**Revealing Complex Genetic Bases of ABC Transporter Mediated Drug Resistance in Yeast using an Engineered Population Strategy**

**Abstract**

Combined genetic perturbations can lead to unexpected effects (‘genetic interactions’) that may be critical towards understanding genotype to phenotype relationships. While genetic interactions may result from any number of perturbations, current approaches makes systematic investigation of the vast number of ‘complex interactions’ involving three or more genes impractical. Here we demonstrate an engineered population profiling strategy to map arbitrarily complex genetic relationships amongst 16 ATP Binding Cassette (ABC) transporters. We engineer and genotype a population of ~7,000 *S. cerevisiae* strains containing random ABC transporter knockout combinations, and perform barcode-indexed growth profiling under 16 drugs. We uncover numerous surprising multi-knockout relationships, leading to a map of context-dependent transporter-transporter relationships, and new roles for *YBT1* and *YCF1*. Amongst the findings, a quadruple deletion of *SNQ2*, *YBT1*, *YCF1*, and *YOR1* confered resistance to the *PDR5* substrate fluconazole, and further investigation pointed to a new model of complex mutual inhibition amongst ABC transporters. Thus, we illustrate the potential for an engineered population profiling strategy to inform understanding of complex traits and gene function.

**Introduction**

A gene’s function in the context of a living organism is often understood by observing the resulting phenotypes upon its perturbation. However, as many genotype to phenotype relationships are a result of the combined influence of many genetic variants, genetic perturbations may not result in straightforward and direct effects. An accurate understanding of a gene’s involvement towards a biological function, even in a controlled setting, must often involve uncovering and accounting for its surprising effects when it is perturbed in the context of other genetic variants. Broadly termed ‘genetic interactions’, the presence of these effects within a set of genes is often informative of mechanistic dependence towards a biological function, such as pathway or complex organization.

While single-knockout effects have been extensively profiled in model organisms such as yeast, the vast number of combinations makes efforts to systematically profile ‘genetic interactions’ are much more challenging. Development of extensive molecular tools in yeast has made possible a genome-wide profile of genetic interactions for 23 million two gene knockout combinations. Such two-gene knockout profiles have been used to accurately cluster genes into functional groups1,2, to better predict the genetic basis of drug resistance3, and to inform order-of-action in biological pathways4. However, it is expected that three-gene interactions alone would greatly outnumber all discovered two-gene interactions if mapped genome-wide5. Consistent with this unexplored potential space, approximately 1,000 genes in yeast show no two-gene interactions under standard conditions, and may require diverse environments and complex knockout combinations in order to understand their function2. Individually, examples of complex genetic interactions involving 5+, 7, and 20+ simultaneous mutations have been reported6–8. A survey of three-way interactions shows that they can arise from both simple and complex biological mechanisms9. Formal approaches to capture genetic interactions in a collection of published multi-perturbation experiments predicts that they may have large effects comparable to, or even greater than, those seen in single perturbations10. Thus, despite the difficulty of mapping higher order interactions, they can have broad applicability towards understanding gene function, biological mechanisms, and can make substantial contributions towards genotype-to-phenotype relationships.

Another practical challenge of studying the effects of multiple knockouts is to efficiently introduce multiple peturbations into a single strain. Many molecular tools in yeast are adapted for the creation of many strains with a few deletions rather than a single strain with multiple deletions11. In contrast, several muliplex automated genome engineering (MAGE) methods have been designed to make modifications at multiple loci simultaneously12,13. However, as the aim of MAGE is often to isolate a phenotype of interest, only a few resulting mutants are typically genotyped and characterized, and the potential high-order interactions leading to the desired phenotype are not explored. In *E. coli*, methods are under development to simultaneously genotype many MAGE mutants at multiple loci, with goal of creating extensive genotype to phenotype maps of these engineered populations14,15. In principle, MAGE-like methods can serve as the basis for creating the genetic variation required to understand phenotypes resulting from multiple genetic modifications, and are likely to do so in the near future.

Independent of the method used, the vast number of potential knockout by knockout combinations makes a genome-wide search of higher order interactions impossible.

and instead requires in-depth exploration within a targeted set of genes. Currently, many approaches studying genetic interactions in yeast individually engineer all combinations of desired genetic variants, which is inefficient for creating many knockout combinations involving a set of genes. Other approaches use controlled crosses of natural populations to introduce genome-wide variation *en masse*, which is then statistically associated to traits of interest3. Such strategies are limited to associating naturally occurring variants of unknown causality and effect, making biological interpretation and in-depth functional exploration difficult and limited. Furthermore, vast statistical power would be required to find complex interactions given this genome-wide scope.

