**Deciphering complex traits via engineered population profiling**

Albi Celaj1,2,3, Marinella Gebbia1, Louai Musa2, Atina Cote2, Minjeong Ko2,6, Jamie Snider1, Victoria Wong1, Tiffany Fong5, Joe Mellor1, Gireesh Seesankar5, Maria Nguyen5, Shijie Zhou1, Igor Stagljar1, Nozomu Yachie4,7, and Frederick Roth1,2,3,6,7

[Author list and order is not final]

1Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario, Canada.

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

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

4Synthetic Biology Division, Research Center for Advanced Science and Technology, the University of Tokyo, Tokyo, Japan.

5McMaster University, Hamilton, Ontario, Canada.

6Department of Computer Science, University of Toronto, Toronto, Ontario, Canada.

7Lead contact

**Lead Contact Information:**

Frederick P. Roth, Donnelly Centre and Departments of Molecular Genetics and Computer Science, University of Toronto, 160 College St., Toronto, ON M5S 3E1, Canada

Phone: +1-416-946-5130; Email: fritz.roth@utoronto.ca

Nozomu Yachie, Research Center for Advanced Science and Technology Synthetic Biology Division, University of Tokyo, Rm 4-420, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan.

Phone: +81-3-5452-5242 (x55242); Email: yachie@synbiol.rcast.u-tokyo.ac.jp

**Abstract**

Many biological traits are controlled by complex systems encoded by multiple genes. Systematically identifying the phenotypic impact of high-order combinations of genetic variants is a daunting challenge. We devised an ‘engineered population profiling’ strategy to examine complex mutant combinations within a targeted gene set, and applied it to the entire set of 16 yeast ATP Binding Cassette (ABC) transporters involved in multidrug resistance. We engineered a uniquely-barcoded outbred population of ~7,000 yeast strains, each carrying gene deletions for a random subset of the 16 ABC transporter genes. After genotyping and barcode identification, we profiled the entire population for resistance to each of 16 drugs. The resulting complex genotype-phenotype map revealed many context-dependent transporter-transporter relationships. Striking findings included discovery of a quadruple knockout combination that unexpectedly conferred resistance to fluconazole (a *PDR5* substrate) and two other azole compounds, yielding a complex inhibition model involving five transporters. Our results show that engineered population profiling applied to targeted gene sets can systematically decipher complex traits.

**Introduction**

The function of a gene in the context of a living system is often understood by observing the effects when it is perturbed or varied. However, functional dependency and redundancy in biological systems results in many traits which cannot be straightforwardly by observing single-variant effects1–4. In many cases, organization of genes into functionally dependent pathways or complexes results in surprising multi-knockout effects, broadly termed ‘genetic interactions’. Thus, accurate understanding of how a gene contributes to a biological function often requires uncovering and accounting for its phenotypic effects in the context of other genetic variants. Studying these interactions can further reveal the organization of functional dependencies in the cell, creating a map of cell function5.

Complex genetic interactions likely mediate the expression of many biological traits. A large survey of triple knockout phenotypes in yeast predicts that three-gene interactions would vastly outnumber all discovered two-gene interactions if mapped genome-wide6. Consistent with this, a survey of individual examples shows that they can arise from many different underlying mechanisms, and may thus be prevalent within many biological systems7. Extending beyond three knockouts, individual examples of complex genetic interactions involving 5+, 7, and 20+ simultaneous mutations have been reported in this model organism8–10, but their general prevalence is unknown. A formal approach to capture complex genetic interactions in a collection of published multi-variant experiments shows that they can have large effects which are comparable to or greater than those seen with single gene knockouts11.

