*Δ*yor1*Δ was not significantly different than the wild-type (p = 0.25), whereas deletion of the two highest-clearance transporters *snq2*Δ*pdr5*Δ resulted in a 9% decrease in resistance (p = 1.2e-70). The model also pointed to differential inhibitory effects between transporters: For example, Snq2 is predicted to be more strongly inhibited by *PDR5* than by *YOR1* (I = -0.96 vs -0.39, Fig 4B, Data S8). Although this might have been gleaned from the observation that *pdr5*Δ yields greater benomyl resistance than does *yor1*Δ

Figure 4


A
Activity of each transporter
G
A
R
16-transporter genotype
Predicted resistance to 16 compounds
I
Influence relationships
E
Efflux activities
B
I weights
E weights
−0.04
−0.08
−0.16
−0.32
−0.64
0.5
1
2
4
8
16
Pdr5
Yor1
Snq2
Ybt1
Ycf1
Pdr18
Vmr1
Bpt1
Nft1
Adp1
Yol075c
Pdr10
Pdr15
Pdr12
Pdr11
Aus1
Antimicrobial/ chemical probe
benomyl
beauvericin
cycloheximide
valinomycin
Azole antifungal
fluconazole
ketoconazole
itraconazole
miconazole
Anticancer
mitoxantrone
camptothecin
tamoxifen
bisantrene
cisplatin
imatinib
Mixed/Other
methotrexate
colchicine
C
r = 0.96
Measured Resistance
Modelled Resistance
0.0
0.2
0.4
0.6
0.8
1.0
D
snq2Δ
pdr5Δ
yor1Δ
ybt1Δ
ycf1Δ
addition of hidden resistance factor
x
r = 0.5 (0.49)
r = 0.86 (0.95)
Observed Resistance – valinomycin
Modeled Resistance – valinomycin
0.80
0.85
0.90
0.95
0.87
0.89
0.91

(Fig 3A), the neural network model provides a clearer statement of the inferred biological relationships.

Potential for iterative refinement of genotype-to-phenotype models

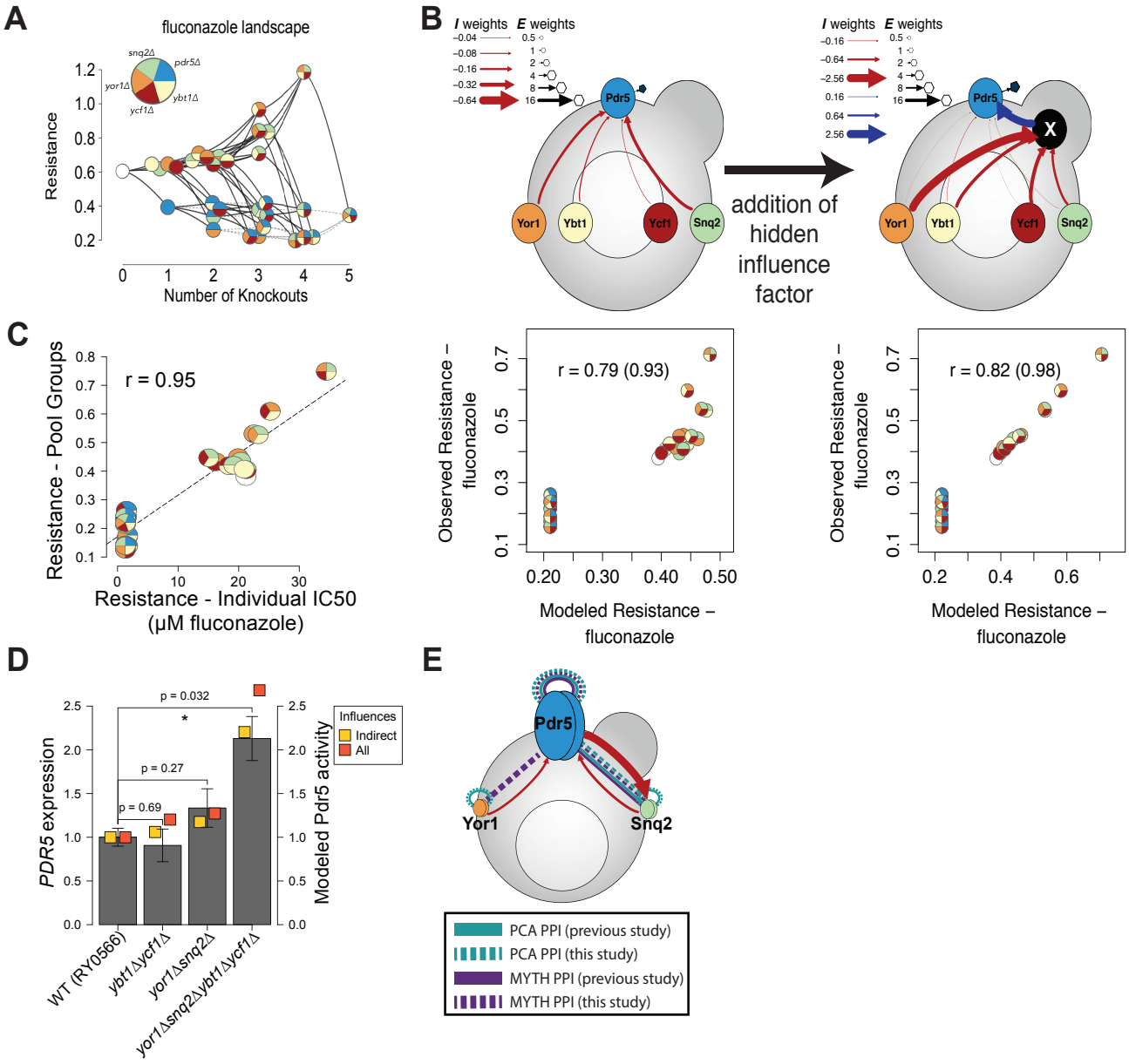
Despite the overall accuracy of the neural network model, some inputs yielded predictions with which departed systematically from observation (Fig S9), suggesting the need for model refinements. For example, valinomycin resistance was quite poorly predicted by the neural network model (r = 0.49, Fig 4D, left panel). As the five-deletion strain was observed to be more valinomycin resistant than the wild-type (Fig 3A, right panel), we considered the possibility that a valinomycin efflux pump (or other resistance mechanism) exists outside of our set of 16 targeted transporter genes, and is negatively influenced by one or more of our 16 studied transporters. Indeed, it had been previously observed that the ABC-16 strain is also more resistant to valinomycin than the wild-type[21,50]. To formalize this possibility, we added one additional 'mystery transporter gene' (always present) and its corresponding activity node to the neural network. Training the neural network only on our valinomycin data yielded a model in which *SNQ2*, *PDR5* and *YBT1* each inhibit a 'mystery transporter' which clears valinomycin, and led to substantially improved modeling of the observed phenotypes (r = 0.86, Fig 3A, right panel). Training a neural network model separately on valinomycin data without an additional factor did not result in similar improvements (r = 0.49, Fig S10A).

Further exploration of complex synergistic resistance to fluconazole

One striking phenotype observed by DCGA was a quadruple deletion—*snq2*Δ *ybt1*Δ *ycf1*Δ *yor1*Δ—with high resistance to fluconazole (Fig 5A), and a similar phenomenon was observed for ketoconazole (Fig S7). Interestingly, the quintuple mutant *pdr5*Δ *snq2*Δ *ybt1*Δ *ycf1*Δ *yor1*Δ (differing from the resistant quadruple genotype only by an additional *pdr5*Δ deletion) showed sensitivity to fluconazole that was comparable to the single-knockout *pdr5*Δ genotype. These results extend previous findings that deletions of *SNQ2* or *YOR1* (either alone or together) increase resistance to fluconazole, and that *snq2*Δ and *yor1*Δ deletions result in increased *PDR5* expression and activity[31]. Generalized linear regression had modeled this phenomenon (in addition to one- and two- gene effects), as the combination of three positive three-gene interactions (all three-knockout combinations of {*yor1*Δ, *snq2*Δ, *ybt1*Δ, *ycf1*Δ} except *snq2*Δ *ybt1*Δ *ycf1*Δ - Fig 3C). The dependence of the resistance resulting from these multi-knockout combinations on the presence of *PDR5* was modeled as three two-way negative interactions: {*pdr5*Δ, *snq2*Δ}, {*pdr5*Δ, *ycf1*Δ}, and {*pdr5*Δ, *yor1*Δ}. These phenotypes further suggest that: 1) fluconazole resistance is increased further by *ybt1*Δ and *ycf1*Δ knockouts in addition to *snq2*Δ and *yor1*Δ; 2) the resistance provided by *ybt1*Δ and *ycf1*Δ is synergistic with that provided by *snq2*Δ and *yor1*Δ; and that 3) resistance of the *snq2*Δ *ybt1*Δ *ycf1*Δ *yor1*Δ knockout strain depends on the presence of a wild-type *PDR5*.

As the neural network had predicted negative influence to Pdr5 from *SNQ2, YBT1, YCF1*, and *YOR1* (Fig 4B), we examined whether this model was congruent with the observed resistances. While the model correctly predicted that the *snq*Δ*ybt1*Δ*ycf1*Δ*yor1*Δ strain would be more resistant to fluconazole than strains carrying any subset of these knockouts, the additive inhibition model of our neural network model under-estimated the resistance of this four-knockout strain (Fig 5B left panel). As the genetic interactions suggested synergistic rather than additive influence on *PDR5*, we formally tested for this possibility. A standard strategy in neural network design, that allows for the non-additive combination of multiple inputs converging on a target node is to have

Figure 5


A
fluconazole landscape
snq2Δ
pdr5Δ
yor1Δ
ybt1Δ
ycf1Δ
Resistance
Number of Knockouts
B
I weights
E weights
Pdr5
Yor1
Ybt1
Ycf1
Snq2
addition of hidden influence factor
X
C
r = 0.95
Resistance - Pool Groups
Resistance - Individual IC50 (μM fluconazole)
r = 0.79 (0.93)
Observed Resistance – fluconazole
Modeled Resistance – fluconazole
r = 0.82 (0.98)
D
p = 0.032
p = 0.27
p = 0.69
PDR5 expression
Modeled Pdr5 activity
Influences
Indirect
All
WT (RY0566)
ybt1Δycf1Δ
yor1Δsnq2Δ
yor1Δsnq2Δybt1Δycf1Δ
E
Pdr5
Yor1
Snq2
PCA PPI (previous study)
PCA PPI (this study)
MYTH PPI (previous study)
MYTH PPI (this study)

those inputs converge additively on one or more 'hidden' nodes, which can then convey a non-additive signal to the original target node. Addition of a single hidden 'modifier' neuron yielded better predictions of fluconazole resistance for the three- and four-knockout strains and led to a model where negative influences partly converge on a hidden factor, which then conveys a positive influence on Pdr5 activity (Fig 5B right panel). Conversely, training a separate neural network on fluconazole data led to similar under-estimation as the original model (Fig S10B).

Before exploring this phenomenon further, we first wished to replicate the initial observations within a single genetic background. Therefore, we generated a single strain for each of the 32 possible combinations of *pdr5*Δ, *snq2*Δ, *yor1*Δ, *ybt1*Δ, and *ycf1*Δ knockouts in a common genetic background (Methods). Fluconazole resistance as estimated by DCGA correlated well with measures of resistance obtained for individual strains—$r = 0.95$ for the fluconazole concentration expected to yield 50% inhibition (IC50; Fig 5C) and $r = 0.89$ for total growth in fluconazole relative to no-drug conditions (Fig S11; Methods). Consistent with DCGA results, *snq2*Δ*yor1*Δ*ybt1*Δ*ycf1*Δ had the highest resistance.

We considered two potential mechanisms by which *SNQ2*, *YOR1*, *YBT1*, and *YCF1* might negatively influence Pdr5 activity: 1) indirect reduction via reduced *PDR5* transcript levels; and 2) direct inhibition of Pdr5 via inhibitory protein interaction.

Inhibition of Pdr5 activity by *SNQ2* and *YOR1* was previously reported, explained by reduced activity of the transcription factor Pdr1 via an unknown mechanism, which in turn yields reduced *PDR5* transcript levels[31]. Using qRT-PCR, we found *snq2*Δ*yor1*Δ 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*Δ*yor1*Δ[31], the change did not reach statistical significance in our hands ($p = 0.27$, Fig 4E) and the previous report did not contain a statistical test. No evidence of mRNA induction in the *ybt1*Δ*ycf1*Δ strain was observed (0.9 fold expression, $p = 0.69$, Fig 4E). More persuasively, we found that *PDR5* mRNA levels were significantly higher in *snq2*Δ*yor1*Δ*ybt1*Δ*ycf1*Δ than the wild type (2.1× increase; $p = 0.032$; Fig 4D). Interestingly, relative expression of *PDR5* tracked well with the relative activity expected by the neural network model, especially when considering only indirect influences (Fig 5D). Although these results can provide only weak evidence for the previous finding that some combination of *SNQ2* and *YOR1* yield reduced *PDR5* mRNA levels, our results more confidently support the idea that the fluconazole resistance *snq2*Δ*yor1*Δ*ybt1*Δ*ycf1*Δ can be explained (at least in part) via increased *PDR5* transcript levels.

Insight into whether Snq2 might inhibit Pdr5 directly via protein interaction comes from a previous study investigating an Snq2-dependent decrease in benomyl resistance imparted by *PDR5* and *YOR1* found evidence for a modest (1.5×) *SNQ2* mRNA induction only in *pdr5*Δ*yor1*Δ, but not for either *pdr5*Δ or *yor1*Δ, despite the fact that each of these single mutants showed increased benomyl resistance[32]. This result suggested the possibility that transporter genes can negatively influence one another by non-transcriptional means, and *PDR5*-mediated repression of Snq2 activity was hypothesized to result at least in part from a direct protein-protein interaction between Pdr5 and Snq2[32]. This 'direct repression' model, in which heterodimerization of Pdr5 and Snq2 transporters draws subunits away from the homodimeric Snq2 complex and thereby reduces total efflux activity of Snq2, drew support from the observed homodimeric interactions of Pdr5, Snq2,

and a heterodimeric interaction between Pdr5 and Snq2. Although the possibility that Snq2 can reciprocally inhibit Pdr5 has not been explored, the possibility of this mechanism is strongly suggested by the previous benomyl resistance study.