In order to explore the potential of higher-order interactions in understanding gene function and phenotype, we had previously deleted 16 ATP Binding Cassette (ABC) transporters in a single yeast strain (ABC-16)11 to serve as a pilot study for future combinatorial engineering approaches. Here we combine a controlled cross with the ABC-16 strain with high-throughput genotyping to tractably map arbitrarily complex genetic interactions under multiple environments. We first back-crossed the ABC-16 strain to a pool of isogenic wild-type parents, each containing a unique DNA barcode16. Then we genotyped 6,709 progeny from this cross *en masse* and linked this genotyping information to a strain-specific barcode, allowing us perform genotype-indexed growth profiling using high-throughput barcode sequencing under 16 drugs16. Overall, we have performed population-level strain engineering within a targeted set of genes in order to determine multi-knockout drug associations in what we term a depthwise association study (DWAS).

We discovered many previously-unknown transporter-transporter relationships in the context of different drugs, even within this highly-studied gene family. Amongst our findings is a complex genetic effect wherein a deletion of *snq2∆*, *ybt1∆*, *ycf1∆*, and *yor1∆* confered unexpected fluconazole resistance which was dependent on *PDR5*. We confirm this complex genetic interaction through individual strain growth profiling, and perform several follow-up experiments guided by the genetic data to inform possible underlying mechanisms. Overall, we demonstrate the potential this strategy to map complex multi-knockout phenotypes and inform understanding of gene function.

**Results**

**Creation of a Genotyped and Phenotyped Engineered Population**

We first mated the ABC-16 strain to a pool of isogenic wild-type strains, each containing two DNA ‘barcode’ tags (UP and DN) and isolated approximately equal numbers of MATa and MATα progeny from this cross into using automated colony picking. We then sought to genotype these progeny *en masse*. First, we PCR amplified the strain-specific barcodes as well as barcodes identifying the deletion of a specific gene (deletion-specific barcodes). These PCR products were tagged so that we could perform high-throughput sequencing and identify which strain-specfic and deletion-specific barcodes came from the same strain our collection (Methods), allowing us to link multiple deletions to a single strain. To calibrate the genotyping results and estimate our success with this method, we also added multiple replicates of 38 individually genotyped strains to our collection (Methods; Fig. 1; Data XX). We estimate a genotyping accuracy of 95% or higher for 12/16 genes, and an overall accuracy of 93.2% (Fig. S1A, Methods). Overall, we obtained useful genotyping data linked to at least one unique DNA barcode for 6,709 progeny with 5,095 unique genotypes, representing approximately 8% of the 216 (65,536) possible knockout combinations within this gene set.

Given our mating-based strategy, we tested the genotyping data for the presence of linkage between ABC transporters, which may affect the knockout combinations obtained. As expected, we found that except for *BPT1* and *YBT1*, all genes were either unlinked or weakly linked (Fig. S1B; r = 0.49 for BPT1 and YBT1). Three pairs of genes – *YOR1-YCF1*, *YOR1-BPT1*, and *SNQ2-PDR5* – had very weak (r ≥ -0.08), but (given the large population size) significant apparent negative linkage, suggesting a possible mild synthetic growth defect under our no-drug growth condition on solid media which may have affected the automated colony picking (Fig. S1B).

As a potential limitation of the high-throughput genotyping method used, only ABC transporter knockouts are tagged with a specific barcode, and thus it is uncertain whether barcode absence implies a wild-type genotype or failure in amplifying a barcode. Conversely, we expect cases where a wild-type is called as a mutant to be comparably negligible. Since over-calling wild type genotypes would lead to a reduction in the average number of knockouts in the pool, we can use the distribution of knockouts in the pool to estimate genotyping accuracy. Indeed, the average number of knockouts in the pool was 7.0, lower than the 8 expected with perfect genotyping. This number is most consistent with an estimated asymmetric genotyping accuracy of 93.8%, which is comparable to the 93.2% derived by comparison to gold standards (Fig S1C, Methods). Thus, we estimate a high pool genotyping accuracy using either method.

After linking each genotype to a strain-specific barcode, we combined our arrayed collection into two separate liquid pools (MATa and MATα) and used high-throughput strain barcode sequencing17 at five time points (0, 5, 10, 15, 20 generations of overall pool growth, Fig. 1) to measure strain prevalence over time. This sequencing was performed for pool growth in the presence of 16 different drugs, as well as a solvent control. Sequencing data from all time points were combined into an overall growth rate metric for each strain, and growth rate in each drug was compared to the solvent control to yield a resistance score (Methods, Fig. 1). After excluding all strains which were not detected at t=0 in the solvent control, we obtained resistance scores for 3,221 strains in the MATa pool and 3,592 strains in the MATα pool. We treat these pools as independent populations. For strains containing two successfully mapped unique DNA barcode (‘UP’ and ‘DN’), we found the resistance scores estimated by either barcode to be highly correlated (r > 0.9 for both MATa and MATα populations) for 8 drugs, and moderately correlated (minimum r > 0.7) for 4 other drugs (Fig S2). This suggests that the resistance metric is not substantially affected by the specific DNA barcode sequence, and that there is sufficient sequencing coverage for high reproducibility. For the remaining 4 drugs, the low correlation appeared to result from a lack of strain-strain variability in resistance rather than large experimental variation, suggesting that the 16 ABC transporters tested did not play a major role in mediating resistance to these compounds (Fig S2).

