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

**Abstract**

Unexpected effects of combined genetic perturbations (‘genetic interactions’) underlie complex traits and are critical to understanding genotype-phenotype relationships. ‘High-order’ genetic interactions­ - those involving three or more genes - ­are likely to greatly outnumber currently well-studied pairwise interactions. However, the large number of potential combinations makes a systematic study of high-order interactions impractical when using approaches based on engineering of individual strains. Here we demonstrate an ‘engineered population’ strategy to map arbitrarily complex genetic interactions in *S. cerevisiae* under multiple environments. We engineered and genotyped a population of ~7000 uniquely DNA-barcoded variants with null variation segregating at 16 loci­ encompassing ATP Binding Cassette (ABC) transporters.  Growth sensitivity of each individual was determined for 16 drugs. We visualized multi-gene fitness effects and identified numerous high-order genetic interactions conferring both drug resistance and sensitivity. Quadruple deletion of *SNQ2*, *YBT1*, *YCF1*, and *YOR1* was found to confer resistance to the *PDR5* substrates fluconazole and ketoconazole. Our study points to a new model for complex mutual inhibition of ABC transporters, and new roles for *YBT1* and *YCF1*. Thus, we illustrate the potential for an engineered population strategy to systematically 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 a resulting phenotype upon its perturbation. However, while some genetic perturbations result in straightforward and direct effects, many genotype to phenotype relationships are, in practice, a result of the combined influence of a large number of genetic variants. An accurate understanding of a gene’s function, even in a controlled setting, must often involve discovering and accounting for its potential unexpected effects when combined with other genetic perturbations. These effects can be broadly termed ‘genetic interactions’, and their presence implies that multiple simultaneous genetic perturbations can be required to reveal phenotypic relationships which are absent when perturbing genes in isolation. Moreover, the presence of genetic interactions within a set of genes is often informative of some dependence in mechanism towards a biological function, such organization into pathways or complexes.

Because of the vast number of combinations, efforts to study ‘genetic interactions’ systematically, even in an organism with extensive availability of molecular tools such as as yeast, have been limited to two gene knockout combinations. These ‘two-gene’ interactions have been used to accurately cluster genes into functional groups[1,2], better predict the genetic basis of drug resistance[3], and inform order-of-action in biological pathways[4]. Understanding two-gene interactions may also be practically important for complex trait prediction in humans, as they may explain up to 80% of the missing heritability in some genome-wide association studies[5].

Despite the success of two-gene interaction studies, many ‘high-order’ genetic interactions, those between three or more genes, have been discovered. It is expected from preliminary studies that three-gene interactions (unexpected three-gene knockout effects) would greatly outnumber all discovered two-gene interactions if mapped genome-wide[6]. Perhaps reflecting this unexplored space and an extensive functional redundancy, approximately 1,000 genes in yeast show no two-gene interactions under standard conditions, and may require diverse environments and/or complex knockout combinations in order to understand their function[2]. Individual examples of high order interactions are known to arise from a variety of biological mechanisms, both simple and complex[7], suggesting their broad applicability towards understanding genotype-phenotype relationships.

While the combinatorial complexity of high order interactions makes a genome-wide search impossible, they may be explored in-depth within a targeted set of genes. However, many current approaches to study genetic interactions involve individually engineering all combinations of desired genetic variants, which is inefficient for creating many knockout combinations involving multiple genes. Alternatively, other studies of genetic interactions use controlled crosses of natural populations to introduce genome-wide variation en masse, which are then statistically associated to traits of interest[3]. However, such strategies are limited to studying naturally occurring variants of unknown effect and causality, making biological interpretation and in-depth functional exploration limited. Furthermore, vast statistical power is required to find complex interactions given a genome-wide scope.

Here we explore a method which combines strain engineering with a controlled cross in order to tractably map arbitrarily complex genetic interactions under multiple environments. Previously, we had deleted 16 ATP Binding Cassette (ABC) transporters in a single yeast strain (ABC-16)[8]. In this study, we have back-crossed the ABC-16 strain to a pool of isogenic wild-type strains, each containing a unique DNA barcode[9]. 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 through high-throughput barcode sequencing under 16 drugs[9]. Overall, we performed population-level strain engineering within a targeted gene set in what we term a target-wise association study (TWAS).

We discover 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 basis of fluconazole resistance wherein a quadruple deletion of *SNQ2*, *YBT1*, *YCF1*, and *YOR1* confers unexpected resistance which is 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 in order to propose several possible underlying mechanisms.

Abstractly, we have used many of the tools available to conventional population studies, but with considerably improved statistical power, and through the use of engineered knockouts, we have created a clear link from association to causation.

ABC transporters have been extensively studied in *S. cerevisiae* and many other organisms, and the 16 known members of the ABC-C and G transporters in our target gene set are widely implicated in drug efflux and antifungal drug resistance, a function shared by other fungi, including human pathogens [10]. and additionally in maintenance of membrane lipid asymmetry and endocytosis of membrane proteins [11–13]. They exhibit properties such as compensatory activation, indicating that they are not simply a system of independent genes working in parallel, but are under extensive feedback [14,15]. Curiously, removal of one or even all of the ABC transporters can even confer drug resistance in some cases [8]. Across other organisms, ABC transporters are one of the largest known protein families, with over 10,000 members spanning all three domains of life known from current genome sequencing efforts [16,17]. Their homologues in humans have a similar extensive functional diversity, as they are widely implicated in resistance to cancer chemotherapy [18], broadly in drug metabolism, and their physiological role in ion maintenance is perturbed in cystic fibrosis [19].