We next further investigated whether the observed Pdr5-dependent decrease in fluconazole resistance provided by *YOR1* might also be mediated by direct physical interactions. This model predicted a heterodimeric interaction between Pdr5 and Yor1, and is made more plausible by the fact that Pdr5 and Yor1 are paralogs, and each can form a homodimer[32,51]. Because all known protein interaction-testing methods miss the majority of real interactions[52], we re-tested for the model-predicted Pdr5-Yor1 interaction using two distinct protein interaction assays: MYTH[32] and PCA[51]. All previously-known MYTH and PCA interactions amongst Pdr5, Snq2, and Yor1 were recovered in their corresponding assays (Fig 4F, S11, S12). Although PCA (Fig S11) did not detect the predicted Pdr5-Yor1 interaction, it was revealed by MYTH (Fig 4F, S12). Given a much-higher baseline abundance of Pdr5 than Snq2[53], a 'heterodimeric repression' model is consistent with the prediction that negative influence of Snq2 by Pdr5 will be greater than negative influence of Pdr5 by Snq2 (I = -0.81 vs -0.25, Fig 4B). This is because a greater proportion of Snq2 would be affected by each heterodimeric interaction than would Pdr5. Taken together, these experiments provide support for two different mechanisms whereby gene deletions can relieve Pdr5 inhibition, one in which deletion of four genes relieves *PDR5* expression, and another in which deletion of *snq2*Δ, *yor1*Δ or both can relieve direct physical inhibition of Pdr5 by Snq2 or Yor1.

Discussion

The use of genetics to dissect and understand complex biological systems has been limited by the difficulty in systematically measuring and interpreting the combinatorial effects of multiple variants. Here we illustrated a method to straightforwardly generate and phenotypically profile a large population with variation segregating at sixteen loci. The complex genetic landscape emerging from this DCGA approach revealed many phenomena that would not have emerged from considering one or two variants at a time. We also demonstrated that the systematically-generated complex phenotypic profiles could be used to automatically derive a computational model which intuitively modeled the way in which variants in some ABC transporters can influence the activity and phenotypic consequences of other transporters. Even within the highly-characterized ABC transporter gene family, DCGA revealed many novel gene functions and gene-gene relationships. These results motivate continued use of this engineered population to study ABC-transporter-mediated drug resistance of other compounds. More broadly, these results illustrate the potential for carrying out DCGAs for other sets of functionally-related genes.

Although variants of the DCGA approach will be required for other systems, it is already straightforwardly adaptable for other yeast strains carrying multiple variants, e.g., a 16-deletion mutant for GPCR pathway-related genes[54]. In other model organisms such as *C. elegans*, methods to introduce many targeted gene knockouts[55] or loss-of-function mutations[56] into a single strain may enable analogous cross-based strategies. The cross-based approach allows mutations to be distributed as needed between the two parents, so that a similar strategy is possible even if all mutations cannot be introduced into a single individual. The single-cross strategy we describe here could easily be replaced with multiple matings between multiple parental strains carrying different subsets of targeted variation, and subsequent inter-crosses between F1 populations.

To enable engineered population profiling without the use of a cross, further molecular tools must be developed to allow for direct introduction of trackable multi-allele diversity into an isogenic population. Such developments would allow DCGA in non-mating model systems, such as human cell lines. Human ABC transporters have roles not only in drug response and chemotherapeutic resistance, but are clearly associated with at least 10 inherited diseases, and likely many others[35]. Thus, an analogous DCGA of human ABC transporters is one of many intriguing possible avenues for understanding genetic systems in depth. Direct population engineering presents more of a challenge than crossing multi-variant parental strains, but technical advances in this area continue to be made[25,27].

Although we generated a sufficient fraction of all possible deletion combinations for the sixteen targeted transporters to detect four and five-gene interactions, further scaling would be needed to exhaustively apply this approach to genotype and identify barcodes for still-larger populations. Recent advances in single-cell sequencing[57] could potentially be applied to this problem.

We demonstrated that DCGA can enable the development of computational approaches that expand the use of genetic inference to understand biological systems. While it is challenging to use manual epistasis analysis to derive biological models when dealing with many complex knockout combinations under multiple environments, automated learning of genotype-to-phenotype models can be used to objectively derive similar biological relationships from genetic data[58,59]. Formal modeling additionally allows evaluation of how well the proposed biological relationships explain the data, and we show that they can also guide iterative extension of the model to capture more complex phenomena as needed. Further modeling work should consider genes which had no weights in the neural network despite exhibiting complex genetic interactions. Profiling more knockout combinations under a greater variety of environments will straightforwardly enable the learning of a more complete model of ABC transporter function. In general, future availability of combinatorial variant profiling data will enable further development of computational modeling of complex genotype-to-phenotype relationships in other systems.

Important to our DCGA approach was the use of strain-specific molecular barcodes. Competitive selection of such barcoded populations, coupled with sequencing, can allow efficient multiplexed measurement of many phenotypes that are more complex than simple growth under different environments[40]. For example, fluorescence-based sorting strategies can convert many assays into a selection that can effect a detectable change in barcode abundance; for example, fluorescent reporters driven by specific promoters can be used to study the effects of knockouts on the activation of signaling pathways, phosphorylation state, epigenetic modifications, or protein abundance[60]. A fluorescence sorting and sequencing strategy could also be used, for example, to directly study the dynamic uptake and efflux of small molecule by incubating cells with fluorophore-conjugated compounds and using changes in barcode frequency at different time points to measure change in cell fluorescence over time[31]. In addition to DNA-based molecular barcodes, methods to genotype large populations of cells after imaging are being developed, and would allow high-content characterization of multi-knockout strains[61]. Thus, with the appropriate design, rich phenotyping for multi-knockout strains may be possible at a large scale.

We envision that the profiling of engineered populations will permit DCGA to dissect, model, and understand multi-gene systems in many living organisms.

Materials and Methods

Yeast strains

RY0622/GM512 (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)*

RY0148 (Barcoder Strain MATα):
MATα lyp1Δ his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 can1Δ::GMToolkit-α (CMVpr-rtTA NatMX4 STE3pr-LEU2) hoΔ::LoxP UP-tag HphMX4 DN-tag Lox2272

Media
SC (SC-His, SC-Leu, SC-Ura)
YPD (+HygroB, +Clonnat, +G418)

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 purified 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. Two PCR reactions were performed on the resulting products to confirm correct synthesis. 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 Qiagen Qiaspin. 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[62]. Transformation was carried out using a previously established protocol[63], 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 recombined barcoder plasmid.

Twenty colonies were tested for 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 4% gel electrophoresis to verify the presence of 263bp and 251bp bands. EXOSAP purification was performed on the PCR products by adding 10µl EXOSAP mix (0.025µl ExoI (0.5U), 0.1µl Antarctic Phosphatase (0.5U), 3.5µl 10X Antarctic Phosphatase Buffer, 6.375µl dH_2O) to 25µl of PCR products and incubating at 37°C for 30min; 80°C for 20min, then diluting with 35µl of DNA-free H_2O to stop the reaction.

Diluted EXOSAP products 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[21]. 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 RY0148 barcoder strain (MAT**α**) as previously described[21], selecting for MAT**a** and MAT**α** progeny separately. Using colony plating, sporulation efficiency was estimated at 24% - 1080 colonies grew in SC, 140 colonies grew in SC –His (MAT**a** haploid selection), and 120 colonies grew in SC –Leu (MAT**α** haploid selection). 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).

To validate the mating and selection strategies, we pooled the MAT**a** and MAT**α** collections and subjected them to cell sorting, confirming haploidy of the overall pool (Fig S2D), and furthermore we tested that samples from each pool do not exhibit any growth in the selection conditions of the opposite mating type.

Pooled strain genotyping

A previously-developed Row-Column-Plate (RCP)-PCR protocol[44] 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[44]. 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. 4µl of overnight yeast culture was mixed with 8µl 0.2M sodium phosphate buffer (pH 7.4), 4µl DNA free dH2O, 0.05µl 5U/µl Zymoresearch zymolyase and incubated at 37°C for 35 minutes. 64µl DNA free dH2O was added to each well to stop the reaction.

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 2µl of lysed colonies using a Hydrocycler with the following program: 95°C for 5min; 23 cycles of 95°C for 60sec, 57°C for 35sec, 72°C for 45sec; 72°C for 2min; 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 pooled and 260µl was purified using a Qiagen Qiaquik Spin kit. DNA yield was quantified using a Nanoquant. From the resulting products from each plate, Illumina adapters containing plate tags were added using an additional PCR reaction as previously described[44]. A pair of PXX_PE1.0 and PYY_PE2.0 primers (Data S2) were added to 3-6µl pooled products (calibrated to ~150ng) from each plate to encode the plate of origin, and were amplified using the following PCR program: 98°C for 30sec; 15 cycles of 98°C for 10sec, 59°C for 15sec, 72°C for 40sec; 72°C for 2min; 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 Qiaquik Spin kit. 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 100µl of this multiplexed sample were run on a 4% gel. Products of the desired size (260-290bp) were isolated from each lane, and purified using a Qiagen gel purify 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, 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 Fig 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. When processing the sequencing data, a single strain was randomly chosen to represent each unique UP and DN tag sequence.

Examining putative wild-type pool strains

For 73 MAT**a** and 131 MAT**α** 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, 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. 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 stated genotype was corrected (Data S2). The 5 unsucessfully genotyped strains, as well as 28 additional strains were discarded from analysis. When calculating linkage and distribution of gene knockouts (Fig S2), 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 there are no wild-type to mutant miscalls, this number is most likely with an asymmetric genotyping accuracy of 93.8%, compared to the 93.2% estimated by comparison to gold standards (Fig S2C).

Indiviual liquid growth profiling

To measure individual strain growth, the $OD_{600\ nm}$ of a 0.0625 $OD_{600\ nm}$ starting culture was measured in the appropriate medium every 15 mins using a GENios microplate reader (Tecan).

Drug testing for growth inhibition

The effects of 16 different drugs on strain growth were tested to find a concentration which inhibits wild type growth by approximately 20% (Data S3). All drugs used were dissolved in 2% DMSO, which was used as a solvent control. Growth was determined by the Average_G metric[64], which represents the average generation time.

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 + 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 $OD_{600\ nm}$ in 10ml. For the solvent control, a 0 generation sample was immediately harvested for sequencing. After growth to approximately 2 $OD_{600\ nm}$, a sample was taken from each drug for sequencing and cells were resuspended in fresh medium to a final concentration of 0.0625 $OD_{600\ nm}$. This process was repeated until 4 generations of samples were collected. Collected samples corresponded approximately to 5, 10, 15, and 20 generations of growth. Harvested samples were subject to genomic DNA extraction using a YeaStar™ Genomic DNA Kit, quantified using a Qubit® 2.0 fluorometer, and diluted to a final

concentration of 20ng/µl. Approximately 350ng of isolated DNA was extracted from each sample and added to 20µL of 2x Platinum PCR SuperMix High Fidelity, 1µL of 10µM F primer, and 1µl of 10µM 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 amplified using the following PCR program: 98°C for 30sec; 24 cycles of 98°C for 10sec, 60°C for 10sec, 72°C for 1min; 72°C for 5min; 4°C forever.

PCR products were subject to gel electrophoresis, and ~210bp bands were isolated, subject to gel purification, and eluted in 60µl tris buffer. DNA yield was quantified in duplicate using a KAPA qPCR assay kit, at 1,000-fold, 10,000-fold, and 100,000-fold dilutions to find a concentration within standard curve range. Samples were pooled to yield approximately equal amounts of DNA, and subject to sequencing using an Illumina NextSeq 500 Mid Output 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 were discarded from the analysis. Additionally, if a sample was discarded for one mating type, the corresponding sample for the opposite mating type was also discarded (e.g. if 'miconazole t=15 MAT**a**' was discarded due to lack of coverage, 'miconazole t=15 MAT**α**' would also be discarded regardless of coverage).

Defining a resistance metric

Following processing of the sequence data, a count c was assigned for each strain s_x in a pool under drug d sequenced at time t ($c_{s_x,d,t}$). The counts in each sample were then converted to a frequency $f_{s_x,d,t}$ by division with the total count for all strains in that sample:

$$f_{s_x,d,t} = \frac{c_{s_x,d,t}}{\sum_{i=1}^{n} c_{s_i,d,t}}$$

If both an UP and DN tag for a given strain were successfully linked to a genotype, $f_{s_x,d,t}$ is estimated from the UP and DN counts and averaged, otherwise the available tag is used. The frequency of each strain was then converted into a 'area under the growth curve (AUC) by first multiplying the frequency at each time point by the expected overall pool growth at that time (2^t, since t is defined by the number of generations) to estimate the individual abundance over time of

each strain, then taking the integral over all measured timepoints 0 to T (the total number of pool generations measured). Frequencies between measured timepoints were linearly interpolated.

$$AUC_{s_x,d} = \int_0^T f_{s_x,d,t} 2^t dt$$

We modelled each strain as growing constantly from an initial abundance (A_0) by an exponential growth rate g in each drug over time t, such that:

$$A_{s_x,d,t} = A_{s_x,d,0} 2^{g_{s_x,d} t}$$

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

$$\int_0^T A_{s_x,d,0} 2^{gt} = A_{s_x,d,0} \frac{g_{s_x,d} T - 1}{g_{s_x,d} T \log(2)}$$

We substitute $AUC_{s_x,d}$ for $\int_0^T A_{s_x,d,0} 2^{gt}$ and $f_{s_x,d,0}$ for $A_{s_x,d,0}$. We then numerically solve for the $g_{s_x,d}$ which satisfies this relationship using the optimize() function in R, setting a minimum of 0 and maximum of 10 for the interval. To obtain the resistance for each strain in each drug ($r_{s_x,d}$), $g_{s_x,d}$ is divided by growth in the DMSO control ($g_{s_x,DMSO}$):

$$r_{s_x,d} = \frac{g_{s_x,d}}{g_{s_x,DMSO}}$$

We note that experimental uncertainty in the collected generation times *t* can introduce biases in the estimation of $r_{s_x,d}$, such that resistance estimates from the MAT**a** and MAT**α** pool can be highly correlated, but may differ in range in some drugs (Fig S5). To avoid un-necessary batch effects in $r_{s_x,d}$ estimated from the MAT**a** and MAT**α** pools, we use the line of best fit derived in Fig S5 to rescale $r_{s_x,d}$ estimates from the MAT**a** pool to match the MAT**α** pool before merging.