**Finding Knockouts of Interest in an Engineered Population**

To target relevant genes mediating resistance to each drug for further exploration, we first used the pool to perform marginal associations (i.e. association of a knockout in a population to drug resistance, aggregated over all genetic backgrounds) using a linear model. This ‘marginal TWAS’ recovered XX% of previously-known knockout associations, which had been limited to involving *snq2∆*, *pdr5∆*,and *yor1∆* (Fig. S3). Previously-unreported associations with *ycf1∆*, *bpt1∆*, and *ybt1∆* were also common amongst the drugs tested, and had comparably small marginal effects. However, while the latter 3 genes had minor overall knockout effects, these may vary by knockout background. We found 62% of significant marginal associations to be reproducible between the MATa and MATα pools (Fig. S3), and that models trained on a population of one mating type did comparably well when predicting knockout phenotypes of the opposite mating type (Fig. S4). Thus, even with preliminary analysis, TWAS recovered many previously-known associations, and suggests roles of three new genes towards resistance to the tested drugs.

**A Complex Fitness Landscape: Visualization of Surprising Multi-Gene Knockout Effects**

After identifying gene knockouts of interest within the drugs tested, we used the population data to explore all combinations of multiple knockout effects amongst *snq2∆*, *pdr5∆*, *yor1∆*, *ycf1∆*, *bpt1∆*, and *ybt1∆*. We grouped the strains in the pool based on all 26 knockout combinations amongst these transporters (allowing variation at the other 10 transporters within each group), in order to create a grouped resistance profile under each drug. Clustering these profiles grouped drugs effluxed by the same main transporter into distinct areas (Fig. S5).

We then used this grouped profile to create a ‘complex fitness landscape’, which shows not only the average resistance of each knockout group, but also visualizes the effects of knocking out a transporter in the context of any other knockout combination within the grouped genes. Even without providing an explicit model and definition of genetic interactions, this landscape revealed many surprising and reproducible genetic relationships in different drugs (Fig. 2A, S6). In the most straightforward cases, multiple knockouts appeared to confer surprising drug sensitivity which was not as apparent when knocking out a smaller set of the underlying genes. Such effects are illustrated by the *snq2∆pdr5∆* group under camptothecin (Fig. S6), and *snq2∆pdr5∆ybt1∆* under mitoxantrone (Fig.2A). Broadly, such ‘complex sensitivity’ patterns suggest parallel action of these transporters in the efflux of their respective drugs.

In less intuitive cases, a knockout appeared to have opposite effects drug resistance effects depending on the genetic background. Under benomyl, for example, the effects of *yor1∆* were highly context depenendent - *yor1∆* was shown to have no significant effect by itself, to increase fitness relative to *pdr5∆* as *yor1∆pdr5∆*, and to decrease fitness relative to *snq2∆* as *yor1∆snq2∆* (Fig. 2A). The *yor1∆pdr5∆* interaction may have been expected given a dependence of benomyl resistance to SNQ2 and an earlier report of a (non-additive) ‘compensatory activation’ of Snq2 given disruption of PDR5 and YOR118. However,the *yor1∆snq2∆* interaction is novel and suggests that Yor1 can serve a secondary role in benomyl resistance, for example by providing some efflux in the absence of Snq2.

In perhaps the most surprising examples, complex knockout combinations appeared to confer considerable drug resistance rather than sensitivity. Under fluconazole and ketoconazole, a quadruple deletion of *snq2∆ybt1∆ycf1∆yor1∆* led to drug resistance, whereas *pdr5∆snq2∆ybt1∆ycf1∆yor1∆* was comparably as sensitive as the *pdr5∆* group (Fig. 2A, S6)*.* Furthermore, combinations of one or two knockouts within these four genes appeared to have only minor effects (Fig. 2A, S6).

Given the potential of these grouped profiles to show unexpected multi-knockout effects, we developed two additional representations to explore the data. In the first representation, we created a ‘radial fitness landscape’ which maintains the knockout relationships present in a traditional fitness landscape representation, but in a compact visualization which allows an overview of all data (Fig. 2B,2C, S7). Furthermore, this representation not only shows all contextual knockout effects, but also draws all paths towards a 6-knockout state. Given that each group represents a population of strains, we also observed the underlying distribution of drug resistance for 25 combinations of *snq2∆*, *pdr5∆*, *yor1∆*, *ycf1∆*, and *ybt1∆* under fluconazole(Fig. 2D), further highlighting the surprising multi-knockout resistance combinations amongst these five genes. Thus, even without an explicit definition of genetic interactions, we uncover unexpected multi-knockout effects amongst the ABC transporters in the context of different drugs.