A challenge to systematically studying multi-variant effects are the vast number of potential combinations. While single-knockout effects have been widely profiled in many organisms, a genome-wide profile of genetic interactions in yeast required the development of tools to individually engineer and profile more than 23 million strains containing two-gene knockout combinations5. Large two-gene knockout profiles like these have been used to accurately cluster genes into functional groups5,12, to better predict the genetic basis of drug resistance13, and to inform order-of-action in biological pathways14. Despite the scale of such efforts, approximately 1,000 yeast genes remain with no single-knockout growth phenotypes and few or no two-gene interactions under standard growth conditions5. While it is expected that many of these genes would reveal a phenotype in a genome-scale triple knockout screen, such a screen would require the engineering and phenotyping of many more strains than is currently feasible. Because of these limitations, general understanding of biological mechanisms leading to complex genetic interactions far behind that of two-gene knockout effects.

To surmount the combinatorial complexity required to study complex multi-knockout phenotypes, methods are needed to efficiently engineer and genotype the required strains. Many molecular tools used to study two-gene knockouts were developed for the engineering of many strains with a few deletions, and are not easily adaptable for the engineering of many multi-deletion strains due to the limited availability of usable molecular markers15. To make modifications at multiple loci simultaneously, several multiplex automated genome engineering (MAGE) methods have been designed16,17. However, such methods often aim to isolate a handful of engineered strains exhibiting a phenotype of interest, and are not designed to isolate and characterize the large number of strains required to create an in-depth genotype to phenotype map. To extend MAGE for these purposes, methods are under development in *E. coli* to allow the genotyping of many engineered strains at multiple loci18,19. 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.

Here we describe a cross-based population ‘engineered population profiling’ strategy to study arbitrarily complex growth phenotypes within a targeted set of genes. We uncover numerous surprising and reproducible multi-knockout relationships, and further explore the biological basis of one these traits, demonstrating the potential of targeted profiling strategies to map phenotypes controlled by multiple genes and inform understanding of gene function.

**Results**

**Engineered Population Profiling: A Hybrid Scheme for Generating Combinatorial Mutants**

Key to our complex multi-mutant phenotype mapping strategy is the engineering of a large population of cells, each containing independently segregating mutations at multiple target loci. A cross between a multi-mutant and wild-type strain results in random multi-mutant progeny, allowing straightforward population engineering, especially in the absence of strong linkage. A similar approach is used with controlled crosses of natural populations to introduce genome-wide variation *en masse*, which is then statistically associated to traits of interest13. However, these genome-wide association strategies are often limited to studying naturally occurring variants of unknown causality and effect, making biological interpretation and in-depth functional exploration difficult and limited. A genome-wide scope further requires prohibitively many strains for finding complex interactions. To overcome these limtiations of population based methods, all desired variation is targeted into a limited multi-mutant strain.

Once a targeted gene set is chosen, a population must be engineered which allows identification, genotyping, and phenotyping of each strain at a large scale. Strain identification is first performed, as an identifier can be linked to both a genotype and a phenotype. Straightforwardly, a sequence of random base pairs can be integrated into the genome of the engineered strains to act as a DNA barcode20. To introduce a unique barcode to each individual in the population, a parent strain can be transformed with a complex barcode pool, so that progeny inheriting one of these barcodes can then be selected. Linking this barcode identifier to a strain genotype can be performed by sequencing it along with the target loci from each isolated strain. This association can be performed at a large scale by adding another molecular tag to the sequencing products to allow pooling and subsequent high-throughput sequencing. Sequencing of strain barcodes over time allows for high-throughput profiling of growth21. Ultimately, this strategy engineers a large population of multi-mutant strains which can be characterized in-depth under multiple conditions.

**Creating a ‘Universal Barcoder Parent’ Pool for Population Engineering**

We aimed to engineer a population by mating a multi-mutant and wild-type ‘parent’ strain, and first required that progeny from this cross are uniquely molecularly barcoded. For this purpose, we transformed a wild-type RY0148 strain with a large pool of genome-integrated unique DNA ‘barcode’ tags. To achieve high barcode complexity, we used a Cre-mediated recombination strategy. We first created a complex pool of Cre-expressing pSH47 plasmids containing two DNA tags (‘UP’ and ‘DN’) in between a resistance marker flanked by Loxp/Lox2272 sites (Fig. S1, Methods). We then transformed a URA3 casssette flanked by Loxp/Lox2272 sites into the HO locus of the RY0148 strain, and further transformed this strain with the plasmid pool. By induction of Cre, this URA3 cassette is replaced *en masse* by the barcoder locus, allowing for the proper selection conditions to result in a pool of ‘universal barcoder parents’ which can be mated with any appropriate multi-mutant yeast strain (Fig. S1, Methods).