Finding complex genetic interactions using a general linear model

The multiplicative model of genetic interactions[65] was applied to the r metric. In this model, the expected resistance of a double knockout strain $x\Delta y\Delta$ in a given drug ($\hat{r}_{x\Delta y\Delta,d}$) is the product of the resistances of the corresponding single knockout strains:

$$1) \quad \hat{r}_{x\Delta y\Delta,d} = r_{x\Delta,d} r_{y\Delta,d}$$

To express this model equivalently in an additive form, we can state this relationship as an exponentiated sum of the log-resistances of the single knockouts - $\log(r_{s_i,d}) = l_{s_i,d}$, so that:

$$2)\quad \hat{r}_{x\Delta y\Delta,d} = \exp(l_{x\Delta,d} + l_{y\Delta,d})$$

We defined a two-gene interaction term $\varepsilon_{x\Delta y\Delta,d}$ 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.

$$3)\quad \varepsilon_{x\Delta y\Delta,d} \equiv \log\left(\frac{r_{x\Delta y\Delta,d}}{\hat{r}_{x\Delta y\Delta,d}}\right)$$

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

$$4)\quad r_{x\Delta y\Delta,d} = \exp\left(l_{x\Delta,d} + l_{y\Delta,d} + \varepsilon_{x\Delta y\Delta,d}\right)$$

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

$$5)\quad \hat{r}_{x\Delta y\Delta z\Delta,d} = \exp(l_{x\Delta,d} + l_{y\Delta,d} + l_{z\Delta,d} + \varepsilon_{x\Delta y\Delta,d} + \varepsilon_{x\Delta z\Delta,d} + \varepsilon_{y\Delta z\Delta,d})$$

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

$$6)\quad \varepsilon_{x\Delta y\Delta z\Delta,d} \equiv \log\left(\frac{r_{x\Delta y\Delta z\Delta,d}}{\hat{r}_{x\Delta y\Delta z\Delta,d}}\right)$$

This definition can be extended analogously for interactions of arbitrary complexity, with ε 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 ΔG, which consists of a subset of ABC transporter knockouts $\{ABC1_\Delta \ldots ABC16_\Delta\}$:

$$7)\quad \hat{r}_{\Delta G \mid \Delta G \subseteq \{ABC1_\Delta \ldots ABC16_\Delta\}} = \exp\left(\sum_{i\in\Delta G} l_i + \sum_{j\subseteq\Delta G \mid |j|\geq 2} \varepsilon_j + c\right)$$

To train this model, ΔG 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 ΔG, the relevant l_i coefficients are added only if the corresponding gene i is knocked out, and the ε_j coefficients are added only if all the genes in subset j are knocked out. For each drug, we fit this model using the glm() function in R, with ε terms to a chosen level of complexity.

To perform the marginal association in Fig S4, we fit a model with only l_i terms, 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 $p \leq 0.05/16$. 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 Fig S3, substituting g for r.

We expanded this approach to train models containing ε terms of up to n-way complexity using a "stepwise search" approach. First, we use the marginal association procedure above to initialize the model at $n = 1$. Then, n is incremented by 1, and all possible n-way interactions between the genes contained in the existing (i.e. $n-1$) model are added as additional ε_j features. Each term in this proposed "n-way" model is tested for significance using Type III Sums of Squares ANOVA, those with $p \geq 0.05$ are discarded, and the model is updated. This "stepwise addition" procedure is repeated until either n reaches 5, or the number of genes in the $n-1$ model is less than n (i.e.

there are no more possible interaction terms). After the stepwise addition 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 $p \leq 0.05/k$, where k is the number of all possible 1-5 gene combinations amongst the marginally associated genes.

Defining a non-linear system model

We will define an 'efflux and compensatory activation' 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:

$$r_{norm_d} \equiv \frac{r_d}{\max(r_d)}$$

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

$$\hat{r}_{norm_d} = \frac{1}{1 + e^{k[d]-a}}$$

Here $[d]$ is the concentration of the given drug, and k,a are an unknown constants which define the dose-response curve (such that $\frac{a}{k}$ 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):

$$\hat{r}_{norm_{G,d}} = \frac{1}{1 + e^{k[d]-a-\sum_{i \in G} C_{i,d}}}$$

Here, G is the set of ABC transporters present in a genotype: $\{ABC1^{+} \ldots ABC16^{+}\}$, and $C_{i,d}$ is the clearance coefficient of a given ABC transporter for a given drug (i.e. $C_{i,d} = k[\Delta d_i]$). Importantly, a dose response curve in this form can be expressed as the activation of a sigmoid neuron, where $k[d] - a$ is collapsed into a single bias term B, and $C_{i,d}$ 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 $C_{i,d}$ to be positive.

We then model compensatory activation between ABC transporters. To do this, we first decompose the clearance coefficient of each ABC transporter $C_{i,d}$. That is, each ABC transporter is also given a degree of activity (a value between 0 and 1) which depends on the genotype - $A_{i,G}$. This activation variable is modelled as being dependent on genotype G, but not the drug d. In this extension, $C_{i,d}$ is the product of $A_{i,G}$ and $E_{i,d}$, a 'maximal' efflux/clearance capacity of a given transporter for a given drug ($C_{i,d} = A_{i,G}E_{i,d}$):

$$\hat{r}_{norm_{G,d}} = \frac{1}{1 + e^{-\sum_{i \in G} A_{i,G}E_{i,d}-B}}$$

We then allow $A_{i,G}$ to capture compensatory activation. That is, $A_{i,G}$ can be influenced by other ABC transporters:

$$A_{i,G} = f\left(\sum_{j \in G, j \neq i} I_j\right)$$

Where I_j are the 'influences' from other ABC transporters. While the form of $A_{i,G}$ may depend on the inhibition mechanism, here we also modelled it as having a sigmoidal form for simplicity:

$$A_{i,G} = \frac{1}{1 + e^{-\sum_{j \in G, j \neq i} I_j - B}}$$

Learning a non-linear system model as a neural network

To create the above model and learn the I, E, and A 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 G 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 $\{ABC1 \ldots ABC16\}$.

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 $i - j$, (I_{i-j}), and the second layer acts to compute the activity state A for each transporter. Specifically, we create a second sigmoid layer of length 16, and connect each transporter in the first layer to each transporter j in the second layer, except where $i = j$, as a transporter cannot inhibit itself in this model. The activity state A for each transporter is then computed by the neurons in the second layer in this network from their inbound influence connection I_j, and a learned bias term b_j. Notably, the neural network multiplies each outgoing inhibitory connection I_i by its corresponding genotype value in G, such that all outgoing inhibitory weights from transporter i are set to 0 if it is knocked out. To analogously set the activation state of each transporter in the second layer A_j to 0 if it is knocked out, each neuron in the second layer is then multiplied element-wise by its corresponding value in G using the layer_multiply() function.

To encode the efflux weights for each transporter-drug pair E_{i-d}, 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 only positive E parameters are learned. Each neuron in this third layer predicts the normalized resistance to each compound $\hat{r}_{norm_{G,d}}$ by multiplying the activation state of each transporter A_j with the learned efflux weights E_{i-d} to compute the clearance coefficients $C_{i,d}$ 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 λ was added to both the I weights and the bias term which defines A for each transporter, using the kernel_regularizer and bias_regularizer parameters in the second layer. Regularization on I achieves sparsity in their weights, as it is otherwise possible, for example, to add I to a transporter which has no E weights, thus learning I parameters which are not supported by any phenotypes. Because the clearance coefficient of each gene for each drug is defined by a product $C = AE$, regularization of the bias

term acts to keep A close to 0.5, effectively setting a prior on A. This prior on A avoids parameterizing $C \approx 0$ by setting a large bias such that $A \approx 0$, which then allows E weights to be added to transporters without affecting phenotype predictions. Thus, regularization of A indirectly enforces sparsity in the E parameters, as each E directly impacts resistance predictions when A is not close to 0. While more complex regularization schemes can potentially impose three separate regularization weights for the I terms, the bias on A, and the E terms, here we found that using a single weight for regularizing both I and the bias for A without any further regularization to the E 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 0.1. 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:

$$|Z_{param}| = \frac{|\mu_{param}|}{\sigma_{param}}$$

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 $|Z_{param}| < 4$ 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 10^{-6} to 10^{-1}. We first searched 13 intervals between 10^{-6} to 10^{-1}, and after observing the mean-squared error of the resulting predictions and the number of model parameters, we searched another 11 intervals between 10^{-4} to 10^{-3} (Fig S8 A-B). We chose a regularization rate of 5×10^{-4}, as any rate higher than this appeared to result in a 'spike' in mean-squared error in both the MAT**a** and MAT**α** pools (Fig S8B), while lowering this rate did not have a clear mean-squared error impact but increased the number of non-zero parameters (Fig 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 i strains and j drugs, given the set of k initial non-zero weights $W_{\{1-k\}}$:

$$(\varepsilon_{initial})^2 = \left(r_{norm_{G_{\{1\dots i\}},d_{\{1\dots j\}}}} - \hat{r}_{norm_{G_{\{1\dots i\}},d_{\{1\dots j\}}}} \mid W_{\{1\dots k\}}\right)^2$$

Then, for each $l \in \{1\dots k\}$, we set $W_l := 0$, and compute the squared residuals in the proposed reduced model:

$$(\varepsilon_{reduced})^2 = \left(r_{norm_{G_{\{1-i\}},d_{\{1-j\}}}} - \hat{r}_{norm_{G_{\{1-i\}},d_{\{1-j\}}}} \mid W_{\{1\dots(l-1),(l+1)\dots k\}}\right)^2$$

Considering only data where W_l made a predictive difference ($\varepsilon_{initial} \neq \varepsilon_{reduced}$ at a numerical tolerance of 10^{-4}), we then compute the paired Mann-Whitney U statistic between $(\varepsilon_{initial})^2$ and $(\varepsilon_{reduced})^2$ to derive a p-value for l, and keep all features with $p < 0.05/k$ in the final model.

Targeted mating and selection to obtain 32 knockouts

The TWAS21230902 strain (*pdr10*Δ *pdr18*Δ *pdr5*Δ *snq2*Δ *ybt1*Δ *ycf1*Δ *yor1*Δ*;* Data S2) was subject to individual strain genotyping, confirming the genotype generated using the RCP-PCR based method. This strain (MAT**α**) was mated with RY0146 (MAT**a**), and was subject to sporulation and MAT**a** haploid selection[21]. Individuals from this cross were arrayed onto a 384 well plate, 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 Mix 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 1.3, 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 Fig S11. To determine the fitted IC50 values in Fig 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[25]. 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[25]. 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)[66]. 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[51]. Additional strains acting required to recreated positive and negative controls were also obtained from this screen (Fig 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. qPCR on these samples was then performed using a SensiFAST™ Real-Time PCR Kit and Ct values were quantified using a CFX machine. cDNA synthesis and qPCR was performed for *PDR5* and *UBC6* (acting loading control).

Availability of Data and Materials

Competing Interests

The authors declare that they have no competing interests.

Acknowledgements

This work was supported by XX.

Author Contributions

N.Y, F.P.R & A.C conceived the experiments. N.Y, M.G, L.M, S.Z & T.F performed experiments. A.C and N.Y analyzed the data. 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 strain, perform genotyping, RCP-PCR overhangs, and pool multiplexing primers.

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

Additional Data S3. Drugs used in this study and their concentration in the pooled growth data.

Additional Data S4. List of primer pairs used to multiplex pooled growth sequencing data.

Additional Data S5. Growth and resistance metrics obtained for all strains in both the MAT**a** and MAT**α** pools.

Additional Data S6. Summary of linear modelling results obtained in this study.

Additional Data S7. Previously-known drug knockout associations within the 16 ABC transporters and 16 drugs studied.

Additional Data S8. Functional interpretations of genetic interactions present in the data.

References

1. Benfey, P. N. & Mitchell-Olds, T. From Genotype to Phenotype: Systems Biology Meets Natural Variation. *Science (80-.).* **320,** (2008).
2. Hartwell, L. Robust Interactions. *Science (80-.).* **303,** (2004).
3. Hartman, J. L., Garvik, B. & Hartwell, L. Principles for the Buffering of Genetic Variation. *Science (80-.).* **291,** (2001).
4. Civelek, M. & Lusis, A. J. Systems genetics approaches to understand complex traits. *Nat. Rev. Genet.* **15,** 34–48 (2014).
5. Costanzo, M. *et al.* A global genetic interaction network maps a wiring diagram of cellular function. *Science (80-.).* **353,** (2016).
6. Shen, J. P. & Ideker, T. Synthetic Lethal Networks for Precision Oncology: Promises and Pitfalls. *J. Mol. Biol.* **430,** 2900–2912 (2018).
7. Horlbeck, M. A. *et al.* Mapping the Genetic Landscape of Human Cells. *Cell* **174,** 953–967.e22 (2018).
8. Costanzo, M. *et al.* The genetic landscape of a cell. *Science* **327,** 425–31 (2010).
9. St Onge, R. P. *et al.* Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. *Nat. Genet.* **39,** 199–206 (2007).
10. Braberg, H. *et al.* Quantitative analysis of triple-mutant genetic interactions. *Nat. Protoc.* **9,** 1867–81 (2014).
11. Tong, A. H. Y. *et al.* Global mapping of the yeast genetic interaction network. *Science* **303,** 808–13 (2004).
12. Taylor, M. B., Ehrenreich, I. M., Rothstein, R., Hu, T. & Mast, J. Genetic Interactions Involving Five or More Genes Contribute to a Complex Trait in Yeast. *PLoS Genet.* **10,** e1004324 (2014).
13. Beh, C. T., Cool, L., Phillips, J. & Rine, J. Overlapping functions of the yeast oxysterol-binding protein homologues. *Genetics* **157,** 1117–40 (2001).
14. Wieczorke, R. *et al.* Concurrent knock-out of at least 20 transporter genes is required to block uptake of hexoses in Saccharomyces cerevisiae. *FEBS Lett.* **464,** 123–8 (1999).
15. Mullis, M. N., Matsui, T., Schell, R., Foree, R. & Ehrenreich, I. M. The complex underpinnings of genetic background effects. *Nat. Commun.* **9,** 3548 (2018).
16. Zhang, Y. *et al.* A transportome-scale amiRNA-based screen identifies redundant roles of Arabidopsis ABCB6 and ABCB20 in auxin transport. *Nat. Commun.* **9,** 4204 (2018).
17. Palmer, A. C. *et al.* Delayed commitment to evolutionary fate in antibiotic resistance fitness landscapes. *Nat. Commun.* **6,** 7385 (2015).
18. Cancer Genome Atlas Research Network *et al.* Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. *N. Engl. J. Med.* **368,** 2059–2074 (2013).
19. Heckl, D. *et al.* Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing. *Nat. Biotechnol.* **32,** 941–6 (2014).
20. Takahashi, K. & Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **126,** 663–676 (2006).