**Formal Modelling of Complex Genetic Interactions: Complex Drug Sensitivity to Knockout Associations**

Given the observation of many apparently-surprising multi-knockout drug resistance phenotypes, we aimed to formally define and model complex genetic interactions within the engineered population. To do this, we developed a linear modelling approach, which extends the multiplicative fitness model of pairwise genetic interactions19 to incorporate interactions of arbitrary complexity (See Methods). We then treated the search for complex genetic interactions as a feature selection problem, and searched for up to to 4-gene interactions within a total space of 2,516 coefficients (See Methods; Fig. 3A). We found 23 2-way, 12 3-way, and 8 4-way interactions which were reproducible between MAT**a** and MAT**α** pools at a stringent Bonferroni corrected p-value cutoff of 0.05 (XX Data S? XX).

This approach formally captured the surprising genetic relationships found in the fitness landscape. For example, *yor1∆* was found to have no main effect under benomyl, to have a positive genetic interaction with *pdr5∆*, and to have a negative genetic interaction with *snq2∆* (XX Fig 3A, Data S? XX). In camptothecin, *pdr5∆* and *snq2∆* each had a minor main effect, as well as a strong negative interaction between them (XX Fig 3A, Data S? XX). Similarly, the surprising *pdr5∆snq2∆ybt1∆* phenotype in mitoxantrone was modelled as a combination of a small marginal effect of *snq2∆*, a negative interaction between *pdr5∆* and *snq2∆*, and a further triple negative interaction upon *YBT1* deletion as *pdr5∆snq2∆ybt1∆* (XX Fig 3A, Data S? XX). A similar ‘redundant transporter action’ phenotypic pattern was observed for *pdr5∆snq2∆yor1∆* in cisplatin (XX Fig 3A, Data S? XX). Under fluconazole, the resistance phenotype was modelled (in addition to one- and two- gene effects involving *pdr5∆*, *yor1∆*, *ybt1∆*, *pdr5∆yor1∆*) as the combination of three positive three-way interactions (*ybt1∆ycf1∆yor1∆*, *snq2∆ybt1∆yor1∆*, *snq2∆ycf1∆yor1∆*). The *PDR5* dependence of these effect were then further modelled as three additional negative four-way interactions(*pdr5∆ybt1∆ycf1∆yor1∆*, *pdr5∆snq2∆ybt1∆yor1∆*, *pdr5∆snq2∆ycf1∆yor1)*.

**Hypothesizing Drug-Dependent Transporter-Transporter Relationshipss from Multi-Knockout Phenotypes**

We sought to intepret the drug-dependent transporter-transporter relationships suggested by the multi-knockout data, using a combination of the grouped fitness landscapes and the formally-determined genetic interactions. The interpreted relationships were highly drug-dependent, and were derived from either two, three, or four knockout phenotypes (XX Fig 3B, XX Data SX XX). All relationships involved either parallel clearance of a drug by two or more transporters, and/or the antagonism of an ABC transporter by the presence of another (Fig 3B). The genetic relationships suggest an extensive role for YBT1 and YCF1 in many drugs, despite none having been previously characterized. YBT1 is modelled to work in parallel with SNQ2, PDR5, or YOR1 depending on the compound, and to antagonize PDR5 in the context of fluconazole resistance (Fig 3B). YCF1 is modelled to antagonize SNQ2 and PDR5 in some compounds, and to work in parallel with SNQ2 in the context of benomyl resistance (Fig 3B).

**Single-Strain Validation of Population-Based Genetic Interactions**

In order to validate the complex genetic interactions present in the population, we sought to generate all knockout combinations of 5 genes identified as important for resistance (*pdr5∆, snq2∆, yor1∆, ybt1∆,* and *ycf1∆*) in individual strains*.* Through a targetted mating and genotyping strategy, all 32 knockout combinations were obtained (see Methods). We first used this collection to confirm the multi-knockout drug resistance observed under fluconazole. Fluconazole resistance of these individual strains correlated well with the pooled resistance (r=XX, Fig XX). As predicted by the TWAS pool, the *snq2∆yor1∆ybt1∆ycf1∆* strain was the fastest growing, and only minor resistance effects were seen for strains with one and two knockout combinations amongst these four genes. To further test the apparent dependence of the increased resistance of the *snq2∆yor1∆ybt1∆ycf1∆* strain on the presence of PDR5, we tested lower concentrations of fluconazole where *pdr5∆* strains still exhibited some growth. Surprisingly, at a low fluconazole concentration (1.3μm), all *pdr5∆snq2∆* strains, including *pdr5∆snq2∆yor1∆-ybt1∆ycf1∆* grew worse than *pdr5∆* strains (Fig XX), further complicating the genotype to phenotype relationship, and suggesting that *snq2∆* can act as a minor efflux pump for fluconazole in addition to its inhibitory role.