**An Engineered Population of Barcoded Combinatorially Mutated Cells**

One set of genes which are likely to exhibit complex multi-knockout effects are the ATP Binding Cassette (ABC) transporters. Part of a large gene family with over 10,000 members across all three domains of life22, ABC transporters mediate functions such as multidrug resistance, disease progression, and basic cellular functions23,24. In yeast, 16 ABC transporters are involved in multidrug resistance with overlapping specificity25, and knockouts have been known to result in compensatory activation of other transporters leading to unexpected resistance26,27. Even in humans, compensatory activation has been observed in ABCC3 upon knockout or disruption of ABCC2 in both rat knockout models28 and in the context of Dubin-Johnson syndrome29. Because of the complex multi-knockout phenotypes expected by redundancy and feedback within this gene family, and their conservation in many organisms, we used a previously-created strain with deletions at 16 ABC transporters (ABC-16)15 to study these genes as a pilot for engineered population profiling.

The ABC-16 strain was mated to the universal barcoder parent pool, and was subject to sporulation and barcoded haploid selection. Automated colony picking was used to isolate ~10,000 MAT**a** and MAT**α** progeny containing random knockouts at 16 loci into a collection of 384-well plates. To genotype these progeny *en masse* and link this information to their DNA barcode, we amplified strain- and deletion-specific barcodes from each individual in our collection and tagged the amplification products with additional sequences indicating the plate, row, and column of origin by adapting our previously-developed row-column-plate PCR (RCP-PCR) strategy30 (Methods; Fig. 1). To verify and calibrate the genotyping results, multiple replicates of 40 individually genotyped ‘gold standard’ strains, as well as two additional control strains were added to the collection at known positions (Methods; Data S2)

Barcode-linked genotyping was performed using high-throughput sequencing and analysis of the tagged amplification products. Using data from calibration strains, we estimated a genotyping accuracy of ≥95% for 12/16 genes, and an overall accuracy of 93.2% (Fig. S1A, Methods). An independent method based on the distribution of knockouts in the genotyping data estimated a similarly high overall accuracy of 93.8% (Fig. S1C, Methods). Based on the pool genotyping data, all genes were either unlinked or weakly linked except for *BPT1* and *YBT1* (Fig. S1B; r = 0.49). This linkage is unsurprising as these genes are separated by 70.1kb on chromosome XII. Surprisingly, three pairs of genes – *YOR1-YCF1*, *YOR1-BPT1*, and *SNQ2-PDR5* – had weak (-0.04 ≥ r ≥ -0.08) but significant apparent negative linkage (Fig. S1B). These small effects may have arisen during the sporulation, haploid selection, or automated colony picking steps. Overall, usable genotyping data linked to at least one unique DNA barcode was obtained for 6,709 progeny with 5,095 unique genotypes.