21. Suzuki, Y. *et al.* Knocking out multigene redundancies via cycles of sexual assortment and fluorescence selection. *Nat. Methods* **8,** 159–64 (2011).
22. Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460,** 894–8 (2009).
23. DiCarlo, J. E. *et al.* Yeast oligo-mediated genome engineering (YOGE). *ACS Synth. Biol.* **2,** 741–9 (2013).
24. Zeitoun, R. I. *et al.* Multiplexed tracking of combinatorial genomic mutations in engineered cell populations. *Nat. Biotechnol.* **33,** 631–637 (2015).
25. Zeitoun, R. I., Pines, G., Grau, W. C. & Gill, R. T. Quantitative Tracking of Combinatorially Engineered Populations with Multiplexed Binary Assemblies. *ACS Synth. Biol.* **6,** 619–627 (2017).
26. Díaz-Mejía, J. J. *et al.* Mapping DNA damage-dependent genetic interactions in yeast via party mating and barcode fusion genetics. *Mol. Syst. Biol.* **14,** e7985 (2018).
27. Wong, A. S. L. *et al.* Multiplexed barcoded CRISPR-Cas9 screening enabled by CombiGEM. *Proc. Natl. Acad. Sci. U. S. A.* **113,** 2544–9 (2016).
28. Jungwirth, H. & Kuchler, K. Yeast ABC transporters – a tale of sex, stress, drugs and aging. *FEBS Lett.* **580,** 1131–8 (2006).
29. Dean, M., Rzhetsky, A. & Allikmets, R. The human ATP-binding cassette (ABC) transporter superfamily. *Genome Res.* **11,** 1156–66 (2001).
30. Kovalchuk, A. & Driessen, A. J. M. Phylogenetic analysis of fungal ABC transporters. *BMC Genomics* **11,** 177 (2010).
31. Kolaczkowska, A., Kolaczkowski, M., Goffeau, A. & Moye-Rowley, W. S. Compensatory activation of the multidrug transporters Pdr5p, Snq2p, and Yor1p by Pdr1p in Saccharomyces cerevisiae. *FEBS Lett.* **582,** 977–83 (2008).
32. Snider, J. *et al.* Mapping the functional yeast ABC transporter interactome. *Nat. Chem. Biol.* **9,** 565–72 (2013).
33. Donner, M. & Keppler, D. Up-regulation of basolateral multidrug resistance protein 3 (Mrp3) in cholestatic rat liver. *Hepatology* **34,** 351–359 (2001).
34. König, J., Rost, D., Cui, Y. & Keppler, D. Characterization of the human multidrug resistance protein isoform MRP3 localized to the basolateral hepatocyte membrane. *Hepatology* **29,** 1156–1163 (1999).
35. Huls, M. *et al.* The breast cancer resistance protein transporter ABCG2 is expressed in the human kidney proximal tubule apical membrane. *Kidney Int.* **73,** 220–225 (2008).
36. Bloom, J. S., Ehrenreich, I. M., Loo, W. T., Lite, T.-L. V. & Kruglyak, L. Finding the sources of missing heritability in a yeast cross. *Nature* **494,** 234–7 (2013).
37. Brem, R. B. & Kruglyak, L. The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 1572–7 (2005).
38. Perlstein, E. O., Ruderfer, D. M., Roberts, D. C., Schreiber, S. L. & Kruglyak, L. Genetic basis of individual differences in the response to small-molecule drugs in yeast. *Nat. Genet.* **39,** 496–502 (2007).
39. Lee, A. Y. *et al.* Mapping the cellular response to small molecules using chemogenomic fitness signatures. *Science* **344,** 208–11 (2014).
40. Kebschull, J. M. & Zador, A. M. Cellular barcoding: lineage tracing, screening and beyond. *Nat. Methods* **15,** 871–879 (2018).
41. Yan, Z. *et al.* Yeast Barcoders: a chemogenomic application of a universal donor-strain collection carrying bar-code identifiers. *Nat. Methods* **5,** 719–725 (2008).

42. Smith, A. M. *et al.* Quantitative phenotyping via deep barcode sequencing. *Genome Res.* **19,** 1836–42 (2009).
43. Giaever, G. *et al.* Functional profiling of the Saccharomyces cerevisiae genome. *Nature* **418,** 387–391 (2002).
44. Yachie, N. *et al.* Pooled-matrix protein interaction screens using Barcode Fusion Genetics. *Mol. Syst. Biol.* **12,** 863 (2016).
45. Smith, A. M. *et al.* Highly-multiplexed barcode sequencing: an efficient method for parallel analysis of pooled samples. *Nucleic Acids Res.* **38,** e142 (2010).
46. KOLACZKOWSKI, M., KOLACZKOWSKA, A., LUCZYNSKI, J., WITEK, S. & GOFFEAU, A. *In Vivo* Characterization of the Drug Resistance Profile of the Major ABC Transporters and Other Components of the Yeast Pleiotropic Drug Resistance Network. *Microb. Drug Resist.* **4,** 143–158 (1998).
47. Shekhar-Guturja, T. *et al.* 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 (2016).
48. Katzmann, D. J., Burnett, P. E., Golin, J., Mahé, Y. & Moye-Rowley, W. S. Transcriptional control of the yeast PDR5 gene by the PDR3 gene product. *Mol. Cell. Biol.* **14,** 4653–61 (1994).
49. Ernst, R. *et al.* A mutation of the H-loop selectively affects rhodamine transport by the yeast multidrug ABC transporter Pdr5. *Proc. Natl. Acad. Sci.* **105,** 5069–5074 (2008).
50. Khakhina, S. *et al.* Control of Plasma Membrane Permeability by ABC Transporters. *Eukaryot. Cell* **14,** 442–453 (2015).
51. Tarassov, K. *et al.* An in vivo map of the yeast protein interactome. *Science* **320,** 1465–70 (2008).
52. Braun, P. *et al.* An experimentally derived confidence score for binary protein-protein interactions. *Nat. Methods* **6,** 91–97 (2009).
53. Newman, J. R. S. *et al.* Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* **441,** 840–846 (2006).
54. Shaw, W. M. *et al.* Engineering a model cell for rational tuning of GPCR signaling. *bioRxiv* 390559 (2018). doi:10.1101/390559
55. C. elegans Deletion Mutant Consortium. Large-Scale Screening for Targeted Knockouts in the Caenorhabditis elegans Genome. *G3 Genes,Genomes,Genetics* **2,** 1415–1425 (2012).
56. Thompson, O. *et al.* The million mutation project: A new approach to genetics in Caenorhabditis elegans. *Genome Res.* **23,** 1749–1762 (2013).
57. Tanay, A. & Regev, A. Scaling single-cell genomics from phenomenology to mechanism. *Nature* **541,** 331–338 (2017).
58. Zupan, B. *et al.* GenePath: a system for inference of genetic networks and proposal of genetic experiments. *Artif. Intell. Med.* **29,** 107–30
59. Ma, J. *et al.* Using deep learning to model the hierarchical structure and function of a cell. *Nat. Methods* **15,** 290–298 (2018).
60. Brockmann, M. *et al.* Genetic wiring maps of single-cell protein states reveal an off-switch for GPCR signalling. *Nature* **546,** 307–311 (2017).
61. Emanuel, G., Moffitt, J. R. & Zhuang, X. High-throughput, image-based screening of genetic variant libraries. *bioRxiv* (2017).
62. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred

kilobases. *Nat. Methods* **6,** 343–5 (2009).
63. Gietz, R. D. & Schiestl, R. H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* **2,** 31–34 (2007).
64. Proctor, M. *et al.* in 239–269 (Humana Press, 2011). doi:10.1007/978-1-61779-173-4_15
65. Mani, R., St Onge, R. P., Hartman, J. L., Giaever, G. & Roth, F. P. Defining genetic interaction. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 3461–6 (2008).
66. Snider, J. *et al.* Detecting interactions with membrane proteins using a membrane two-hybrid assay in yeast. *Nat. Protoc.* **5,** 1281–1293 (2010).

Figures

Figure 1. Overview of the engineered population profiling process.
We created a barcoded wild-type pool (Fig S1) to enable construction of an engineered population using any multi-mutant strain. In this study, this pool was mated with a 16 ABC transporter knockout strain (ABC-16). The genotype at 16 ABC transporters is indicated by the squares drawn in each cell (black = knockout, white = wild type). Diploids from this cross were subject to sporulation and barcoded haploids were then selected. Each haploid inherits either a wild-type or knockout allele at these 16 loci. Single colonies were picked and arrayed onto a series of 384-well plates. *En masse* genotyping was performed on this collection using an RCP-PCR[44] strategy, which uses a combination of row and column tags to allow identification of PCR products arising from the same well in each plate (Methods). An additional PCR reaction adds a plate tag (not shown). High throughput sequencing of pooled RCP-PCR products allows large scale genotyping and identification of a strain-specific DNA barcode for many strains. Strains with a sucessfully determined barcode and genotype are transferred into two liquid pools based on mating type (MAT**a** or MAT**α**), and grown under each of 16 drugs, as well as a solvent control. High throughput sequencing of strain-specific DNA barcodes at t = 0, 5, 10, 15, and 20 generations of growth reconstructs the resistance of each strain to each drug, profiling the engineered population.

Figure 2. An exploration and assessment of multi-knockout fitness landscapes within a 6-gene group.
A Comparison of MAT**a** and MAT**α** group resistance profiles in camptothecin and ketoconazole. Individuals were grouped by their genotype at *pdr5*Δ, *snq2*Δ, *ybt1*Δ, *ycf1*Δ, and *yor1*Δ. The 5-locus genotype of each group is indicated by the legend. Individuals in each group vary at the remaining 11 loci. Each point represents the mean resistance of the indicated group in the MAT**a** pools (x-axis) and MAT**α** pools (y-axis). Profiles for the remaining drugs are shown in Fig S5.
B Distribution of MAT**a**–MAT**α** correlations of the grouped resistance profiles amongst all drugs tested.
C A radial landscape of benomyl resistance. The graph is centered by the 5-gene wild-type group, with outward extensions adding cumulative knockouts. Each section is coloured by the average resistance of the corresponding 5-gene group relative to the 5-gene wild type. Extensions to 1, 2, and 5 total knockouts are shown. Sections are coloured by the mean resistance of each group relative to the 5-gene wild-type. The colour scale is centered by the mean 5-gene wild-type resistance and extends by half of the observed difference between the 5% and 95% percentile resistance in that drug in both directions (blue for increased resistance, orange for decreased resistance).

D As in B, showing radial fitness landscapes for 10 additional drugs. The remaining 5 drugs are shown in Fig S6.

Figure 3. Exploration and formalization of surprising multi-gene knockout phenotypes.
A A linear landscape of resistance to benomyl, mitoxantrone, and valinomycin in amongst 5-gene groups. The 5-gene genotype of each group is indicated by the legend. Groups are arranged on the x-axis by the number of knockouts (with jitter added to improve clarity), and the y-axis by average drug resistance. Groups separated 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. Linear landscapes for all pools are drawn in Fig S7.
B Distribution of valinomycin resistance amongst all *ybt1*Δ, *yor1*Δ, *snq2*Δ, *ycf1*Δ, and *pdr5*Δ knockout groups. Group genotype is indicated for each line using the same legend as in A). All *pdr5*Δ groups (dark blue) are paired with their corresponding *PDR5+* equivalent (grey).
C A linear model was used to formally determine significant gene knockout and genetic interaction effects mediating resistance to the tested drugs (see Methods). Linear model terms which were significant (Bonferroni adjusted $p < 0.05$) in a given drug are coloured according the legend on the right. Maximum and minimum scale values are determined by the median absolute deviation of the log(resistance) in that drug. Non-significant terms are coloured in grey. ε terms represent n-way interactions (see Methods). Coefficients are sorted by term complexity. Term complexity is also indicated by the grey colour scale on the top of the heatmap.

Figure 4. Modeling and interpreting a complex genetic landscape.
A A neural network model was created to infer transporter-drug and transporter-transporter relationships from the engineered population profiles. The 16-transporter genotype (***G***), is given as input to the model as a binary variable (1 = presence, 0 = absence for each transporter) , and the activity of each transporter (***A***) is computed by the set of learned transporter-transporter influence weights (***I***), and is multiplied element-wise by ***G***. Resistance to each of the 16 tested compounds (***R***) is then computed by transporter-drug efflux weights (***E***). Appropriate weights for ***I*** and ***E*** are learned using stochastic gradient descent and backpropagation using the engineered population profiling data such that mean-squared error is minimized between ***R*** and measured resistance. In addition, a positive constraint is placed on ***E*** and regularization is added to the model (Methods).
B Weights learned by the neural network model after training and pruning are shown. All non-zero ***I*** weights learned by the model were negative.
C Comparing the normalized resistance of each strain measured by engineered population profiling to resistances modelled by the neural network.
D Comparing the neural model in valinomycin to the observed resistances for each five-gene knockout group. The neural network weights (top) are shown for the original model (top-left) and one trained with an extra always-present node in the activity layer to model potential influence of a hidden resistance factor (top right). At the bottom, strains were grouped by knockout genotypes at *pdr5*Δ, *snq2*Δ, *ybt1*Δ, *ycf1*Δ, and *yor1*Δ. Each point represents the mean resistance of a group of strains containing the 5-locus genotype indicated by the legend, either as modeled by the corresponding neural network (x-axis) or as measured in the data (y-axis). Correlation in the top left is shown for all data, then only for the 5-locus groups in parentheses.

Figure 5. Further modeling and exploring of ABC-16 mediated fluconazole resistance.
A As in Figure 3A, a linear landscape of fluconazole resistance is shown .
B Comparing the neural model in fluconazole to the observed resistances for each five-gene knockout group. The neural network weights (top) are shown for the original model (top-left) and one trained with an extra always-present 'hidden' node between the ***G*** and ***A*** layer to model potential non-linear influence of Pdr5 (see Methods for details, top right). At the bottom, strains were grouped by knockout genotypes at *pdr5*Δ, *snq2*Δ, *ybt1*Δ, *ycf1*Δ, and *yor1*Δ. Each point represents the mean resistance of a group of strains containing the 5-locus genotype indicated by the legend, either as modeled by the corresponding neural network (x-axis) or as measured in the data (y-axis). Correlation in the top left is shown for all data, then only for the 5-locus groups in parentheses.
C Comparing the IC50 of fluconazole derived from single-strain growth experiments to the normalized resistance expected by in the grouped pool data (mean resistance is shown for each group). Strain genotype is indicated by the legend.
D Measuring the mRNA expression of *PDR5* in wild-type (RY0566), *ybt1*Δ*ycf1*Δ, *snq2*Δ*yor1*Δ, and *snq2*Δ*yor1*Δ*ybt1*Δ*ycf1*Δ strains. *PDR5* mRNA expression was measured using qRT-PCR and normalized relative to *UBC6*. Values represent the ratio of *PDR5* expression compared to the average in the wild-type. Error bars indicate standard deviation. Three replicates were used in each experiment. p-values were calculated using a t-test. Overlaid are the corresponding Pdr5 activity values from the neural network in the top-right panel of Figure 5B, considering only influences going through the hidden node (yellow), or all influences (orange).
E Comparing the modeled *PDR5* repression by *YOR1* and *SNQ2* with with protein-protein interactions found using MYTH and PCA. Interactions were measured in both this study (Fig S11, S12) and previous studies[32,51]. Learned ***I*** weights from 4B are overlaid.

Figure S1. Creation of a parent barcoder pool.
A Engineering of 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 then 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.
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 known reference 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 grey. 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 grey.

Figure S3. Reproducible marginal gene knockout growth effects in the pool.
A linear model was used to formally determine significant gene knockout effects mediating growth in 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 coloured according the legend on the left. Other terms are coloured in grey.

Figure S4. Reproducible marginal gene knockout resistance effects in the pool.
A linear model was used to formally determine significant gene knockout effects mediating resistance to 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 coloured according the legend on the left. Other terms are coloured in grey.

Figure S5. Reproducibility of grouped genotype resistance.
Strains were grouped on knockout genotypes at *pdr5*Δ, *snq2*Δ, *ybt1*Δ, *ycf1*Δ, and *yor1*Δ. Each point represents a group of strains containing the 5-locus genotype indicated by the legend. Strains in each group vary at the remaining 11 loci. Each point represents the mean resistance of each group in the MAT**a** (x-axis) and MAT**α** (y-axis) pools.