**Exploring Mechanisms of Pdr5 inhibition by YOR1 and SNQ2**

The fluconazole growth data matches earlier reports of compensatory activaion of Pdr5p upon knockout of *SNQ2* and *YOR1*, and a similar activation of Snq2p upon knockout of *PDR5* and *YOR1*18. While one study reports that transcriptional activation by Pdr1 is responsible for these effects through an unknown mechanism18, another study failed to find changes in Snq2p protein abundance or localization upon knockout of *PDR5*, and suggested that physical interaction between these two proteins may lead to their mutual repression20.

We noted that homodimers of Pdr5p, Snq2p, and Yor1p have been detected using the mDHFR protein complementation assay (PCA)21,22, and that the Pdr5p homodimer has also been detected using the membrane yeast-two-hybrid (MYTH)20 assay. Furthermore, an interaction has been reported between Pdr5p and Snq2p with both assays20,21. The observation of these homo- and hetero-dimers are consistent with PPI-based repression model, as deletion of any of these genes may shift ABC transporter complex levels towards a greater abundance of homodimers. However, this model would also predict a Pdr5p-Yor1p interaction, which has not been previously reported. To test whether this interaction may have been missed in genome-wide screens, we re-tested a potential Pdr5p-Yor1p interaction using both PCA and MYTH. We confirmed this interaction by MYTH but not PCA (XX Fig. ?? XX), lending support to the PPI-based repression model. Given that the repression appears to be bi-directional for Pdr5p and Snq2p, both in our data (XX Fig ?? XX) and from previous reports18, we hypothesize that the homodimeric form of Pdr5p may be responsible for the efflux of its respective drugs, and that the formation of Pdr5p::Snq2p and Pdr5p::Yor1p heterodimers may deplete the concentration of Pdr5p::Pdr5p homodimers available for efflux.

**Discussion**

We show the potential of profiling the effects of combinatorial gene deletions of arbitrary complexity under multiple conditions. Many observable ABC transporter phenotypes were found to be highly dependent the drug stress and on knockout background. Amongst the inferred ABC transporter relationships, we found extensive novel roles for *YBT1* and *YCF1* in mediating resistance and sensitivity to the tested drugs. The functions of these two genes may have been previously missed due to a lack of known direct substrates, and because many of their functions (e.g. contributing to compensatory activation of *PDR5*, efflux of methotrexate and mitoxantrone) become readily apparent only in certain knockout backgrounds. While we had initially engineered random variation amongst 16 ABC transporters, we were able to identify a smaller set of genes of interest and individually engineer all knockout combinations within this set for in-depth study. A similar approach may be used to study knockout effects in other sets of genes such as pathways or gene families, and may uncover similar previously-uncharacterized functions.

A challenge in studying complex gene knockout phenotypes is that they may be difficult to interpret and can arise from a variety of biological mechanisms9. Here we show that with some knowledge of gene function, complex multi-knockout effects can be used to construct clear follow-up experiments and models. For example, while many potential mechanisms may give rise to multi-knockout phenotypes that grow slower than expected (i.e. ‘sythetic lethality’), we use prior knowledge of ABC transporter function to interpret unexpected multi-knockout sensitivity in the context of a drug as suggesting parallel efflux. Cases where knockouts conferred increased resistance rather sensitivity to a drug were likewise interpreted as contributing to compensatory activation amongst the remaining ABC transpoters, given previous reports. In the case of fluconazole, these multi-knockout phenotypes then allowed us to investigate potential mechanisms of repression, leading to understanding of this phenomenon. Thus, we demonstrate the ability to use complex multi-knockout phenotypes to construct clear mechanistic investivagation and understanding of gene function.

In the future, the use of existing genome engineering tools could make similar studies straightforward. Here, the availability of the ABC-16 strain allowed us to use a mating-based strategy to efficiently introduce mutations to a population of cells at intermediate frequency. While modern tools such as CRISPR may allow engineering of similar multi-knockout strains, we foresee two challenges in the broad expansion of this approach. Firstly, the ABC-16 strain and its progeny do not exhibit major growth defects under standard conditions. Cases where multiple knockouts cause considerable growth defects in a standard condition would require the use of conditional alleles or other tools to suppress constitutive growth defects. Secondly, alternatives to the mating and random segregation based strategy would allow extension of this methodology to systems where mating is not possible, such as human cell lines. In these systems, one may use genome engineering tools to introduce mutations directly into a population at intermediate frequency. Such developments would allow broad expansion of this methodology into many more systems. Here we present a proof-of-concept system to show the value of such expansions.