Using TWAS, we were able to map unexpected multi-knockout effects that inform biological mechanism. Amongst our results, we found that a quadruple deletion of *SNQ2*, *YBT1*, *YCF1*, and *YOR1* confers unexpected *PDR5*-depdendent fluconazole and ketoconazole resistance. Genetically, this mechanism was best understood as a four-way parallel inhibition of these four genes on *PDR5.* This phenotype could not be modelled as an additive combination of single, double, or triple knockout effects, in total requiring a 4-way genetic interaction for accurate prediction. Starting with this complex genetic interaction, we were able to support to a model in which Yor1 and Snq2 directly impede the activity of Pdr5 in parallel through a protein-protein interaction mediated mechanism, while Ybt1 and Ycf1 compete with Pdr5 efflux by sequestering its potential substrates in the vacuole. Thus, we systematically uncovered complex genetic interactions and interpreted them to inform novel biological mechanism even within a well-studied gene family.

**Materials and methods**

**-Making the barcoder strain**

We inserted a URA3 cassette flanked by LoxP and Lox2272 sites into the HO locus of the RY0418 strain containing the GMToolkit-α construct. We then inserted a plasmid (which one?) containing two random DNA sequences (how long?) containing with two flanking regions (‘U1’ and ‘U2’ for the uptag and ‘D1’ and ‘D2’ for the downtag) acting as DNA barcodes. These DNA barcodes flanked a HphMX4 resistance marker. In the backbone of the plasmid, we added Cre under a GAL1 promoter and URA3 in the backbone for counterselection. We transformed this plasmid to our modified RY0418 strain, and by adding galactose, we were able to induce Cre-mediated recombination. We then used 5-FOA to counterselect cells which have not been barcoded or still contain URA3 at the HO locus. We then grew cells under YPD and hygromycin to select for the presence of the barcode.

**-Mating barcoder strain with GM**

We used the same strategy as the green monster paper to mate the alpha barcoder strain with the green monster strain. We plated the cells on SC + Kan + Nat to select for diploids, which then underwent sporulation. Progeny were grown on SC –His to select for a strains, and SC –Leu to select for α strains.

**-Simulation for number of progeny needed**

For our simulation, we performed a toy feature selection strategy wherein one variable is added at a time, and its increase in performance is evaluated by 10 fold cross-validation. In this simulation, 70% of the data is taken as the training set, while the other 30% of the data is taken as the test set. If addition of a single term decreases performance, or if the difference in performance is within the standard deviation amongst ten folds, we stop regression. We start by adding single gene terms, then when the algorithm converges, we create interactions amongst these chosen terms to search for two gene interactions and above. Then, we perform a significance test on the linear model such that the Bonferroni corrected p value is 0.05. Using a simple pool of 170 strains grown under benomyl, we used the fitness value range as well the expected error as the median absolute deviation present in the data. We modelled a simulation where there are three genes involved in drug reistance and two of these genes have a main effect, there are two interaction terms, and there is a triple interaction term amongst the three genes. We created a precision recall curve using this preliminary feature selection method and found that 3000 strains achieved (at minimum) approximately 70% recall and 100% precision for all terms.

**-Row column plate PCR protocol**

We plated the resulting pools into solid media, and using a Qpix robot to pick colonies, we picked 14x384 well plates of the ‘a’ mating type (5,376 colonies) and 16x384 well plates of the alpha mating type (6,144 colonies). The selected progeny were then grown in YPD+HygroB in order to select for barcoded haploids. In order to genotype each well, we performed a ‘Row-Column PCR’ strategy wherein the constant U1 and U2 and D1 and D2 sequences flanking barcodes denoting the deletion of each gene, as well as he Us1-Us2 and Ds1-Ds2 constant regions flanking the barcode identifying each strain were amplified using primer pairs containing a sequence denoting the row position of the plate along with the column position of the plate. In order to denote the which plate the samples came from, they also contained a Plate-specific Illumina adapter. PCRs were performed on a hydrocycler. We performed a MiSeq run on a PCR performed on one plate and compared it to manual genotyping of 59 strains wherein a complementary primer pair for each gene as well as its knockout was amplified in order to have complementary genotyping result. In order to make sure that the right barcode was associated with the given genotype, the two barcodes were also identified by two sets of primers in the HO locus (one pair joining the 3’HO locus with the Ds1 sequence, another pairing the 5’HO locus with the Us2 sequence). We found that 94% of genotypes were correctly identified in this set when comparing to the Row-Column-Pate PCR.

**-Computational analysis for Row Colum Plate PCR**

Need Nozomu’s input

**-Strain pooling and sequencing**

Need Marinella’s input

**-Sequence data processing and defining a growth metric**

Paired Illumina primers were added to denote each sample, and demultiplexing was performed by searching for exact matches with the Illumina barcodes for each sample. Within each sample, we searched for the corresponding strain mating type and barcode type matching the sample identity (e.g. we searched only for ‘a’ strains with the ‘DN’ tag if this was indicated in the sample id), allowing for a maximum of two mismatches, and at least two additional mismatches separating the closest matched strain with the second closest. A quality control step allowing for matching to any sample and tag combination was also performed to ensure samples were not mis-labelled, and any samples which had substantial matches to something other than what was indicated in the description were discarded. Furthermore, when comparing a and alpha replicates of a pool, if a time-point was discarded for one mating type it was also discarded for the other one in order to make the growth metrics more comparable. If a strain was called for both the forward or reverse read, it had to be the same strain called for both or it was discarded. The data was pooled to a separate ‘UP’ and ‘DN’ barcode count for each strain. Strains not containing at least one valid Unique DNA barcode were excluded from the analysis. Two sequencing runs (containing the t=0, 5, 15, and t=0, 10, 20 samples) were analyzed in this way and the results were combined into one table. Counts at each time point were converted to a frequency (counts per strain/ total counts per sample), and the counts for the UP and DN tags were added together. Strains not present at t=0 were discarded from the analysis. The t=0 frequency was used to normalize the frequencies for the t=5,10,15,20 timepoints, in total yielding a normalized frequency for each time point corresponding to: (counts per strain at t=x/ total counts at t=x)/(counts per strain at t=0/ total counts at t=0). As the time points were defined by the total growth of the pool rather than the time elapsed, the normalized frequencies were multiplied by 2t to yield a metric that is proportional to the growth of the strain as a fraction of the pool as a whole. The log2 of the ‘virtual area under this curve’ (log2 vAUC) was then taken as the growth metric for each strain, and for each drug was further subtracted by its corresponding log2 vAUC metric in DMSO. This metric was then divided by total time of growth in order to make metrics of samples taken at different time points more comparable.