Figure S6. A radial fitness landscape in six additional drugs.
A radial fitness landscape in six drugs showing all multi-knockout paths. Each graph is centered by the 5-gene wild-type group, with outward extensions adding cumulative knockouts. Each section is coloured by the average resistance of its corresponding knockout group relative to the 5-gene wild type. Extensions to 1, 2, and 5 total knockouts are shown. Sections are coloured by the mean resistance of each group relative to the 5-gene wild-type. The colour scale is centered by the mean 5-gene wild-type resistance and extends by half of the observed difference between the 5% and 95% percentile resistance in that drug in both directions (blue for increased resistance, orange for decreased resistance).

Figure S7. A linear landscape of resistance to 16 drugs.
A A linear landscape of resistance to all tested drugs in the amongst 5-gene groups. The 5-gene genotype of each group is indicated by the legend. Groups are arranged on the x-axis by the number of knockouts (with jitter added to improve clarity), and on the y-axis by average drug resistance. Groups separated by single knockouts 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.

Figure S8. Neural network evaluation
A Number of reproducible network parameters ($Z \geq 4$ estimated from 10 iterations, Methods) as a function of the regularization rate λ. 13 intervals are plotted from 10^{-6} to 10^{0} (left), and 11 intervals are plotted from 10^{-4} to 10^{-3} (right). Values between intervals are linearly interpolated.

B As in S8A, 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 modelled 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 S9. Neural network performance for single drugs
Strains were grouped on knockout genotypes at *pdr5*Δ, *snq2*Δ, *ybt1*Δ, *ycf1*Δ, and *yor1*Δ. Each point represents the mean resistance of a group of strains containing the 5-locus genotype indicated by the legend, either as modeled by the neural network (x-axis) or as measured in the data (y-axis). Correlation in the top left is shown for all data, then only for the 5-locus groups in parentheses.

Figure S10. Neural networks trained in single environments
A Comparing the neural model in valinomycin to the observed resistances for each five-gene knockout group. The neural network weights (top) are shown for a model trained only on valinomycin data. At the bottom, strains were grouped by knockout genotypes at *pdr5*Δ, *snq2*Δ, *ybt1*Δ, *ycf1*Δ, and *yor1*Δ. Each point represents the mean resistance of a group of strains containing the 5-locus genotype indicated by the legend, either as modeled by the corresponding neural network (x-axis) or as measured in the data (y-axis). Correlation in the top left is shown for all data, then only for the 5-locus groups in parentheses.
B As in A, showing a neural network trained only on fluconazole data.

Figure S11. Comparing drug resistance measured from single-strain experiments to the grouped pool data.
Resistance of individual strains containing each of 32 knockout combinations at *pdr5*Δ, *snq2*Δ, *ybt1*Δ, *ycf1*Δ, and *yor1*Δ was measured and compared to the resistance to the pool data. Pool strains were grouped based on genotype at these 5 loci, median $\log_2$-resistance was determined for each group in MAT**a** and MAT**α** pools, and these values were averaged to obtain a single pool value. Strain genotype is indicated by the legend. Growth of individual strains was measured at 1.9, 3.9, 7.8, 15.6, 23.4, 31.2, 35, and 40μm of fluconazole.

Figure S12. Measuring all protein-protein interactions between Pdr5, Snq2, and Yor1 using mDHFR PCA.
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[51]. 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 S13. Measuring protein-protein interactions of Pdr5 with Snq2 and Yor1 using MYTH. NubG-PDR5, NubI-PDR5, Ost1-NubG, and Ost1-NubI strains were retrieved from a previously constructed genomic prey library[25] and were each transformed with plasmids containing clones of *PDR5*, *YOR1*, *SNQ2*, or an artificial bait fused to Cub (YOR1-L2, PDR5-L2, SNQ2-L2, Artificial L2 bait). 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.

Figure S1


A
Desired barcode cassette for in-yeast-assembly
5' SacI
LoxP
US1
US2
HphMX4
DS1
DS2
Lox2272
SacI 3'
UP tag
DN tag
Size: 1732 bp
Step 1: Amplify the HphMX4 cassette from pIS420 adding the US2 and DS1 components of the UP and DN tag
Size: 1826 bp
Step 2: Add UP and DN tag barcodes (25 random bp), and US1/DS2 regions
Size: 1958 bp
Step 3: Add LoxP/Lox2272 sequences and homology to pSH47 SacI site

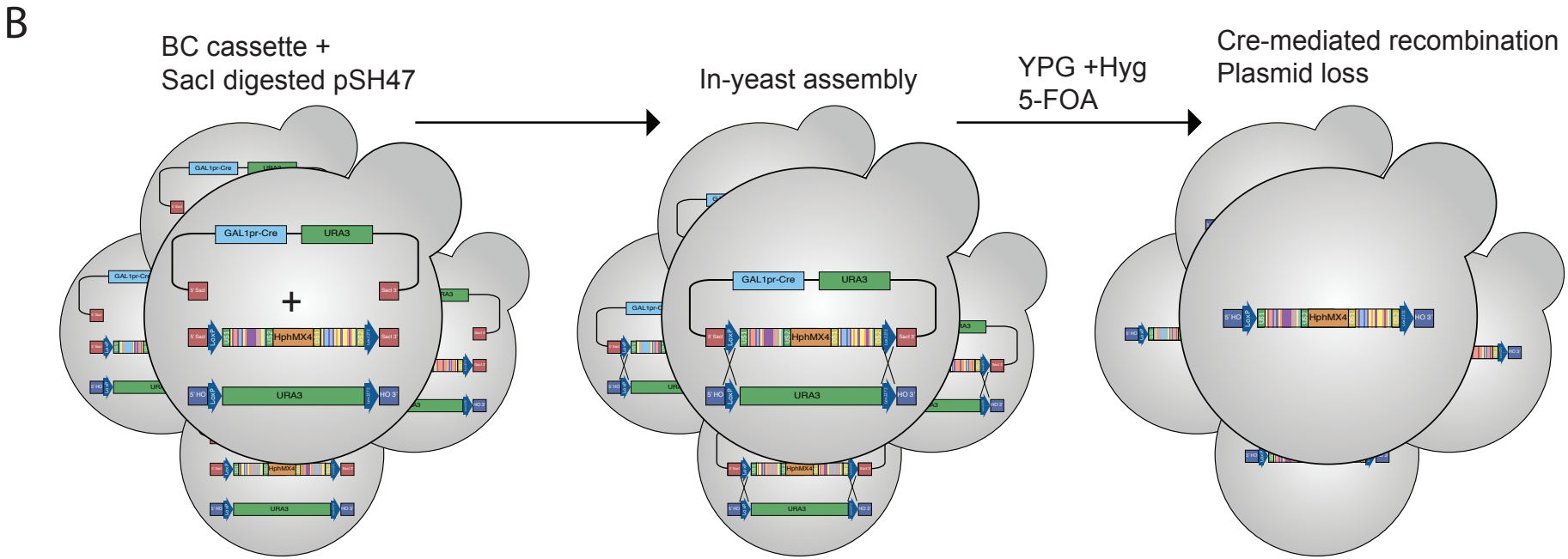

B
BC cassette +
SacI digested pSH47
In-yeast assembly
YPG +Hyg
5-FOA
Cre-mediated recombination
Plasmid loss
GAL1pr-Cre
URA3
HphMX4

Figure S2

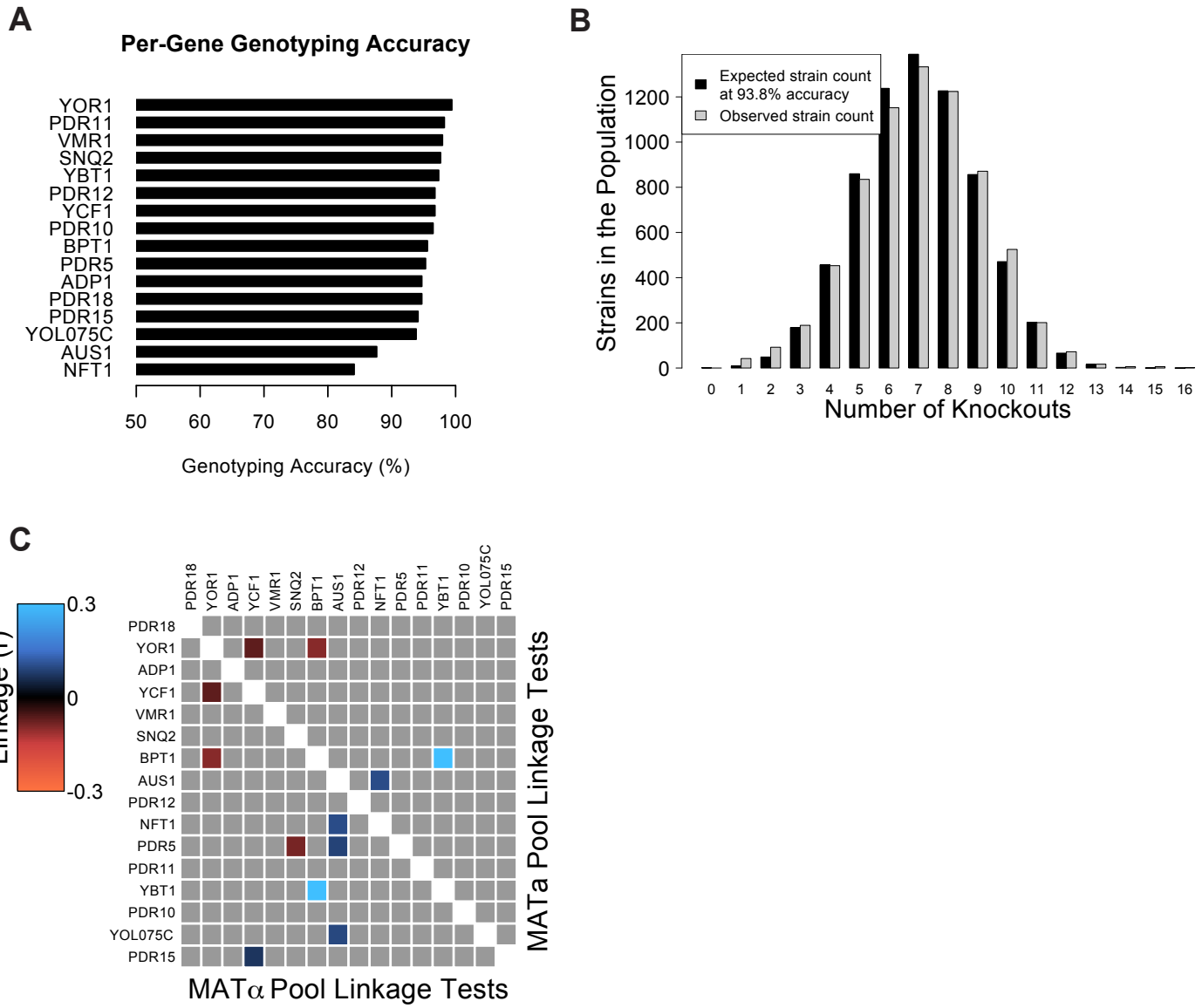

A
Per-Gene Genotyping Accuracy
YOR1
PDR11
VMR1
SNQ2
YBT1
PDR12
YCF1
PDR10
BPT1
PDR5
ADP1
PDR18
PDR15
YOL075C
AUS1
NFT1
50
60
70
80
90
100
Genotyping Accuracy (%)
B
Expected strain count at 93.8% accuracy
Observed strain count
Strains in the Population
0
200
400
600
800
1000
1200
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Number of Knockouts
C
Linkage (r)
0.3
0
-0.3
PDR18 YOR1 ADP1 YCF1 VMR1 SNQ2 BPT1 AUS1 PDR12 NFT1 PDR5 PDR11 YBT1 PDR10 YOL075C PDR15
MATa Pool Linkage Tests
MATα Pool Linkage Tests

Figure S3

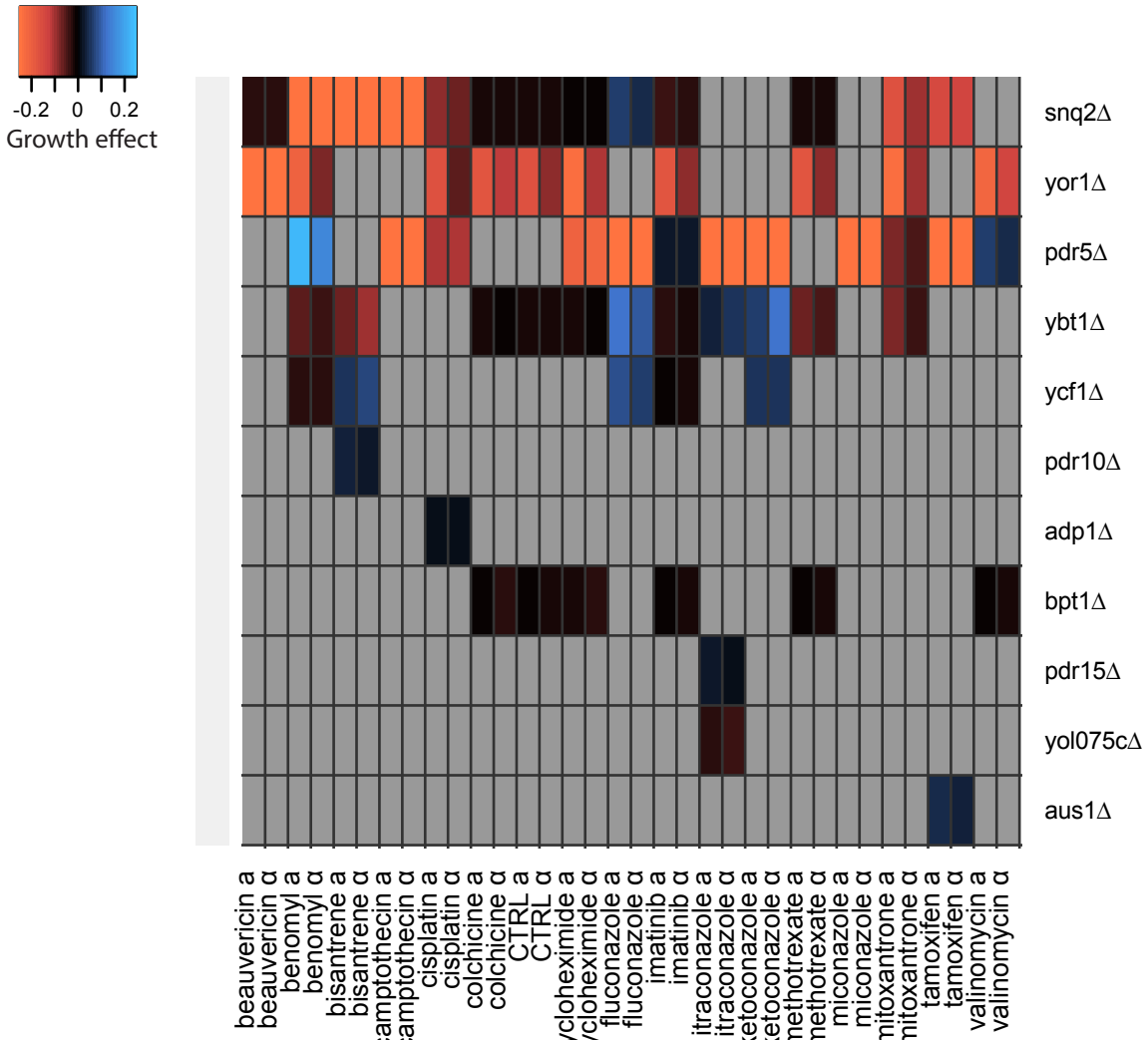

-0.2
0
0.2
Growth effect
snq2Δ
yor1Δ
pdr5Δ
ybt1Δ
ycf1Δ
pdr10Δ
adp1Δ
bpt1Δ
pdr15Δ
yol075cΔ
aus1Δ
beauvericin a
beauvericin α
benomyl a
benomyl α
bisantrene a
bisantrene α
camptothecin a
camptothecin α
cisplatin a
cisplatin α
colchicine a
colchicine α
CTRL a
CTRL α
cycloheximide a
cycloheximide α
fluconazole a
fluconazole α
imatinib a
imatinib α
itraconazole a
itraconazole α
ketoconazole a
ketoconazole α
methotrexate a
methotrexate α
miconazole a
miconazole α
mitoxantrone a
mitoxantrone α
tamoxifen a
tamoxifen α
valinomycin a
valinomycin α