**Materials and Methods**

**Yeast Strains**

RY0622 (Green Monster MAT**a**):

*MAT***a** *adp1Δ snq2Δ ycf1Δ pdr15Δ yor1Δ vmr1Δ pdr11Δ nft1Δ bpt1Δ ybt1Δ ynr070wΔ yol075cΔ aus1Δ pdr5Δ pdr10Δ pdr12Δ can1Δ::GMToolkit*-**a** (*CMVpr-rtTA KANMX4 STE2pr-Sp-his5*) *his3Δ1 leu2Δ0 ura3Δ0 met15Δ0*.

RY0148 (Barcoder Strain MATα):

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

**Media**

**DNA Oligomers Used**

LoxP:

Lox2272:

US1:

US2:

DS1:

DS2:

U1:  
U2:

D1:

D2:

5’HO:

3’HO:

**PCR Conditions**

For each PCR reaction performed, cells were lysed in a 96-well format in 2 μl of buffer (0.1 M sodium phosphate (pH 7.4) and 1 unit zymolyase (ZymoResearch) overlaid with mineral oil, incubating them at 37 °C for 20 min and then at 95 °C for 5 min.

**High Throughput Sequencing**

**Creating the Barcoder Plasmid**

A plasmid backbone expressing GAL1pr-CRE was combined with a ‘barcoder locus’ flanked by LoxP and Lox2272 sites using in-yeast-assembly. This barcoder locus consists of a random 25bp DNA sequence (‘UPTAG’) in between two common primer regions (US1 and US2), followed by a HphMX4 cassette, and then by another random 25bp DNA sequence (‘DNTAG’) in between two common primer regions (DS1 and DS2).

**Generating a Barcoder Strain**A linear KanMX4 cassette flanked by LoxP and Lox2272 sites was integrated into the HO gene of the RY0148 strain through transformation to serve as the locus for barcode integration. This strain was further transformed with the barcoder plasmid, and transformants were selected using HygroB containing media. These transformants were then subjected to a 4 day galactose induction to allow for Cre expression. All strains which had either lost the plasmid or were unsuccessful in Cre-mediated recombination were counter-selected by growth in the presence of 5-FOA. Strains surviving the counterselection were then grown in YPD in the presence of HygroB in order to select for successful integration of the plasmid barcode locus. Ten colonies were tested for barcode integration using two sets of primers to verify the strain barcode-specific UPTAG and DNTAG - 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). PCR products were analyzed using gel electrophoresis (Supplementary Figure XX).

**Mating, Sporulation, and Haploid Selection**

For mating, we combined MAT**a** and MATα cell cultures 1:1 in <50 μl YPD at 1 OD600 nm, centrifuged the mixture at 735*g* for 5 min and incubated it at 30 °C for 24 h. We transferred 10 μl of the mating mixture to 500 μl GNA medium (starting OD600 nm ≅ 0.1) containing G418 and Nat, and cultured it for 24 h. This allowed for the selection of diploids owing to resistance to G418 and Nat conferred by *GMToolkit*-**a** (*KanMX4*) and *GMToolkit*-α (*NatMX4*), respectively. We transferred 20 μl of the 1-d culture to 500 μl of fresh GNA medium containing G418 and Nat (starting OD600 nm was ~0.2) and cultured for 5–7 h to bring cells to the log phase (OD600 nm was ~1) before sporulation. At this point, we rinsed cells three times with 500 μl of minimum sporulation medium (XX cite) and resuspended them in 1 ml of minimum sporulation medium containing 7.5 μg ml−1 lysine, 7.5 μg ml−1 leucine, 5 μg ml−1 histidine, 5 μg ml−1methionine, and 1.25 μg ml−1 uracil (to meet auxotrophic requirements). The sporulation mixture was rotated at room temperature (20–25 °C) for 1d and then at 30 °C for 3d (XX cite). We treated 125 μl of this mixture with zymolyase (2 units) at 30 °C in a 50-μl reaction (100 mM sodium phosphate buffer (pH 7.4) and 1 M sorbitol) for 1 h, followed by 5 min of treating with NP-40 (added to the zymolyase reaction to achieve a final concentration of 0.01%), and stopped the zymolyase reaction by adding 500 μl of water and placing it on ice. This mixture was sonicated at the output setting of 1 for one minute with the 50% duty cycle using Sonifier 450 (Branson), split equally into two tubes and centrifuged at 735*g* for 5 min.   To the pellet, we added 100 μl of SC without histidine (SC –His) or SC –Leu to separately select for haploid cells of MAT**a** or MATα, respectively, for 16 to 24 h, using the haploid selection markers (*STE2pr-Sp-his5* and *STE3pr-LEU2*).