**Resistance Profiling of An Engineered Population**

After assessing the genotyping results and determining a barcode identity for many strains, we profiled resistance to a collection of drugs. Arrayed collections were transferred to two liquid pools separated by mating type (MAT**a** and MAT**α**), and treated as independent populations in subsequent analyses. These pools were grown in the presence of 16 different anticancer and antifungal drugs (Data S3), as well as a solvent control. Using high-throughput strain barcode sequencing21, strain frequency was estimated at five time points (0, 5, 10, 15, and 20 generations of overall pool growth, Fig. 1) and an overall growth metric was determined for each strain. Resistance was determined by comparing growth in each drug to the solvent control (Methods, Fig. 1). Resistance scores were determined for 3,221 MAT**a** and 3,592 MAT**α** strains (Data S5), excluding those absent at the onset of growth. To evaluate sequencing complexity, we measured correlation between the resistance scores generated by the UP and DN barcode tags. Resistance scores for 8 drugs were highly correlated (r > 0.9 for both MAT**a** and MAT**α** populations), and moderately correlated (minimum r > 0.7) for 4 other drugs (Fig. S2 A,B). The remaining 4 drugs also had low strain to strain variability in resistance, suggesting that the 16 ABC transporters tested did not play a major role in mediating resistance to these compounds (Fig. S2C).

**Grouped Profiles Recreate Consistent Drug Resistance Relationships**

To explore the multi-knockout data, we first identified a set of genes relevant for resistance to each drug. For this purpose, we used a linear model to determine a marginal association of each knockout to drug resistance. This approach tests for significant knockout effects when aggregated over all genetic backgrounds in the population. While prior reports had been largely focused on the effects of *snq2∆*, *pdr5∆*,and *yor1∆*, we found many unreported associations with *ycf1∆*, *ybt1∆*, and *bpt1∆* to be common amongst the drugs tested (Fig. S3, Data S6). Comparably, the latter three knockouts had small marginal effects, indicating either minor roles towards resistance to the tested drugs, or a greater dependence on knockout background. Of the significant marginal associations found, 62% were reproducible between the MAT**a** and MAT**α** pools (Data S6). From these reproducible associations, 16 of 21 previously-known single-knockout resistance phenotypes were verified (Fig. S3; Data S7).

Relevent gene identification allowed us to group strains by their genotype at *snq2∆*, *pdr5∆*, *yor1∆*, *ycf1∆*, *ybt1∆*, and *bpt1∆*, allowing for background variation at the other 10 transporters. Taking the average resistance of each group yielded a reduced profile for each population, which can cluster drugs by profile similarity (Fig. S4). We found grouped profiles to be highly reproducible between different populations grown in the same drug (r2 > 0.8 for 11 of 16 drugs, Fig. 2A, S6), indicating high reliability in the average resistance measurements. In some drugs, the resistance scores were strikingly similar for the majority of these 64 groups (Fig. 2B). We used these grouped resistance profiles to understand overall patterns in each drug by developing a radial representation which allows for a broad data overview that can be further explored to study all possible trajectories towards a six-knockout state (Fig. 2C). These radial overviews highlight the main ABC transporters conferring resistance to each drug, and multi-knockout phenotypes which were more resistant than the wild-type group were apparent. Reflecting the high profile reproducibility, these path structures were strikingly consistent for populations tested on the same compound, while differing greatly between compounds (Fig. 2D, S6).

**Engineered Population Data Reveals High-Order Combinatorial Drug Resistance Effects**

To further explore the multi-knockout relationships in depth, we used the grouped profiles to recreate a ‘complex fitness landscape’ to allow visualization of transporter knockout effects in the context of any other knockout combination (Fig. 3A). We first evaluated the ability of these profiles to reproducibly capture previously-known surprising multi-knockout traits in the context of benomyl resistance. Specifically, we investigated the previously-reported *SNQ2*-mediated benomyl sensitivity, and ‘compensatory activation’ of Snq2 given disruption of *PDR5* and *YOR1*26. In both the MAT**a** and MAT**α** pools, we confidently recapitulated the *snq2∆* sensitivity (Fig. 3A, S7, *p* = 9.2e-24 MAT**a**, 6.3e-44 MAT**α**; Wilcoxon rank sum test), and consistent with compensatory activation, we found the *pdr5∆* and *pdr5∆yor1∆* strains to exhibit increased resistance compared to the wild type (Fig. 3A, S7, *p* = 3.8e-33 MAT**a**, 4.0e-54 MAT**α** for *pdr5∆* and *p* = 2.0e-32 MAT**a**, 1.4e-53 MAT**α** for *pdr5∆yor1∆*). Additionally, this map revealed that the effects of *yor1∆* were highly context dependent – *yor1∆* was not found to have a significant effect relative to the six-gene wildtype group (*p* = 0.69 MAT**a**, *p* = 0.33 MAT**α**), to increase fitness relative to *pdr5∆* as *yor1∆pdr5∆* (*p* = 2.0e-32 MAT**a**, *p* = 1.57e-37 MAT**α**), and to decrease fitness relative to *snq2∆* as *yor1∆snq2∆* (*p* = 2.6e-09 in MAT**a** and *p* = 1.3e-05 in MAT**α**; Fig. 3A, S7). The *yor1∆snq2∆* phenotype 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.