**-Linear modeling and feature selection**

In order to model genetic interactions, we must first define the expectation of how the effects of two genes are to be combined. While several methods have been evaluated systematically in yeast [20], we will use the multiplicative model that describes the effect of two genes on colony size:

As we have an analogous metric that is meant to estimate total growth, subtracted by a control (in this case, growth of the same strain under no drug), we sought to use this model. A log transformation that allows us to express this as a sum:

, written simply as:

An interaction between term is therefore a deviation from this expectation:

As we move to higher order interactions, we define the expectation as the sum of all single gene effects and all pairwise error terms:

Similarly, a triple interaction is then a deviation from this expectation:

While the suitability of this model depends on the definition of the phenotype, the ability to map genotype to phenotype using linear regression has many advantages, both in computational efficiency and its congruence with classical quantitative genetics models [21]. In the end, we used linear regression, also allowing a term to normalize plate of origin effects in the pool. In order to perform feature selection (also write a paragraph here describing how you tested this), we performed a two-step iterative strategy combining a marginal search, fast feature elimination, significance testing, and model convergence. In the first step of this algorithm, genes were tested for main effect using a Bonferroni corrected p-value cutoff of 0.05. Of the main effects, we assembled up to 4 gene interactions made of these pairs. We also attempted 3 gene interactions, but the 4 gene interactions gave us more significant reproducibility between samples, so we searched up to 4 even though the simulation was designed to only consider up to 3 way interactions, as we were able to capture such interactions reproducibly without much sacrifice in sensitivity for 3 way interactions (XX Supplementary Figure 6 XX). From the 4 gene interaction terms, we performed Elastic Net regression (alpha=0.5, beta=0.5) with 10 fold cross validation to chose the optimal lambda term. All zero coefficients from this model were eliminated. Next, a linear model was fit using all remaining terms, and stepwise feature elimination was performed using Type III sums of squares Anova until all terms had Bonferroni corrected p < 0.05 (the number of hypotheses was taken as all possible interactions possible given the genes of marginal effect). Given the stochastic nature of the feature selection, the terms in the final model were decomposed into single coefficients and the model was re-built again taking into account all interactions within these new chosen terms. This was repeated until the genes used to build the model converged into an identical set of initial genes between two runs.

**-(Genotype enrichment)**

TBA if part of paper

**-Mani’s Strategy**

**-Generating sub-progeny to measure complex genetic effects**

In order to generate all 32 strains needed for confirming the 5 way genetic interactions, the TWAS21230902 was manually genotyped to confirm its genotype from the row-column-plate PCR, and then mated to RY0566. As per the Green Monster protocol, diploids were selected using Kan and Nat, and A haploids were selected on –his +Kan media (and alpha –leu +Nat if we end up using both genotypes) **-Individual growth assays**

Ask Marinella

**-Measuring drug efflux**

In order to measure drug efflux, cells were grown to an overnight culture , resuspended in sterile PBS without glucose, incubated with 10μm Rhodamine 6G for 15 minutes, then washed again with sterile PBS at OD=0.3. To start efflux, cells were again resuspended in PBS+glucose and the images were analyzed over time with the Opera high throughput microscopy system. Background corrected log intensity (taken as the average intensity from t x to y) per cell was taken over time and a slope of the log of the intensity when it is at least 50 units above background was estimated as a way to measure efflux rate for each sample.

**-PPI assays (MYTH)**

Ask Jamie Snider

**Results**

**Designing an Engineered Population**

We envisioned three potential strategies for engineering widespread genetic variation into a target population (XXFigure 1AXX). In the first strategy, a simple cross is made between two strains – one containing all of the desired engineered variation, and another that is identical to the first strain at all but the desired loci. This is the most straightforward method, but is confounded by the presence of tightly linked loci. The second strategy involves a series of coordinated crosses in order to introduce mutations from multiple “parental” strains in a population, and is useful when a single strain containing all of the desired variation does not exist. This population engineering strategy was used to combine natural variation from many existing mouse strains into a multitude of recombinant inbred lines that can be further expanded to create a population of up to 1 million genetically distinct individuals with a known genotype [22]. In the third strategy, genetic variation is introduced directly into a population of cells using technologies such as MAGE [23] or its analog in yeast[24]. The use of such technology has the advantage of being able to modify linked loci and is compatible for situations where a genetic cross is not feasible (such as human cell lines). However, reliably modifying multiple genes in a population at an intermediate frequency has not been shown with such technologies and would require further technological development.

We have previously created an ABC-16 strain with all of our desired genetic variation, and using a cross to a wild type strain we are able to generate our engineered population. First, we performed a power calculation to determine how many individuals are needed to reasonably study complex genetic variation. To estimate the size of the population needed to model a mixture of single, double, and triple interaction effects, we created a simulation in which three genes contribute to fitness with a main effect, two two-gene interactions, and one triple interaction. Using a toy linear model feature selection and cross-validation strategy, we determined that a population size of ~3000 would allow us to recover the modelled interactions at approximately 70% recall with an estimated 100% precision (which is achieved with well under 1000 strains). As such, we aimed to create two independent pools of 3000-4000 strains each in order to assess our results with two independent populations and determine biological reproducibility and robustness to stochastic variation in population composition.