**Generating Barcoded Random Knockout Progeny**

Mating, sporulation, and haploid selection were performed between the RY0622 ‘Green Monster’ strain (MAT**a**) and the barcoder strain (MATα). 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 (MATα haploid selection). The haploid pools were plated as a form of quality control. From the Sc-His pool, 440 colonies grew in Sc-His, 0 grew in Sc-Leu, 468 grew in YPD+G418, 13 (~2%) grew in YPD+Nat; from the Sc-Leu pool, 0 colonies grew in Sc-His, 1600 grew in Sc-Leu, 50 (~3%) grew in YPD+G418, 1650 grew in YPD+Nat. 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 robot, 5,078 colonies were picked from the MAT**a** pool and 5,844 colonies were picked from the MATα pool onto 384 well plates.

**Individual Strain Genotyping**

To genotype each strain at the 16 ABC transporter loci, two PCR reactions were performed for each locus - one to determine the presence of GFP cassette integration, and another to determine the presence of the wild type gene.

For the first set of reactions, locus–specific PCR primers from the 5′ flanking sequences of each gene were paired with a common primer complementary to the *GFP* cassette (See Table SX). For the second set of reactions, the 5’ locus-specific primers were paired with primers which were complementary to an internal sequence of their corresponding gene (One of two such pair sets was used per gene, see Table SX).

PCR products were analyzed using gel electrophoresis.

**Creating A ‘Gold Standard’ Genotyped Set**

To create a ‘Gold Standard’ genotyped set, 40 progeny strains (19 MAT**a** and 21 MATα), as well as 2 barcoded RY0148 strains and duplicates of a RY0622 ‘Green Monster’ strain and were randomly picked from the collection and subjected to individual genotyping. The strain-specific UPTAG and DNTAG were PCR amplified using two sets of primers – 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). The PCR products were subjected to Sanger sequencing.

From the set of 40 ‘Gold Standard’ genotyped strains, 289 MAT**a** and 300 MATα were added to the 384 well plates in order to calibrate the pooled strain genotyping protocol.

**Pooled Strain Genotyping**

In order to perform *en-masse* genotyping of the random knockout progeny using high throughput sequencing, we adapted our previously-developed Row-Column-Plate (RCP)-PCR protcol23. In order to uniquely tag PCR products originating from the same well, all forward primers had a 5’ tag encoding the well row (R), and all reverse primers had a 5’ tag encoding the well column (C).

For each well in the collection, four PCR reactions were performed with the following primers: R+U1 and C+U2 to amplify DNA barcodes encoding the UPTAGs for each gene deletion; R+D1 and C+D2 to amplify the deletion-specific DNTAGs; R+US1 and C+US2 to amplify the strain-specific UPTAG; R+DS1 and C+DS2 to amplify the strain-specific DNTAG. PCR reactions were performed using a Hydrocycler as previously described23, along with the addition of a plate-specific Illumina sequencing adapter (P tag).

**Analysis of Pooled Strain Genotyping Data**

The pooled PCR products were sequenced using an Illumina MiSeq, and the reads were demultiplexed into individual samples corresponding to a plate and well of origin using a Perl script.

For each sample, we developed a pipeline to determine the strain-specific TAG sequences and genotype from the reads. The parameters of the genotype caller from this pipeline were trained to best match the genotyping calls of strains known to be in the ‘Gold Standard’ set, and the cross-validated accuracy is reported in the text for each gene.

UPTAG or DNTAG identity and a corresponding genotype was successfully determined for 7,195 samples. For 7,030 samples, the UPTAG or DNTAG was unique, and for 165 samples, both the UPTAG and DNTAG 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 random strain was chosen to represent each unique UPTAG and DNTAG sequence.

**Examining Putative Wild-Type Pool Strains**

For 73 unique MAT**a** strains and 131 MATα strains, pooled genotyping had called the strains as wild-type. We isolated many of these strains and tested for the presence of any gene knockout by growth in SC-Ura. Out of 96 MATα 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. Out of 45 MATastrains, all exhibited growth in SC-Ura. Individual genotyping was performed for these MATa strains, and was successful for 40 of 45 strains. These strains had their stated genotype was corrected (Additional File XX). The 5 unsucessfully genotyped strains, as well as 28 additional strains not genotyped, were discarded from analysis. When calculating linkage and distribution of gene knockouts, the wild-type MATα strains were excluded from analysis as they were likely parental strains rather than progeny arising from mating.

**Population Growth Profiling by High-Throughput Sequencing**

All barcoded random knockout progeny were combined into two separate liquid YPD pools, separated by mating type. The effects of 16 different drugs on strain growth were tested to find a concentration which inhibits overall growth by approximately 20%. The determined doses used in the pooled experiment are listed in Additional File XX. All drugs used were dissolved in XX% DMSO, which was used as a solvent control.

0.0625 OD600 nm from a YPD only pool (corresponding to approximately 1250 cells per strain) were added media containing drug/solvent. For the solvent control, a 0 generation sample was immediately harvested afterwards. After growth to 2 OD600 nm, a sample was taken from each drug and 0.0625 OD600 nm of cells were added to fresh medium. This process was repeated three times. Collected samples corresponded approximately to 5, 10, 15, and 20 generations of growth.