After validating the ability of the ‘complex fitness landscapes’ to capture previously-known single- and multi-knockout phenotypes in benomyl, we explored surprising genetic relationships under all tested drugs. In the most straightforward cases, sensitive multi-knockout groups were found where removal of a smaller number of the underlying genes had much less apparent effects. These cases are illustrated, for example, by the *snq2∆pdr5∆* group under camptothecin (Fig. S7), and by the *snq2∆pdr5∆ybt1∆* and *snq2∆pdr5∆yor1∆* groups under mitoxantrone (Fig. 3A, S7). Broadly, such complex sensitivity patterns suggest that multiple transporters are able to redundantly efflux a given drug. In several other cases, the fitness landscapes involved complex multi-knockout patterns mediating both drug resistance and sensitivity. In many compounds, multiple paths led to ‘peaks’ conferring more resistance than the wild type state (Fig. S7). Under fluconazole and ketoconazole, a quadruple deletion – *snq2∆ybt1∆ycf1∆yor1∆* – led to resistance, whereas *pdr5∆snq2∆ybt1∆ycf1∆yor1∆* was comparably as sensitive as the *pdr5∆* group (Fig. 3A, S7; )*.* Furthermore, combinations of one or two knockouts within these four genes resulted in minor effects. These findings extend a previously-reported compensatory activation of *PDR5*26, but require multiple knockouts and the involvement of *YCF1* and *YBT1*.

Because each group contains a heterogeneous population of cells with additional background knockouts, we further dissected the underlying distribution within each group. In the context of fluconazole resistance (Fig. 3B), this revealed that variance in resistance within the *pdr5∆* groups is much less than groups containing *PDR5* (e.g. ; *p* = 2.8e-07, 2.5e-26, Bartlett Test). The lack of variability in the *pdr5∆* groups may be explained by the unconditionally high fluconazole sensitivity of these strains, whereas the variability in the *PDR5* groups may reflect still-unidentified and more complex ABC transporter knockout effects in these genetic backgrounds.

After observing many striking and reproducible multi-knockout drug resistance phenotypes, we aimed to formally capture the complex genetic interactions present within the engineered population. We used a linear model to extend the multiplicative model of combined genetic effects31 to incorporate potential interactions (denoted by ε) of arbitrary complexity (see Methods). We treated the search for complex genetic interactions as a feature selection problem, and tested for the presence of up to to 4-gene interactions within a total space of 2,516 coefficients (See Methods; Fig. 3C). We found 23 2-way, 12 3-way, and 8 4-way interactions which were consistent between the MAT**a** and MAT**α** pools at a stringent Bonferroni corrected *p*-value cutoff of 0.05 (Data S6). To verify generalizability, we tested the ability of models made from pooled resistance data of one mating to predict knockout phenotypes of the opposite mating type (Fig. S8).