**Creating a Genotyped and Phenotyped Engineered Population**

We first mated the ABC-16 strain to a pool of isogenic wild-type parents, each containing two DNA tags (‘barcodes’). After isolating approximately equal numbers MAT**a** and Mat**α** progeny from this cross, we performed large-scale genotyping using high-throughput sequencing of tagged PCR products, which was calibrated by comparison to 38 strains genotyped using individual PCR reactions (Methods; Fig. 1). We estimate an accuracy of 95% or higher for 12/16 genes, and an overall accuracy of 93.2% (XX Fig. S, Methods XX). 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.

We then tested the genotyping data for the presence of linkage between ABC transporters, and the distribution of knockouts within the progeny. We found that with the exception of *BPT1* and *YBT1*, all of the genes were either unlinked or negligibly linked (XX Fig. S1 XX). Three pairs of genes – *YOR1-YCF1*, *YOR1-BPT1*, and *SNQ2-PDR5* – further had weak, but (given the large sample size) significant apparent negative linkage, suggesting a mild synthetic growth defect under our no-drug growth condition (XX Fig. S1 XX). The average number of knockouts in the pool was XX, which is most consistent with an estimated asymetric genotyping accuracy of XX% (compared to XX% derived by comparison to gold standards; XX Fig S1 XX, Methods).

After linking each genotype to a strain-specific barcode, we pooled our arrayed collection in order to estimate competitive growth rates using high throughput barcode sequencing[25] at five different timepoints (0, 5, 10, 15, 20 generations of overall pool growth, Fig. 1). The two pools were grown in the presence of 16 different drugs and a solvent control. Sequencing data from all timepoints were combined to yield a growth score, which was used to compare each drug to the solvent control in order to yield a resistance score for every strain (Methods, Fig. 1). After excluding all strains which were not present in the solvent control, we obtained resistance scores for 3,186 strains in the MATa pool and 3,523 strains in the Matα pool. Here we treat these pools as biological replicates. Taking advantage of the fact that some strains have two unique DNA barcodes (‘UP’ and ‘DN’), we found the resistance score to be highly correlated amongst different barcodes for the same strain in XX drugs (XX Fig SX XX).

**Pooled Resistance Scores Allow Reproducible Drug-Gene Associations**

As a preliminary test of using TWAS to associate gene knockouts to drug sensitivity, we first used our pools to recover marginal associations (i.e. association of a knockout in a population to drug resistance without taking into account any genetic interactions) using a linear model. This ‘marginal TWAS’ recovered XX% of previously-known knockout associations, which involved *snq2∆*, *pdr5∆*,and *yor1∆*. Previously-unreported associations with *ycf1∆*, *bpt1∆*, and *ybt1∆* were also common amongst the drugs tested, albeit with relatively small marginal effect sizes (XX Fig. SX XX). We found XX% of significant marginal associations to be reproducible between the MATa and MATα pools (XX Fig. SX XX), and further found that models trained on a pool of one mating type had comparable predictive power when applied to the opposite mating type for the same drug (XX Fig. SX XX). In this limited test scenario, TWAS recovered many previously-known associations, and suggests a role of three new genes towards resistance to the tested drugs.

**Visualization of Multi-Gene Knockout Effects in a “Complex Fitness Landscape”**

Given the marginal TWAS results, we used the population data to explore all multiple-knockout effects amongst *snq2∆*, *pdr5∆*, *yor1∆*, *ycf1∆*, *bpt1∆*, and *ybt1∆*. First, we grouped strains based on all 26 knockout combinations amongst these transporters (allowing variation at the other 10 transporters within each group) in order to create a reduced population profile for each drug (XX Fig. SX XX). This profile was used to cluster drugs based on their multi-knockout sensitivity patterns, and grouped separate pools for 15/16 drugs as closest neighbours, indicating strong profile reproducibility (XX Fig. SX XX).

We then converted this grouped profile into a ‘complex fitness landscape’, which shows not only the average resistance of each knockout group, but also predicts the effects of knocking out any of the transporters in the context of any other knockout combination within the grouped genes. Even without providing an explicit model of genetic interactions, this landscape revealed many surprising reproducible genetic relationships in different drugs (XX Fig. ?? XX). In the most straightforward cases, multiple knockouts appeared to confer surprising drug sensitivity which was not clearly present upon knocking out a smaller set of the underlying genes. Examples of this effect are *snq2∆pdr5∆* under camptothecin, and *snq2∆pdr5∆ybt1∆* under mitoxantrone (XX Fig. ?? XX). Broadly, such patterns suggest parallel action of these transporters in the efflux of their respective drugs.

In more complex cases, a knockout appeared to have opposite drug resistance effects depending on the genetic background. Under benomyl, for example, *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∆* (XX Fig. ?? XX). While the first two results may be 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 YOR1[14], the latter is a surprising and previously-unreported effect which may suggest that Yor1 can serve a secondary role in benomyl resistance, such as providing low-affinity efflux in the absence of Snq2.

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

We further explored two alternative representations of these grouped profiles. In the first representation, we created a ‘radial fitness lanscape’ which provides a compact visualization of the grouped profile data while maintaining the knockout relationships present in the more traditional ‘linear’ landscape (XX Fig ?? XX). This representation allows for both an overview of population resistance data, and quick identification of surprising multi-knockout effects, such as the ones observed under fluconazole (XX Fig ?? XX). Furthermore, once a potential surprising effect is discovered, it may then be useful to observe the underlying distriubtion of drug resistance within each group. We illustrate this for all 25 combinations of *snq2∆*, *pdr5∆*, *yor1∆*, *ycf1∆*, and *ybt1∆* under fluconazole(XX Fig ?? XX). In many cases the underlying populations within each group do not appear to be normally-distributed, suggesting further context-dependent roles for the other 10 transporters in this drug.