Figure S4

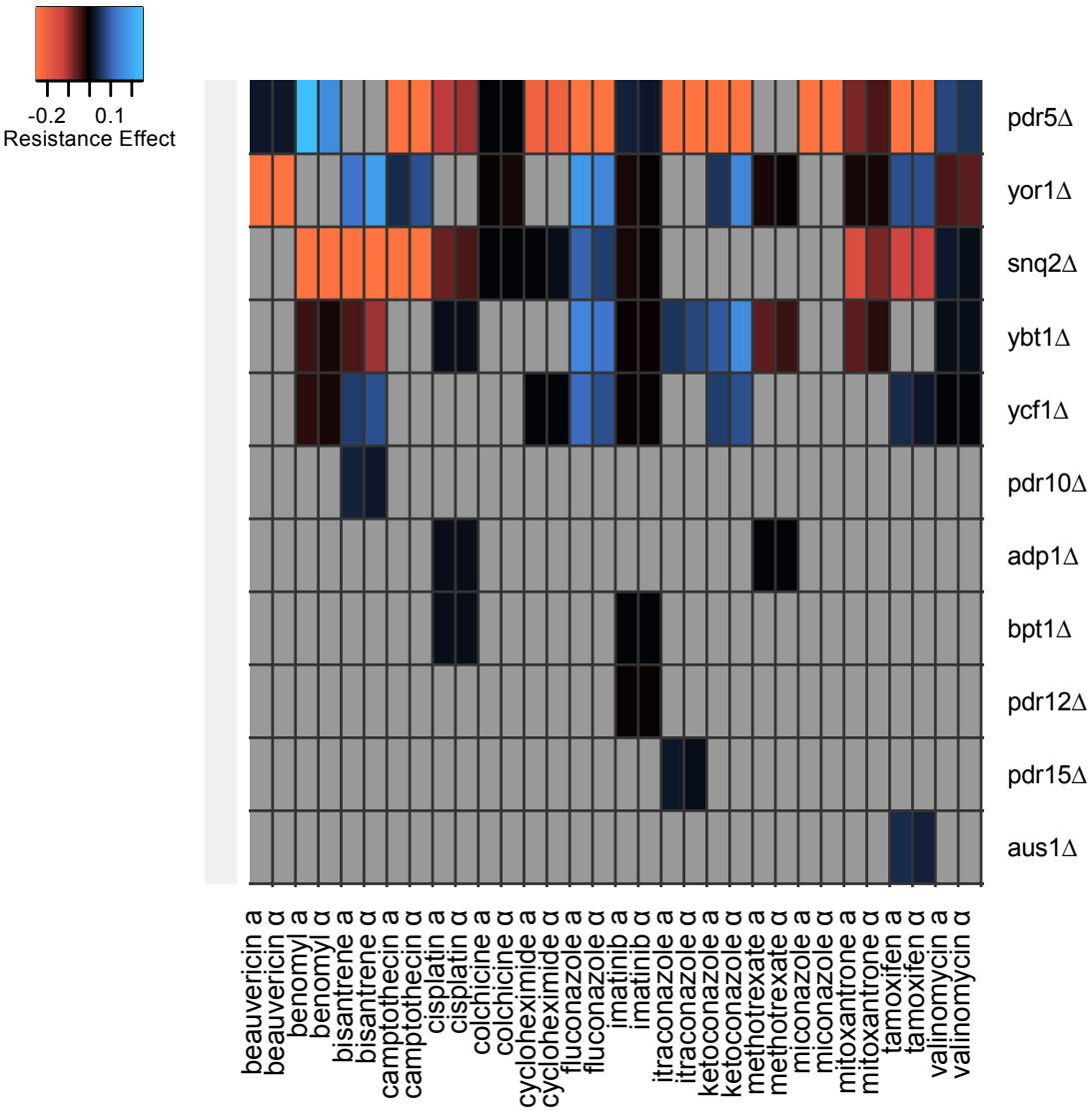
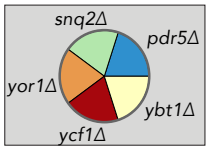

-0.2
0.1
Resistance Effect
pdr5Δ
yor1Δ
snq2Δ
ybt1Δ
ycf1Δ
pdr10Δ
adp1Δ
bpt1Δ
pdr12Δ
pdr15Δ
aus1Δ
beauvericin a
beauvericin α
benomyl a
benomyl α
bisantrene a
bisantrene α
camptothecin a
camptothecin α
cisplatin a
cisplatin α
colchicine a
colchicine α
cycloheximide a
cycloheximide α
fluconazole a
fluconazole α
imatinib a
imatinib α
itraconazole a
itraconazole α
ketoconazole a
ketoconazole α
methotrexate a
methotrexate α
miconazole a
miconazole α
mitoxantrone a
mitoxantrone α
tamoxifen a
tamoxifen α
valinomycin a
valinomycin α

Figure S5


snq2Δ
pdr5Δ
yor1Δ
ybt1Δ
ycf1Δ

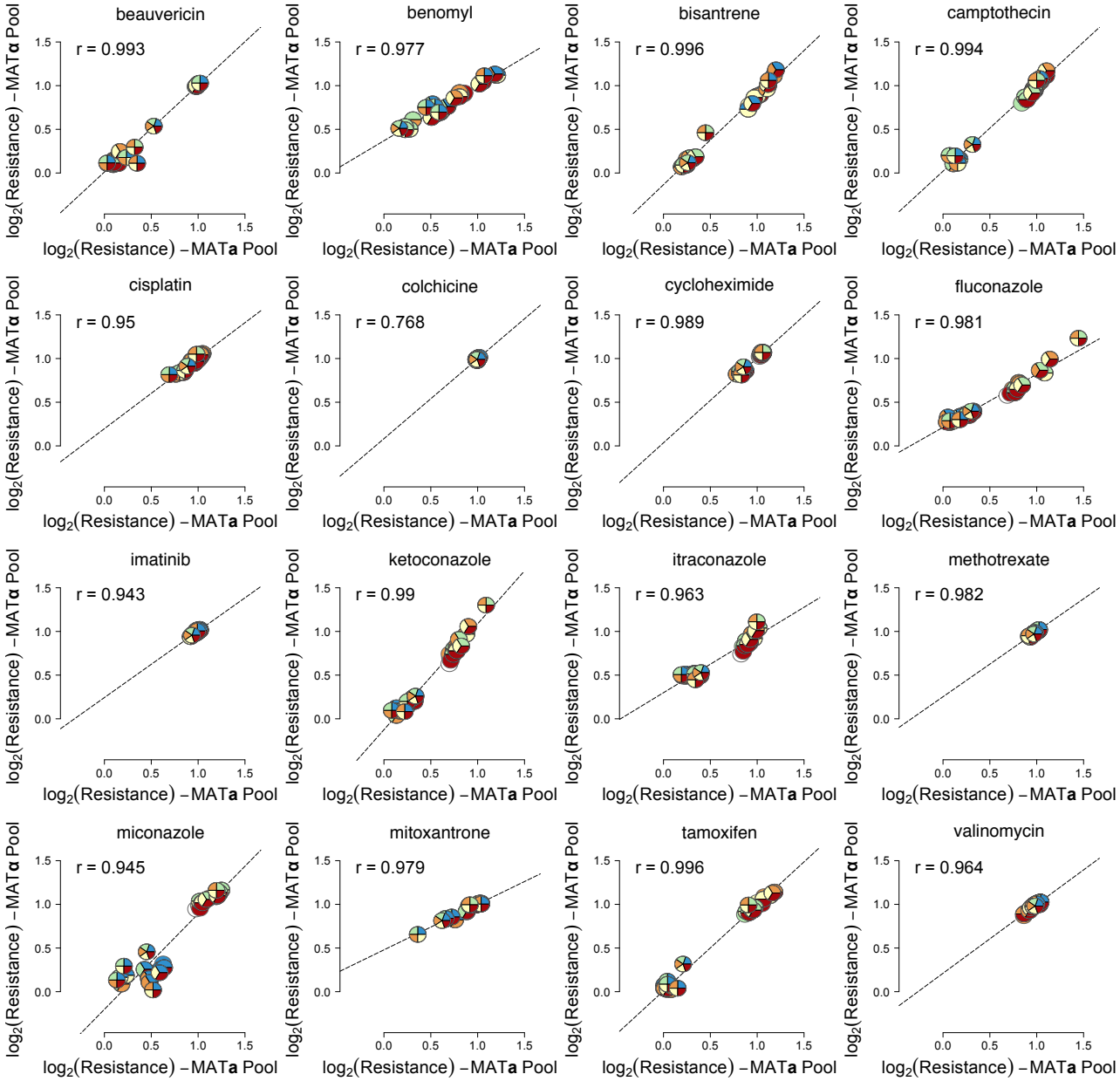

beauvericin
r = 0.993
benomyl
r = 0.977
bisantrene
r = 0.996
camptothecin
r = 0.994
cisplatin
r = 0.95
colchicine
r = 0.768
cycloheximide
r = 0.989
fluconazole
r = 0.981
imatinib
r = 0.943
ketoconazole
r = 0.99
itraconazole
r = 0.963
methotrexate
r = 0.982
miconazole
r = 0.945
mitoxantrone
r = 0.979
tamoxifen
r = 0.996
valinomycin
r = 0.964
log2(Resistance) – MATα Pool
log2(Resistance) – MATa Pool
0.0
0.5
1.0
1.5

Figure S6

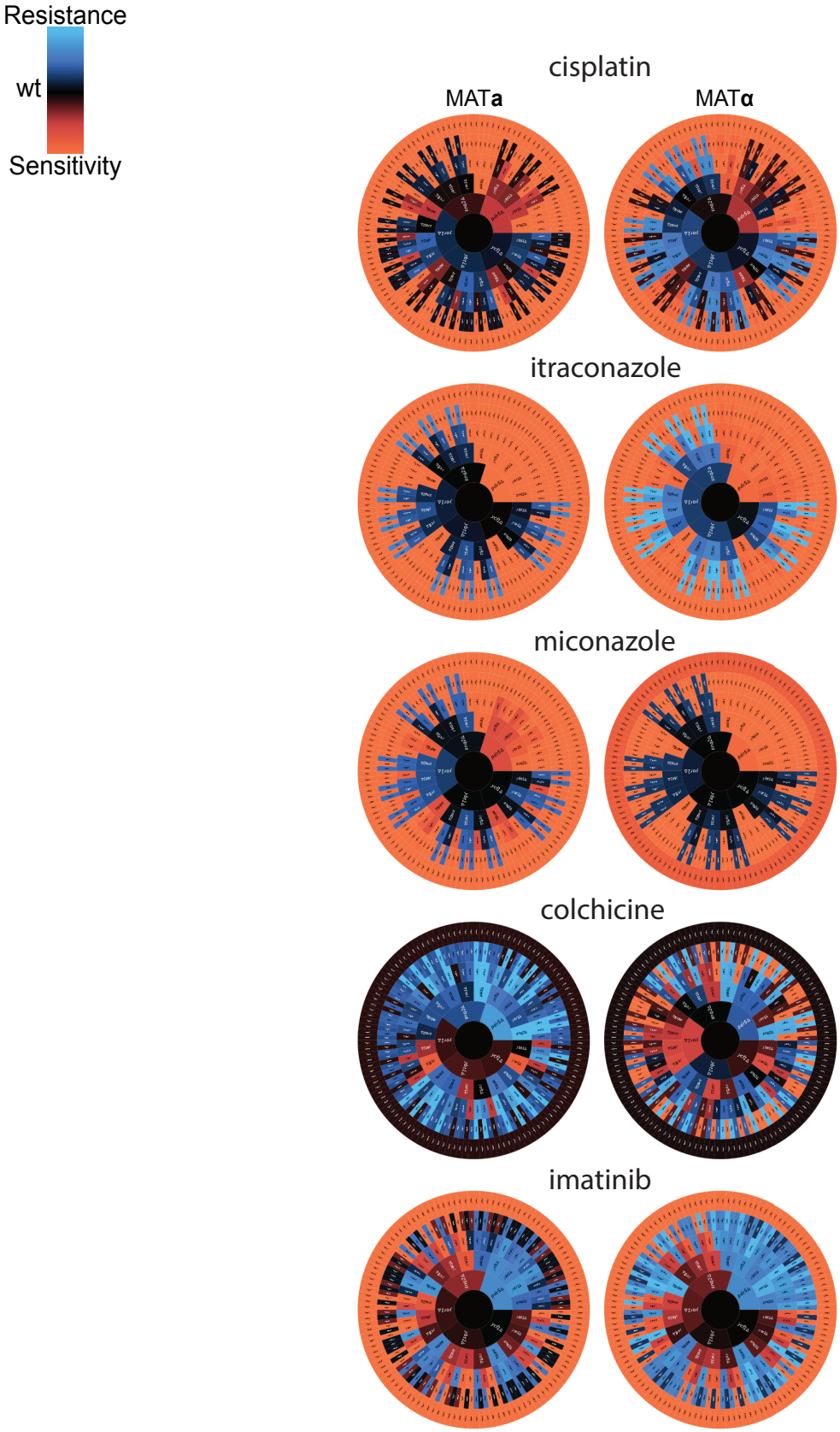

Resistance
wt
Sensitivity
cisplatin
MATa
MATα
itraconazole
miconazole
colchicine
imatinib