The formal genetic interaction approach captured many of the apparently-surprising phenotypes found in the fitness landscapes. 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∆* (Fig. 3C, Data S6). In camptothecin, *pdr5∆* and *snq2∆* each had a minor main effect, as well as a strong negative interaction between them (Fig. 3C, Data S6). Similarly, the surprising phenotype in mitoxantrone was modelled as a combination of a small marginal effect of *snq2∆*, a negative interaction , and further triple negative interactions upon *YBT1* and *YOR1* deletions as and (Fig. 3C, Data S6). This pattern of genetic interactions suggests not only that these three genes efflux mitoxantrone in parallel, but also a differential efflux capacity between them (Snq2 has the only single-knockout defect, so it may have the largest efflux potential). A similar ‘parallel transport’ phenotype was observed for *pdr5∆snq2∆yor1∆* in cisplatin (Fig. 3C, Data S6). Under fluconazole, the resistance phenotype was modelled (in addition to one- and two- gene effects involving *pdr5∆*, *yor1∆*, and *ybt1∆*) as the combination of three positive triple interactions (, , ). The dependence of these effects on *PDR5* was then further modelled as three additional negative four-way interactions(, , ).

**Detailed Validation of Population-Based Genetic Interactions**

Using the grouped fitness landscapes and formally-determined genetic interactions, we interpreted a set of transporter-transporter relationship models. These relationships were highly drug-dependent, and were derived from either two, three, or four knockout phenotypes (Fig. 4A, Data S8). We modelled all relationships as involving 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. 4A). These models suggest an extensive role for *YBT1* and *YCF1* in many drugs – *YBT1* is hypothesized to work in parallel with *SNQ2*, *PDR5*, or *YOR1* in the context of mitoxantrone, itraconazole, and methotrexate resistance, and to antagonize *PDR5* in the context of fluconazole resistance (Fig. 4A). *YCF1* is modelled to antagonize *SNQ2* in bisantrene resistance, to antagonize *PDR5* in fluconazole and ketoconazole, and to work in parallel with *SNQ2* in the context of benomyl resistance (Fig. 4A). Because these genes had absent or subtle single-knockout effects in many drugs, these relationships are likely to have been missed in traditional knockout screens. While *BPT1* had roles in mediating resitance to some compounds, its phenotypic patterns were not consistent with antagonism of any other ABC transporter, suggesting that it mediates its effects through interactions with another system32.

In order to validate the complex multi-knockout phenotypes present in the population, we used targeted crosses to generate individual strains containing all knockout combinations of *pdr5∆, snq2∆, yor1∆, ybt1∆,* and *ycf1∆* (Methods). With this population, we tested the multi-knockout resistance patterns for XX drugs under non-competitive liquid growth environments. It was found that XX. Fluconazole resistance of these individual strains correlated well with the pooled resistance (r=XX, Fig XX). As predicted by the pool data, 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.

A targeted collection of individual strains readily allows testing of the relationships of multiple knockouts to a drug response at multiple concentrations. We used this collection to better evaluate the dependence of the increased resistance observed in the *snq2∆yor1∆ybt1∆ycf1∆* strain on the presence of *PDR5.* We reasoned that at concentrations where the *pdr5∆* strain exhibits some growth, partial rescue effects which are missed at higher concentrations will be apparent. At such concentrations (1.3μm; XX% inhibition of *pdr5∆*), it was found that all *pdr5∆snq2∆* strains, including *pdr5∆snq2∆yor1∆ybt1∆ycf1∆* grew worse than *pdr5∆* strains (Fig XX), while there was no partial rescue evident in the *pdr5∆yor1∆ybt1∆ycf1∆* strain. This suggests that *snq2∆* can simultaneously act as a minor efflux pump for fluconazole in addition to its inhibitory effect, allowing its knockout to result in different drug resistance phenotype at different concentrations. Ultimately, these data support a model where multiple gene deletions modulate the level or activity of Pdr5.