However,this format does not make apparent the effects of any knockout on fitness under a variety of genetic backgrounds

We looked at the fitness effects of different knockout combinations within these 6 genes (ignoring variation at the other 10 loci) and found that in many cases, the two pools clustered together with each other rather than a different pool with the same drug (XX Supplementary Figure 5 XX ).

**Visualization and Estimation of Multi-Gene Effects in a “Complex Genetic Landscape”**

In order to gain an understanding of multi-gene effects, we developed a means by which to analyze and visualize the estimated multiple knockout phenotypes across a variety of drugs. Again we focused our analysis to 6 genes – YBT1, YCF1, SNQ2, PDR5, YOR1, BPT1, as these were most consistently involved in drug resistance in our initial linear model (XX Supplementary 4 XX). While a clustered heatmap containing the average fitness values of all possible knockout combinations within this set was initially created and made it possible to visualize the growth rates of a multitude of knockout combinations (XX Supplementary 5 XX), this format did not make apparent the effects of any knockout on fitness under a variety of genetic backgrounds. We instead developed a circular plot wherein movement from the center towards the outer ring represents additional knockouts (XX Figure 2D XX). This representation is a two dimensional representation of a six dimensional “fitness landscape”, as all possible paths towards possible sequential knockout combinations amongst these 6 genes are enumerated and arranged in a two-dimensional representation with a clear trajectory. In either format, the presence of complex knockout combinations that confer both drug resistance and sensitivity was apparent, and was consistent with previous findings of single and double ABC-16 transporter knockouts conferring drug resistance[8]. For example, we recapitulated prior results showing that the knockout of PDR5 and YOR1 in benomyl leads to increased resistance ( XX Figure 2D XX) [14]. Similarly, because each gene has a similar probability of being knocked out as well as being present in any given strain, it is possible to envision an ‘inverse’ landscape wherein the effects of sequential additions of genes to the ABC-16 strains are visualized (XX Supplementary Figure XX), as such a view allows a clearer characterization of single-transporter effects if multiple transporters work redundantly to efflux a drug of interest.

**Prediction of Sensitive and Resistant Knockout Combinations**

To take advantage of the engineered genetic complexity, we predicted the most sensitive and resistant knockout combinations for any given drug. In order to do this without explicitly modelling all complex genetic effects, we looked for non-random allele frequencies amongst the most sensitive and resistant strains in the population. In many drugs, it was predicted that the wild type strain would not be the most resistant, and similarly that there are knockout combinations which would render a strain more sensitive than the green monster. For example, in valinomycin the most sensitive strains are predicted to be ones which have YOR1 knocked out, but have SNQ2, PDR5, YCF1, and YBT1 present (XX Figure XX). To test one our predictions, we took a (something knockout strain) and tested its growth in the (presence of some drug) it was verified that this strain grew better than wild type and better than any single knockout. Similarly, we found that (some strain) grew worse under (some drug) than the green monster. Therefore, our pool can be used either to intelligently construct a hypersensitive strain if the green monster is not sufficient (for example it would not be sufficient in…), or to study ABC transporter mediated drug resistance in XYZ drugs. Furthermore, the existence of strains which are more sensitive than the 16 gene knockout may suggest the presence of a secondary compensatory system which interacts with the ABC transporters tested, and this is consistent with previous reports [8].

**Engineered Population Profiling Reveals Extensive Complex Genetic Interactions** Given the unexpected fitness patterns in multi-gene knockouts, we developed a feature selection strategy to determine if there are any fitness effects which cannot be explained by single or double knockout effects. Given the extensive use of the multiplicative model of genetic interactions when measuring total growth by colony size[26] and its analogy to our sequencing-based total growth rate, a logarithmic transformation of our metric allows us to model deviations from expected fitness using a linear model, and readily extend this to definition to capture deviations in the expected fitness of three or more gene knockouts.

While our population size was designed to capture 3-gene interactions at high power, we attempted to expand our modelling to determine if there are more complex interactions which can be captured without a substantial loss of statistical power. We found that up to 4-gene interactions (a total search space of 2,516 coefficients – similar in size to the number of strains in each pool), which were found in both of our a and alpha pools at a stringent Bonferroni corrected p-value cutoff of 0.05, without a substantial loss in 3- and 2-gene interactions (XXSupplementary Figure 6A XX, XX Figure 3B XX). A similar result was observed when modelling gene add-in effects instead of gene knockout effects (XX Supplementary Figure 6B XX, XX Figure 3C XX). We compared the distribution of the effect sizes of complex interactions, and found them to be similar to single and double gene effects (XX Supplementary Figure 6 XX). This indicates that consequences from unexpected multi-gene knockout effects may be similar in magnitude on average as the effects of knocking out a given gene, but do not have as many ‘outlier’ effects.

We revised our previous linear-model based reproducibility estimates using our complex linear models. Overall, we found that XX..the linear models had similar performance (maybe) on its trained pool compared to a test set (have to also normalize by how much those gene levels have in common in actuality), showing both strong reproducibility as well as lack of overfitting. When attempting to cluster conditions based simply on linear model coefficients, we did not achieve as impressive performance, as is expected due to the extensive coefficient search space. Still, in many cases the a and alpha cases clustered strongly together. (Better not to put this)

We also undertook a secondary strategy where potential interactions were determined without the use of an explicit multiplicative model. This strategy takes advantage of the fact that we have a large pool that allows us to construct an extensive number of virtual “quadruplets” wherein we can find 1) a parental strain with an arbitrary background, 2) two “children” strains that vary from the parental strain by the knockout of one gene, and 3) one “grandchild” strain which contains both knockouts of the children strains and is otherwise identical to the parent. By observing the difference in fitness of the grandchild strain to the parent strain and comparing it to the combined difference of the children to the parent strain, any deviations from the expected patterns when performing this analysis for all gene-gene pairs were classified as genetic interactions. Ultimately, we obtained similar interactions with this strategy as we did using our linear model feature search for two gene interactions, showing the robustness of our pairwise genetic interactions to differences in methodology. (Add sentence saying why we chose to go forward with our strategy – e.g. because the linear model version is more readily extensible to higher order interactions).