Figure S7

A


snq2Δ
pdr5Δ
yor1Δ
ybt1Δ
ycf1Δ

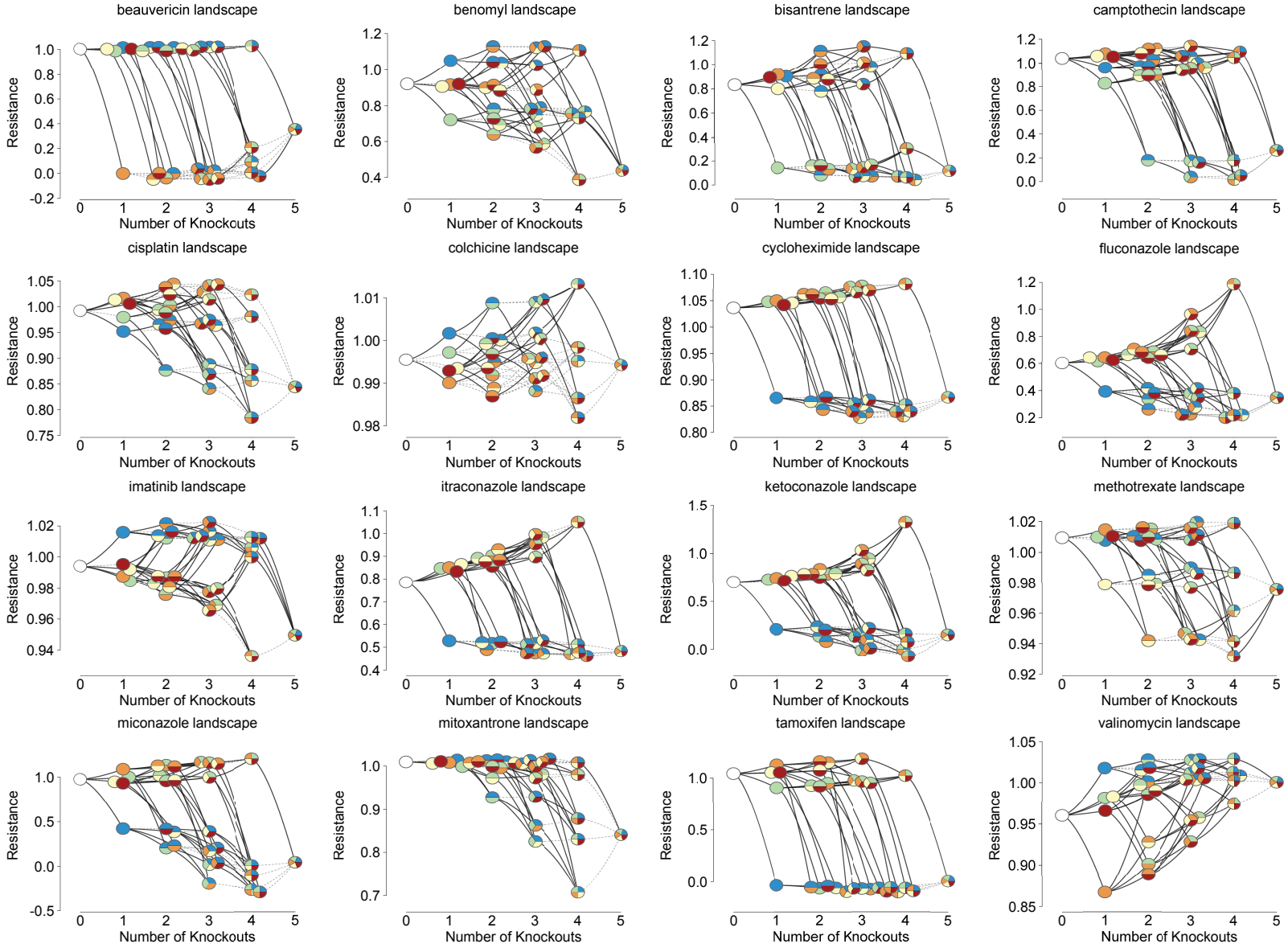

beauvericin landscape
benomyl landscape
bisantrene landscape
camptothecin landscape
cisplatin landscape
colchicine landscape
cycloheximide landscape
fluconazole landscape
imatinib landscape
itraconazole landscape
ketoconazole landscape
methotrexate landscape
miconazole landscape
mitoxantrone landscape
tamoxifen landscape
valinomycin landscape
Resistance
Number of Knockouts

Figure S8

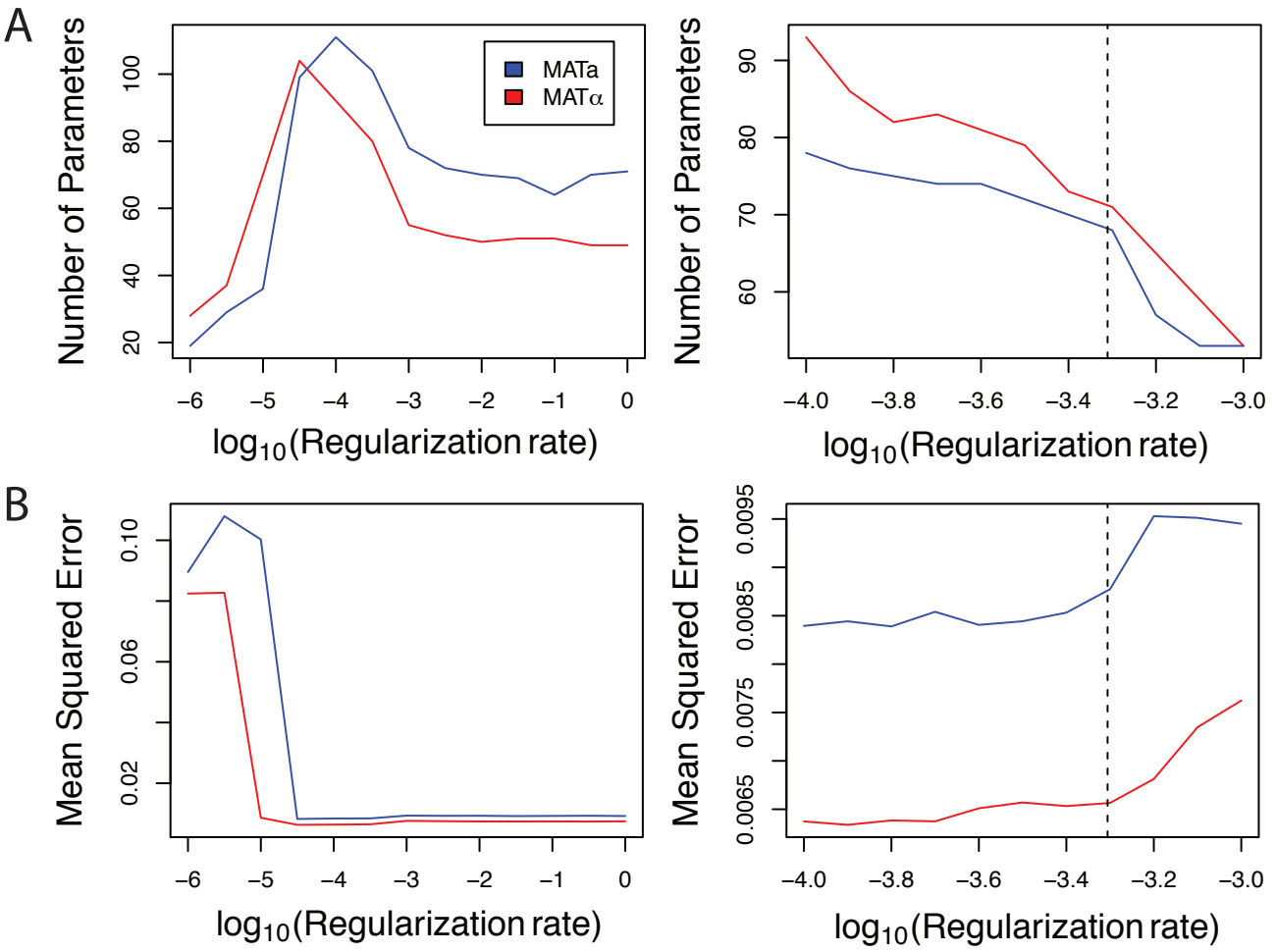

A
Number of Parameters
MATa
MATα
log10(Regularization rate)
B
Mean Squared Error
log10(Regularization rate)
C
r = 0.95
r = 0.96
Measured Resistance (MATa)
Measured Resistance (MATα)
Modelled Resistance (MATa)
Modelled Resistance (MATα)
D
r = 0.98
Model weights (MATα)
Model weights (MATa)

Figure S9

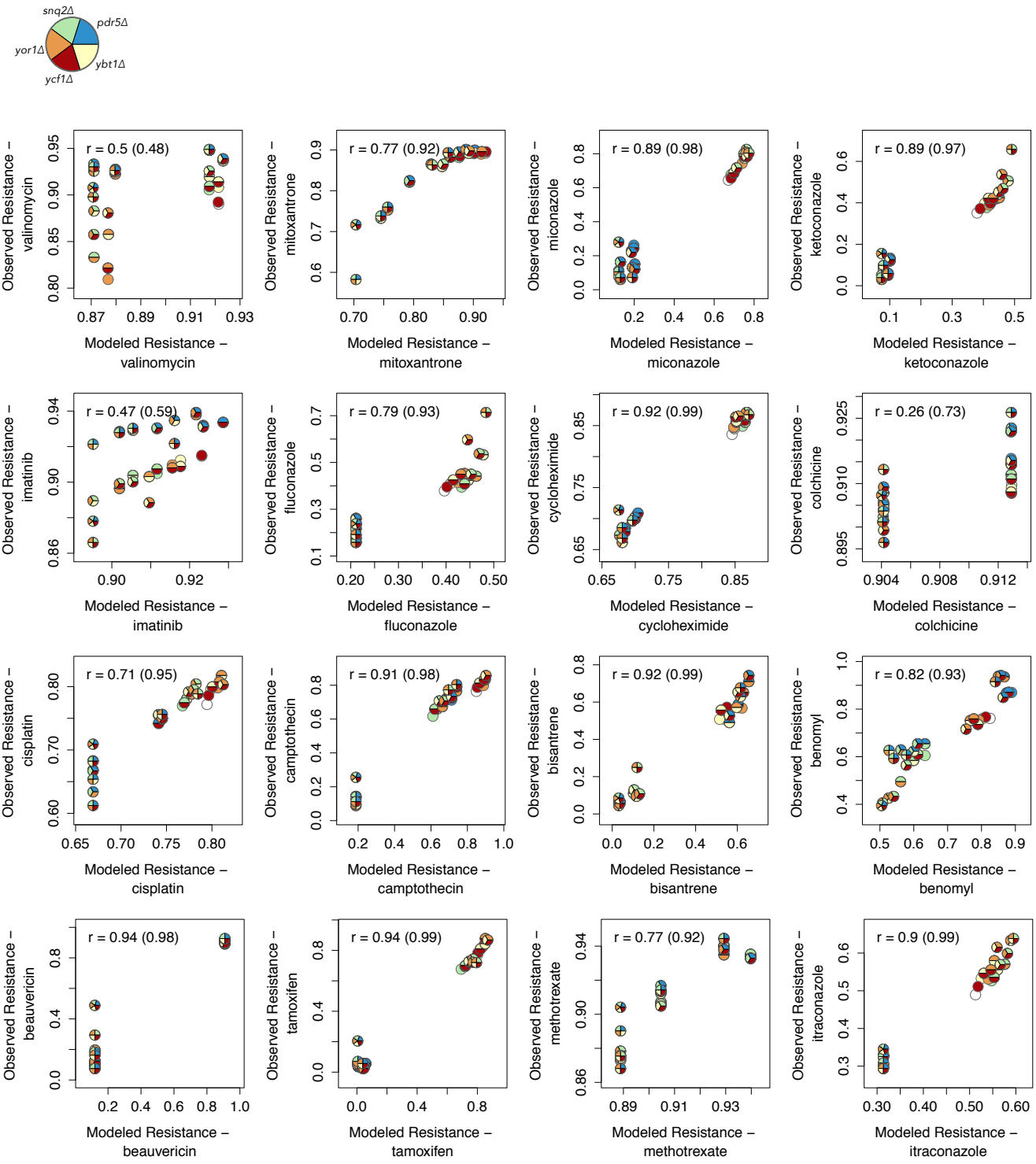

snq2Δ
pdr5Δ
yor1Δ
ybt1Δ
ycf1Δ
r = 0.5 (0.48)
Observed Resistance – valinomycin
Modeled Resistance – valinomycin
r = 0.77 (0.92)
Observed Resistance – mitoxantrone
Modeled Resistance – mitoxantrone
r = 0.89 (0.98)
Observed Resistance – miconazole
Modeled Resistance – miconazole
r = 0.89 (0.97)
Observed Resistance – ketoconazole
Modeled Resistance – ketoconazole
r = 0.47 (0.59)
Observed Resistance – imatinib
Modeled Resistance – imatinib
r = 0.79 (0.93)
Observed Resistance – fluconazole
Modeled Resistance – fluconazole
r = 0.92 (0.99)
Observed Resistance – cycloheximide
Modeled Resistance – cycloheximide
r = 0.26 (0.73)
Observed Resistance – colchicine
Modeled Resistance – colchicine
r = 0.71 (0.95)
Observed Resistance – cisplatin
Modeled Resistance – cisplatin
r = 0.91 (0.98)
Observed Resistance – camptothecin
Modeled Resistance – camptothecin
r = 0.92 (0.99)
Observed Resistance – bisantrene
Modeled Resistance – bisantrene
r = 0.82 (0.93)
Observed Resistance – benomyl
Modeled Resistance – benomyl
r = 0.94 (0.98)
Observed Resistance – beauvericin
Modeled Resistance – beauvericin
r = 0.94 (0.99)
Observed Resistance – tamoxifen
Modeled Resistance – tamoxifen
r = 0.77 (0.92)
Observed Resistance – methotrexate
Modeled Resistance – methotrexate
r = 0.9 (0.99)
Observed Resistance – itraconazole
Modeled Resistance – itraconazole