**Pdr5 is inhibited by XX through XX**

Given the genetic evidence of repression of Pdr5 by the presence of *SNQ2*, *YOR1*, *YBT1*, and *YCF1*, we sought to understand the underlying mechanisms. Inhibition of Pdr5p by *SNQ2* and *YOR1* has been previously reported, and thought to be mediated by transcriptional activation by Pdr126, although the mechanism underlying this compensatory activation is unknown. To investigate whether *PDR5* mRNA abundance was elevated in the *snq2∆yor1∆ybt1∆ycf1∆* compared to the wild type and *snq2∆yor1∆*, we performed qRT-PCR in both the solvent control and fluconazole. In both the solvent control and fluconazole, *PDR5* transcript abundance was elevated in the *snq2∆yor1∆ybt1∆ycf1∆* strain relative to the wild type strain (*p* = 0.039 in DMSO; *p* = 1.6 x 10-3 in fluconazole; Fig XX). However, we did not find evidence of increased *PDR5* transcript levels in the *snq2∆yor1∆* strain (*p* = 0.366 in DMSO; *p* = 0.15 in fluconazole; Fig. XX). This is consistent with the lack of improved fluconazole resistance of this strain (*p* = XX, Fig. XX) and suggests strain-specific differences in the induction phenomenon. Thus, while we observed transcriptional activation of *PDR5*, here we required additional knockouts in *ybt1∆* and *ycf1∆*.

As an alternative mechanism of *PDR5* repression by Snq2 and Yor1, we investigated the possibility of a protein-protein interaction based mechanism. A previous study investigating an analogous compensatory activation of Snq2 by deletion of *PDR5* and *YOR1* found unchanged protein abundance and localization upon knockout of these two genes, and suggested that physical interaction between these two proteins may lead to their mutual repression27. We also noted that homodimers of Pdr5p, Snq2p, and Yor1p have been detected using the mDHFR protein complementation assay (PCA)33,34, and that the Pdr5p homodimer has also been detected using the membrane yeast-two-hybrid (MYTH)27 assay. Furthermore, an interaction has been reported between Pdr5p and Snq2p with both assays27,33. If these proteins require homodimerization to efflux their given drugs, or the heterodimer forms are otherwise unable to perform this function, then it would be predicted that deletion of any of these genes may shift ABC transporter complex levels towards a greater abundance of their active form. However, this model also predicts the presence of a previously-unreported Pdr5p-Yor1p interaction. We re-tested a potential Pdr5p-Yor1p interaction, and found it using MYTH but not PCA (Fig. ?? XX), lending consistency to the PPI-based repression model.

Given both evidence for transcriptional feedback and consistency with the PPI-based repression model, we tested whether the transcriptional induction is sufficient for the observed increase in resistance. To do this, we removed all endogenous control of *PDR5* by expressing it in a plasmid under the control of a constitutive promoter in the *pdr5∆*, *pdr5∆snq2∆yor1∆*, and *pdr5∆snq2∆yor1∆ybt1∆ycf1∆* strains. However, this still allows for mechanisms such as PPI-based repression to take effect. We found XX.

that the repression appears to be bi-directional for Pdr5p and Snq2p, both in our data (XX Fig ?? XX) and from previous reports26, 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**

Because of combinatorial complexity and lack of suitable molecular tools, characterization and understanding of high-order multi-mutant phenotypes has been challenging and limited. Here we develop a method to straightforwardly engineer and profile a large population of high-order multi-mutant strains, and demonstrate its ability to uncover many unexpected multi-knockout effects and yield new roles for the 16 ABC transporters studied. In this gene family, knockout phenotypes were highly dependent on environment and knockout background, motivating similar studies within other gene groups.

While principles guiding mechanistic understanding of complex genetic interactions will improve with their further characterization, here the available prior knowledge allowed functional interpretation and construction of clear follow-up experiments and models. Complex multi-knockout resistance phenotypes were hypothesized to result from a combination of parallel clearance and feedback between transporters. For example, novel context-dependent roles were found for *YBT1* and *YCF1* in mediating both resistance and sensitivity to the tested compounds, despite their subtle or absent single-knockout effects in a wild-type background. In an illustrative case, the complex drug resistance phenomenon observed in fluconazole extends previous reports of compensatory activation between ABC transporters26,27 to include these genes. Several drugs also exhibited multi-knockout sensitivity patterns involving up to four transporters, suggesting overlapping substrate specificity and motivating the use of this method in studying ABC-transporter mediated clearance of other drugs over single-knockout approaches. Overall, this method yielded many new functions even within a highly-characterized gene family.