**Gene Knockouts Mediate Population-Level Variation**

Given our population approach, we are able to investigate not only the effects of multiple gene knockouts on drug resistance and strain fitness, but also how the presence of each gene buffers against the effects on background variation, even if the causes of the underlying variation cannot be explicitly modelled. For example, knocking out a set of genes may not have a considerable effect on mean fitness, but may make the strain more affected by background variation. To test this, we tested which knockout combinations significantly increased or decreased population variance. It was found that…

**Analyzing a Complex Genetic Interaction Between SNQ2, YBT1, YCF1, YOR1, and PDR5 Using a Gene Add-In Model**

One strong example of a multiple gene deletions conferring strong drug resistance is the ∆ybt1∆ycf1∆yor1∆snq2 quadruple deletion in the presence of either fluconazole or ketoconazole. It is predicted both by our linear modelling and by looking at the enrichment of genotypes in the most resistant strains that this knockout combination should be amongst the most resistant to fluconazole and ketoconazole (XX Figure 3C, Figure 2E XX). Furthermore, based on a multiplicative model of gene effects, this resistance reproducibly requires a 4-gene interaction term to explain fully in both the a and alpha pools (XX Figure 3 B-C XX).

In order to develop a hypothesis for the mechanistic basis of this complex resistance genotype, we analyzed the coefficients in our ‘add back regression’ model. Under this model, we hypothesize that a single gene effect represents the ability of a given transporter to efflux the drug or otherwise mediate resistance (or sensitivity) independently of any other. Interactions between transporters under this model could represent parallel action, inhibition of one transporter by another, or the dependence of two transporters to efflux a single drug. Under fluconazole, this model predicts that in the green monster background, only the addition of PDR5 would affect growth rate (XX Figure 3D XX). However, given the presence of PDR5, the further addition of SNQ2, YBT1, YCF1, or YOR1 to the strain impede growth, and we hypothesize that this is because these genes act to inhibit or otherwise directly impede the activity of PDR5 (XX Figure 3D XX). Furthermore, there are positive three gene coefficients when any of these two genes are added in combination in the presence of PDR5 – this is consistent with a model where these four genes are inhibiting PDR5 in parallel such that their combined inhibition is not the sum of their single inhibitions. The significant four gene effects further support this to show there is a complex ‘parallel’ system of repression on PDR5.

We analyzed the reverse regression coefficients for other other drugs, applying the same logic was applied to PDR5. It was found that relationships of activation, inhibition, and parallel action, and otherwise uncharacterized complex feedback (such as the ‘parallel inhbition’ described above) were common amongst ABC transporters (XX Figure 3E XX) and are extensively context dependent, as different drugs show different relationships amongst ABC transporters (XX Supplementary figure XX). For example, amongst SNQ2 and PDR5, our model predicts that SNQ2 inhibits PDR5 in fluconazole, that the two genes work together in camptothecin, and that SNQ2 activates PDR5 in tamoxifen (XX Supplementary figure XX).

**Confirming a Complex Genetic Interaction (may be added to above header)**

In order to determine if the interaction of the SNQ2, YBT1, YCF1, YOR1 quadruple deletion under fluconazole is robust when measured in a single strain growth assay, we constructed all 25 (32) knockout combinations of PDR5, SNQ2, YBT1, YCF1, YOR1 by mating one of the strains in the TWAS pool with a wild type strain (RY0566) and performing targeted genotyping to identify all strains of interest. We confirmed that the quadruple deletion from this cross is very resistant to fluconazole when compared to wild type, and that this effect is dependent on PDR5 – in fact, a five gene knockout is much more sensitive to fluconazole than a PDR5 knockout, suggesting that the removal each of these genes confers resistance by modulating PDR5 activity, but that they are able to confer mild resistance by themselves. However, consistent with the fact that even the 16 knockout strain is viable, none of the knockout combinations showed considerable fitness defects when grown in the presence of DMSO (XX Supplementary Fig XX). As these transporters were involved in resistance to many drugs, we also performed detailed growth assays in the presence of other drugs as informed by the complex population data. We confirmed the prediction that (Strain X is resistant to X drug) and using this sub-population we further predict that (XYZ).We performed a literature search on the roles of SNQ2, YBT1, YCF1, and YOR1 in relation to PDR5 function. Both compensatory non-additive activation of SNQ2 by a PDR5 and YOR1 double knockout and activation of PDR5 by an SNQ2 and YOR1 knockout have been documented[14], which is consistent with a model of parallel inhibition. However, to our knowledge the role of YBT1 and YCF1 has not been documented either in terms of azole resistance or PDR5 function, but their localization to the vacuolar membrane is well documented[27]. We hypothesize that these pumps may work in parallel to localize either azoles or PDR5 substrates in general to the vacuole, where they are sequestered unavailable for PDR5-based export and therefore compete for export in an inefficient manner.

**Verifying Complex Feedback Betweeen ABC Transporters**

-We focused on the SNQ2/PDR5 pair, especially comparing fluconazole and tamoxifen

-First, we measured PDR5 levels in both drugs, and found (XX)

-Then we induced SNQ2 and PDR5 using an artificial promoter and found that XX (we could also grow it in different concentrations in order to make this more clear)

-Thus (can be explained by differential transcription/translation or it cant), and (can be explained by dose-response relationship or it can’t)

-We then measured PDR5 dependent efflux using a radiolabeleld (fluconazole/tamoxifen) and found that (XX)

**-**In order to determine if complex genetic interactions can be mediated at the level of drug efflux, we wanted to observe drug efflux mechanics across all progeny. One common marker for efflux activity is the uptake and export of the rhodamine 6G dye, which is a known PDR5 substrate. Using a high throughput fluorescence microscopy based approach, we measured the cellular accumulation and glucose-dependent efflux across all progeny over the course of 90 minutes across many replicates. When analyzing the genetic data in the context of PDR5 based efflux rather than fluconazole based data, we discovered that ….