Figure S10

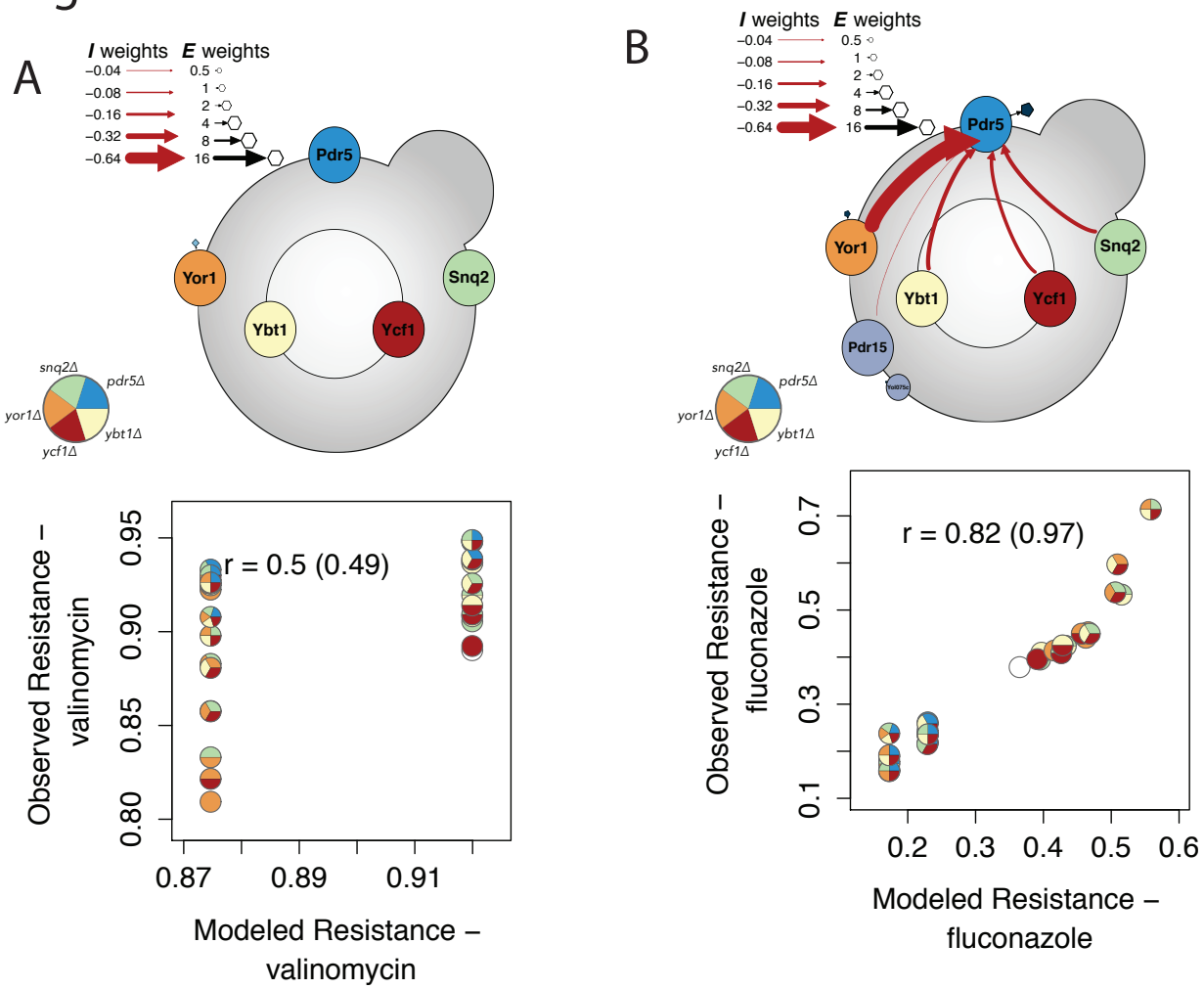

A
I weights
E weights
−0.04
−0.08
−0.16
−0.32
−0.64
0.5
1
2
4
8
16
Pdr5
Yor1
Snq2
Ybt1
Ycf1
snq2Δ
pdr5Δ
yor1Δ
ybt1Δ
ycf1Δ
r = 0.5 (0.49)
Observed Resistance – valinomycin
Modeled Resistance – valinomycin
0.80
0.85
0.90
0.95
0.87
0.89
0.91
B
I weights
E weights
Pdr5
Yor1
Snq2
Ybt1
Ycf1
Pdr15
r = 0.82 (0.97)
Observed Resistance – fluconazole
Modeled Resistance – fluconazole
0.1
0.3
0.5
0.7
0.2
0.3
0.4
0.5
0.6

Figure S11

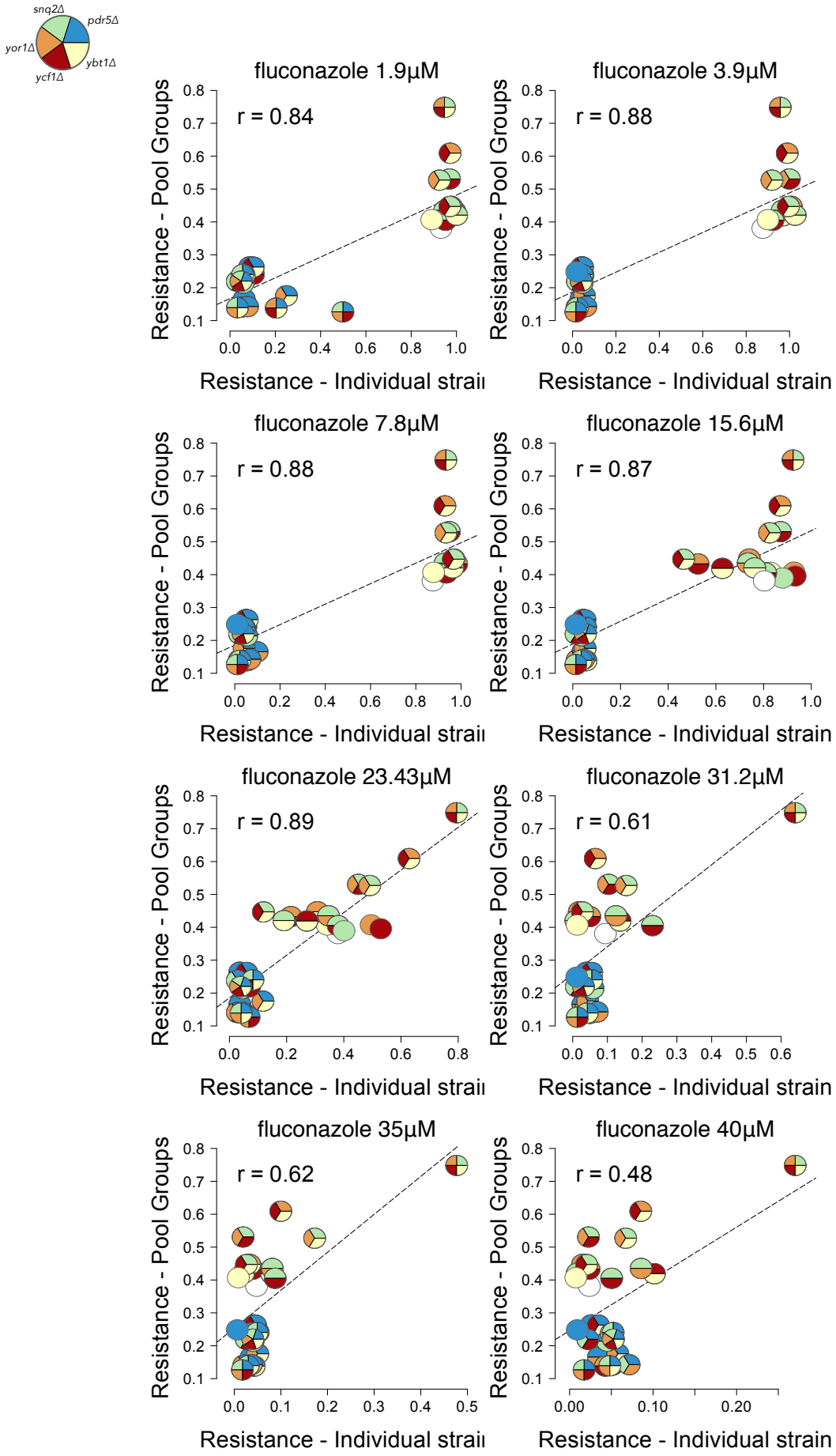
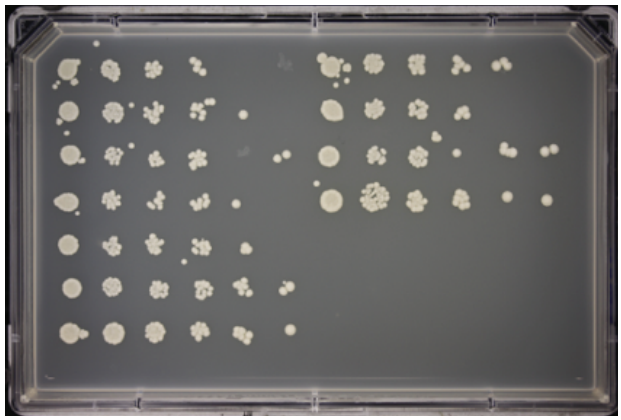
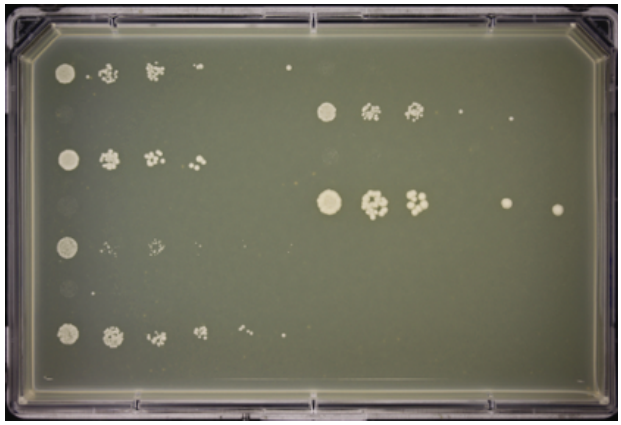

snq2Δ
pdr5Δ
yor1Δ
ybt1Δ
ycf1Δ
fluconazole 1.9µM
r = 0.84
fluconazole 3.9µM
r = 0.88
fluconazole 7.8µM
r = 0.88
fluconazole 15.6µM
r = 0.87
fluconazole 23.43µM
r = 0.89
fluconazole 31.2µM
r = 0.61
fluconazole 35µM
r = 0.62
fluconazole 40µM
r = 0.48
Resistance - Pool Groups
Resistance - Individual strain

Figure S12

DMSO



MTX + DMSO



MTX + DMSO + 46.8μM fluconazole

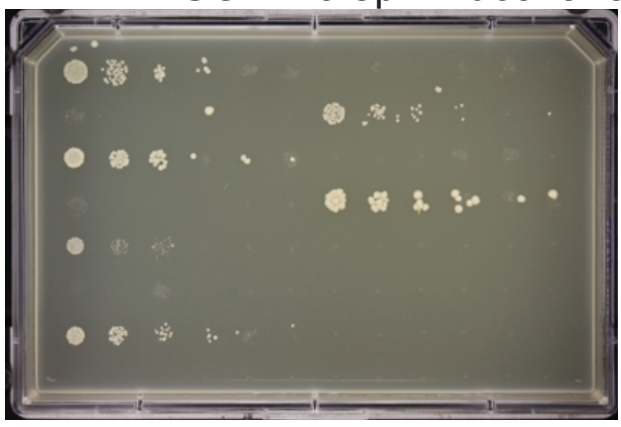
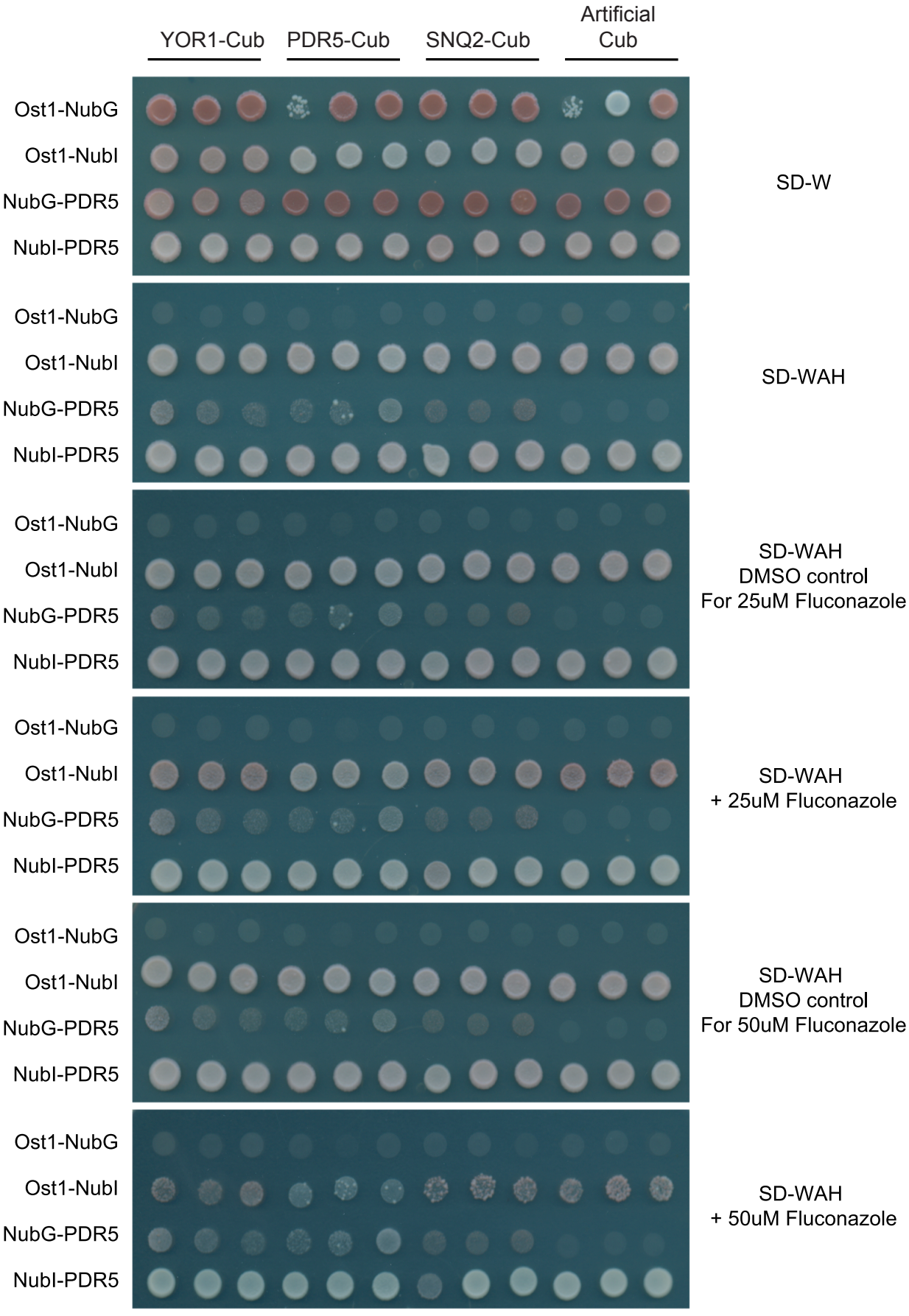


Figure S13


YOR1-Cub
PDR5-Cub
SNQ2-Cub
Artificial Cub
Ost1-NubG
Ost1-NubI
NubG-PDR5
NubI-PDR5
SD-W
SD-WAH
SD-WAH DMSO control For 25uM Fluconazole
SD-WAH + 25uM Fluconazole
SD-WAH DMSO control For 50uM Fluconazole
SD-WAH + 50uM Fluconazole