Harvested samples were subject to genomic DNA extraction.

Two separate PCR reactions were performed to amplify the UPTAG and DNTAG using the same process as in the analysis of pooled strain genotyping data.

An additional Illumina primer was added to the amplified products. Approximately 200ng of amplified DNA was extracted from each sample and added to 20μL of 2x Phusion Master Mix with HF buffer, 1μL of 10μM F primer, and 1uL of 10μM R primer. F and R primer pairs were PXX+US1/ PYY+US2 and PXX+DS1/PYY+DS2 for the strain-specific UPTAG and DNTAG, respectively. PXX and PYY correspond to the pairs of plate-specific Illumina sequencing adapters which were used to demultiplex the samples (See Data SX).

PCR products amplified using the following PCR program: 98C for 30sec; 98C for 10sec, 60C for 10sec, 72C for 1min (14 times); 72C for 5min; 4C forever

**Sequence Data Processing**

Paired-end Illumina sequencing data was first demultiplexed 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 demultiplexed sample (corresponding to a mating type + timepoint + drug combination), it then searches for a strain barcode within each pair of reads which matches the sample mating type, allowing up to two ungapped mismatches with the putative strain barcode and no fewer than two additional ungapped mismatches separating 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). Furthermore, if a successful match was found for both the forward and reverse reads (i.e. the UP and DN tags), the matches had to correspond to the same strain or the read was discarded. A version of this script was also run which does not restrict the mating type of the potential barcode match to verify that samples were not mislabelled or contaminated.

Any samples with less than 200,000 reads were discarded from the analysis. Additionally, if a sample was discarded for a sample of one mating type, the identical 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 MATα’ would also be discarded regardless of coverage).

**Defining a Resistance Metric**

If both an UP and DN tag for a given strain were successfully linked to a genotype, the UP and DN counts in a given sample were averaged to yield a strain count in that condition. Otherwise either the available tag was used as the count for each strain in a sample. The counts in each sample (corresponding to a pool in drug sequenced time ) were then converted to a frequency through division by the total count for all strains in that sample:

The frequency of each strain was then converted to a ‘virtual area under the growth curve’ () by first multiplying the frequency at each time point by the expected overall pool growth at that time (, since is defined by the number of generations), then taking the integral (using the trapezoid rule to extrapolate missing frequencies between measured timepoints) over all measured timepoints to (the total number of generations measured).

We then define the drug resistance score of a strain as the ratio of its total growth in a drug compared to its growth in the solvent control, per generation:

**Finding Complex Genetic Interactions using a Linear Model**

The multiplicative model of genetic interactions19 was applied to the metric. In this model, the expected resistance of a double knockout strain in a given drug () is the product of the resistances of the corresponding single knockout strains:

To use the multiplicative definition additively, the log of the resistance metric was taken , so that:

A two-gene interaction term is then defined as the deviation of the observed double mutant fitness from this expectation:

When modelling an expected triple knockout fitness, all relevant two-gene interaction terms are added:

Similarly, a three gene interaction term is the deviation from this expectation:

This definition can be analogously extended for interactions of arbitrary complexity.

Practically, the above definition allows modelling of genetic interactions in the pool using linear regression. The terms are given as interaction terms between the corresponding knockouts. Since the genotype of each strain in the population is encoded using a binary variable, if and only if all the corresponding genes in the term are knocked out.

For each sample, a linear model was trained relating genotype to log resistance , including all terms up to a chosen level of complexity. Additionally, each trained model contained terms stating the plate of origin for each strain.

To perform marginal association, we fitted an initial linear model with no terms, and performed stepwise feature elimination (eliminating the gene with the highest value at each step) until all included terms had a significance level of . Linear model term significance was tested using the Type III Sums of Squares ANOVA implementation given in the car package in R.

To train models containing terms of up to -way complexity, additional steps were performed. First, an initial set of a genes was chosen by testing whether a knockout of the 16 ABC transporters resulted in a significant population-level difference in log-resistance, first in the context of the overall population, then in the context of all additional gene knockouts (where is the desired level of gene-gene interaction complexity to be fit in the overall model). Significant differences were evaluated using a t-test, requiring a significance level of , where is the total number of contexts tested for that gene. A linear model including genes passing this test and all of their way interactions was then fit. Many terms were first heuristically eliminated using cross-validated LASSO regression implmented using the cv.glmnet function in R (using the default parameters of alpha = 1, nfolds = 10, type.measure = ‘deviance’) to choose a λ value which minimizes the mean cross validated error and then selecting all terms with a nonzero coefficient at that λ value. Remaining terms were subjected to stepwise elimination as above, at a significance threshold of where is the number of terms in the initial linear model.

**Targeted Mating and Selection to Obtain 32 Knockouts**

**Liquid Growth Assays**

**Analysis of Liquid Growth Data**

**Myth Assay**