Given that SNQ2 and PDR5 have a protein-protein interaction and both form homodimers, we hypothesized that heterodimerization of SNQ2 with PDR5 creates a non-functional transporter and this is the mechanism by which SNQ2 inhibits PDR5. In a similar fashion, we should expect a heterodimer to form between PDR5 and YOR1, and since YOR1-based repression of SNQ2 is evident when analyzing the growth rates of ∆snq2, ∆yor1, and ∆yor1∆snq2 under camptothecin, we also tested for a similar potentially inhibitory protein-protein interaction between SNQ2 and YOR1. By small scale retesting using the MYTH assay it was found that…

An alternative mechanism is transcriptional activation by PDR1/3. In the context of compensatory activation of SNQ2 by a PDR5 and YOR1 knockout, it is unclear whether transcriptional activation is the explanatory mechanism. If this were the case, we would determine that we should see compensatory activation only if PDR5 is expressed under its native promoter. In order to determine whether this is the case, we expressed PDR5 on a plasmid under a conditional promoter in all PDR5 knockout strains to determine if any of the effects of the deletions under a wild type PDR5 strain are recapitulated, either at the level of cell fitness or drug efflux. It was found that…

**Discussion**

In this study, we expanded upon a previously created 16 deletion strain in order to study the effects of combinatorial gene knockouts through an engineered population strategy. To our knowledge, this is the first study to systematically explore genetic interactions of arbitrary complexity in diverse environments. Amongst our results, we recapitulated previous reports of compensatory activation between ABC transporters, such as those between YOR1, SNQ2, and PDR5, and expand this to include previously unreported roles of YBT1 and YCF1 in multiple drug resistance. Even within this well-studied gene system, the compensatory effects of the latter two genes may have been previously missed due to a lack of known direct substrates (and therefore fewer studies), or because their effects become readily apparent only upon multiple knockouts, thus underscoring the importance of a similar approach to studying highly parallel pathways or large gene families.

While systematic studies of two gene interactions under standard growth conditions have been instrumental to discovering functional relationships between yeast genes, we required the use of multiple knockouts in the presence of a relevant drug in order to build our model. While complex genetic interactions may still be challenging to interpret in general[7], we show that with some knowledge of gene function, complex genetic interactions can be used to construct clear follow-up experiments and models. We applied the same rules of epistasis as those used when constructing a regulatory network (rather than a substrate-dependent pathway, which follows the inverse pattern), and make straightforward extensions to these rules to allow for interpreting multiple knockout effects at a population level.

Besides uncovering complex regulatory relationships between ABC transporters, we also expected to find cases of drugs which depend on multiple transporters. While the basis of drug resistance in the ABC transporters has been often understood as each drug having a specific transporter, we have found cases such as camptothecin where clearance appears to be the action of two pumps of similar clearance activity, or in the case of (drug x) where it appears that multiple gene knockouts are required for a clear signature of sensitivity. Thus, our pool can be used to study the ABC-16 efflux dynamics of a drug of interest in cases where such an effect is highy pleiotropic.

Our major limitation compared to traditional association studies is our 16 gene scope, as many factors besides ABC transporters are responsible for drug clearance and may interface with our system [28]. We have shown that the green monster is surprisingly not always the most sensitive strain for any given compound, and that often the presence of a given gene confers sensitivity rather than resistance even when all other ABC transporters in our system are knocked out, suggesting secondary drug resistance mechanisms which may be activated when the ABC-16 system is compromised.

While the availability of a 16 knockout strain made our population engineering strategy more straightforward, the development of multiplex engineering using tools such as CRISPR allow the expansion of our methodology to different organisms and gene systems. For example, the C and G clades in this study have human homologues, allowing for a study of drug efflux and substrate specificity in human cells lines, as these genes are implicated in cancer multidrug resistance.

While growth rate is an easy phenotype to measure en-masse, our method is expandable to allow scoring of more complex or specific phenotypes. For example, cellular morphology has been scored within the yeast knockout collection and the same assay can be adopted for use in an engineered population, given that the scale required is similar to performing a genome-wide screen. Even within our study, we may have expanded the drug efflux assay to a population-wide scale. Another potential extension is the use of multiple conditional alleles in order to study gene systems in which multiple knockouts would prove lethal. Many conditional allele collections exist such as the yeast temperature sensitive collection (which has expanded to include X number of genes), as well as other technologies such as an auxin inducible degron system [29]. Furthermore, while we have engineered null variation into every ABC transporter, this need not necessarily be the case - a similar system could be applied to study the interactions of multiple alleles from multiple genes to identify allele-specific complex interactions, or to perform a similar analysis in diploid yeast strains to study complex haploinsufficiency amongst multiple genes.

**Works Cited**

1. Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, et al. The genetic landscape of a cell. Science. 2010;327: 425–31. doi:10.1126/science.1180823

2. Costanzo M, VanderSluis B, Koch EN, Baryshnikova A, Pons C, Tan G, et al. A global genetic interaction network maps a wiring diagram of cellular function. Science (80- ). 2016;353.

3. Bloom JS, Ehrenreich IM, Loo WT, Lite T-LV, Kruglyak L. Finding the sources of missing heritability in a yeast cross. Nature. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2013;494: 234–7. doi:10.1038/nature11867

4. St Onge RP, Mani R, Oh J, Proctor M, Fung E, Davis RW, et al. Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. Nat Genet. 2007;39: 199–206. doi:10.1038/ng1948

5. Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: Genetic interactions create phantom heritability. Proc Natl Acad Sci U S A. 2012;109: 1193–8. doi:10.1073/pnas.1119675109

6. Tong AHY, Lesage G, Bader GD, Ding H, Xu H, Xin X, et al. Global mapping of the yeast genetic interaction network. Science. American Association for the Advancement of Science; 2004;303: 808–13. doi:10.1126/science.1091317

7. Braberg H, Alexander R, Shales M, Xu J, Franks-Skiba KE, Wu Q, et al. Quantitative analysis of triple-mutant genetic interactions. Nat Protoc. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2014;9: 1867–81. doi:10.1038/nprot.2014.127

8. Suzuki Y, St Onge RP, Mani R, King OD, Heilbut A, Labunskyy VM, et al. Knocking out multigene redundancies via cycles of sexual assortment and fluorescence selection. Nat Methods. 2011;8: 159–64. doi:10.1038/nmeth.1550

9. Smith AM, Heisler LE, Mellor J, Kaper F, Thompson MJ, Chee M, et al. Quantitative phenotyping via deep barcode sequencing. Genome Res. 2009;19: 1836–42. doi:10.1101/gr.093955.109

10. Cannon RD, Lamping E, Holmes AR, Niimi K, Baret P V, Keniya M V, et al. Efflux-mediated antifungal drug resistance. Clin Microbiol Rev. 2009;22: 291–321, Table of Contents. doi:10.1128/CMR.00051-08

11. Gulshan K, Moye-Rowley WS. Vacuolar import of phosphatidylcholine requires the ATP-binding cassette transporter Ybt1. Traffic. 2011;12: 1257–68. doi:10.1111/j.1600-0854.2011.01228.x

12. Decottignies A, Grant AM, Nichols JW, de Wet H, McIntosh DB, Goffeau A. ATPase and multidrug transport activities of the overexpressed yeast ABC protein Yor1p. J Biol Chem. 1998;273: 12612–22. Available: http://www.ncbi.nlm.nih.gov/pubmed/9575223

13. Jungwirth H, Kuchler K. Yeast ABC transporters-- a tale of sex, stress, drugs and aging. FEBS Lett. 2006;580: 1131–8. doi:10.1016/j.febslet.2005.12.050

14. Kolaczkowska A, Kolaczkowski M, Goffeau A, Moye-Rowley WS. Compensatory activation of the multidrug transporters Pdr5p, Snq2p, and Yor1p by Pdr1p in Saccharomyces cerevisiae. FEBS Lett. 2008;582: 977–83. doi:10.1016/j.febslet.2008.02.045

15. Snider J, Hanif A, Lee ME, Jin K, Yu AR, Graham C, et al. Mapping the functional yeast ABC transporter interactome. Nat Chem Biol. 2013;9: 565–72. doi:10.1038/nchembio.1293

16. Kovalchuk A, Driessen AJM. Phylogenetic analysis of fungal ABC transporters. BMC Genomics. BioMed Central; 2010;11: 177. doi:10.1186/1471-2164-11-177

17. Dassa E, Bouige P. The ABC of ABCs: a phylogenetic and functional classification of ABC systems in living organisms. Res Microbiol. 2001;152: 211–229. doi:10.1016/S0923-2508(01)01194-9

18. Fletcher JI, Haber M, Henderson MJ, Norris MD. ABC transporters in cancer: more than just drug efflux pumps. Nat Rev Cancer. 2010;10: 147–56. doi:10.1038/nrc2789

19. Dean M, Rzhetsky A, Allikmets R. The human ATP-binding cassette (ABC) transporter superfamily. Genome Res. 2001;11: 1156–66. doi:10.1101/gr.184901

20. Mani R, St Onge RP, Hartman JL, Giaever G, Roth FP. Defining genetic interaction. Proc Natl Acad Sci U S A. 2008;105: 3461–6. doi:10.1073/pnas.0712255105

21. Cockerham CC. An Extension of the Concept of Partitioning Hereditary Variance for Analysis of Covariances among Relatives When Epistasis Is Present. Genetics. 1954;39: 859–82. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1209694&tool=pmcentrez&rendertype=abstract

22. Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD, Beatty J, et al. The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nat Genet. 2004;36: 1133–7. doi:10.1038/ng1104-1133

23. Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, et al. Programming cells by multiplex genome engineering and accelerated evolution. Nature. Macmillan Publishers Limited. All rights reserved; 2009;460: 894–8. doi:10.1038/nature08187

24. DiCarlo JE, Conley AJ, Penttilä M, Jäntti J, Wang HH, Church GM. Yeast oligo-mediated genome engineering (YOGE). ACS Synth Biol. American Chemical Society; 2013;2: 741–9. doi:10.1021/sb400117c

25. Smith AM, Heisler LE, Mellor J, Kaper F, Thompson MJ, Chee M, et al. Quantitative phenotyping via deep barcode sequencing. Genome Res. 2009;19: 1836–1842.

26. Baryshnikova A, Costanzo M, Kim Y, Ding H, Koh J, Toufighi K, et al. Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. Nat Methods. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2010;7: 1017–24. doi:10.1038/nmeth.1534

27. Paumi CM, Chuk M, Snider J, Stagljar I, Michaelis S. ABC transporters in Saccharomyces cerevisiae and their interactors: new technology advances the biology of the ABCC (MRP) subfamily. Microbiol Mol Biol Rev. 2009;73: 577–93. doi:10.1128/MMBR.00020-09

28. Lee AY, St Onge RP, Proctor MJ, Wallace IM, Nile AH, Spagnuolo PA, et al. Mapping the cellular response to small molecules using chemogenomic fitness signatures. Science. American Association for the Advancement of Science; 2014;344: 208–11. doi:10.1126/science.1250217

29. Morawska M, Ulrich HD. An expanded tool kit for the auxin-inducible degron system in budding yeast. Yeast. 2013;30: 341–51. doi:10.1002/yea.2967