The creation of a universal ‘barcoder’ parent pool allowed the use of a cross-based method to efficiently introduce mutations into a population of uniquely-identifiable cells, and for straightforward adaptation of this method for use with other multi-gene deletion strains in yeast. In previous yeast studies, profiles of individuals from controlled crosses have been used to study the genetics of complex traits such as gene expression35 and resistance to small molecules36. In more complex model organisms such as mice, initiatives like the collaborative cross have made use of controlled crosses between inbred strains to create numerous defined heterozygotes that are used to study the variability and genetics of mouse traits37. The engineered population design extends natural population based cross methods by targeting genetic variation to a defined set of genes, creating a clear link from association to causality. Efforts to introduce targeted gene knockouts38 or loss-of-function mutations39 into model organisms such as *C. elegans* enables the use of this strategy with engineered multi-knockout parental strains. Such parental strains can be generated by a series of targeted matings, or with modern genome engineering tools such as CRISPR. Furthermore, a cross-based strategy allows flexibility in the distribution of mutations between the two parents, which is useful in cases where their introduction into a single strain strain would cause considerable defects or lethality.

As an extension to cross-based methods, development of new tools would allow introduction of multi-allele diversity directly into a population. Direct population engineering allows for the study of complex multi-knockout phenotypes in model systems where mating is not possible, such as human cell lines. For example, this extension would further enable analogous studies of human ABC transporter function. We note that methods which aim to achieve direct population-level engineering are under development in *E. coli*19. However, the direct introduction of variation to many strains at multiple loci at intermediate frequency is a major area of future development.

A greater ability to genotype many strains would constitute a considerable improvement to future engineered population profiling methods. The RCP-PCR based method used here allowed large-scale genotyping, but first required the isolation and storage of strains into a defined collection. Linking genotype to a DNA barcode in a direct pooled format without isolating and performing a separate PCR reaction on each strain would therefore allow profiling at a substantially larger scale than is currently possible. We note that despite the complex multi-knockout phenotypes discovered, the majority of knockout combinations amongst the 16 ABC transporters (~92%) remain to be characterized, and even a study within this limited set of genes would benefit from the ability to perform more in-depth genotyping. Furthermore, such methods would enable the discovery of combinatorially complex interactions involving multiple alleles at multiple genes, enabling detailed genetic maps which reflect the variation present in real populations, while allowing for precise allelic control4.

In addition to the context-dependent growth measured in this study, many traits can be measured and linked to a specific genotype *en masse* using a barcode-based strategy. For example, to understand drug efflux, cells may be exposed to a fluorescence-conjugated drug, and fluorescence-activated cell sorting at different time points coupled with barcode sequencing may recreate efflux dynamics for many strains. In a similar fashion, barcode sequencing coupled with cell sorting may be used to reconstruct many phenotypes, for example the effects of knockouts on the activation of signalling pathways, phosphorylation state, epigenetic modifications, or protein abundance41. Extending DNA barcode approaches, methods are being developed to genotype large populations of cells after imaging, which would allow high-content characterization of multi-knockout strains42. With the appropriate design, multi-knockout strains obtained by population engineering can be characterized for many phenotypes.

We envision the use of this method as part of a broader approach to understand complex multi-knockout traits, allowing for in-depth genetic characterization of biological systems in multiple organisms. Even within a relatively small and well-characterized gene group, many functions and gene-gene relationships remained unknown. As techniques for the introduction and characterization of large scale genetic variation improve in scale, allelic variety, and complexity of phenotypic characterization, we expect similar methods to be readily applied towards a greater variety of gene families. Given current estimated prevalence of complex genetic interactions, we expect they will be better characterized in the future and readily related to biological function. Ultimately, this will allow a much more in-depth understanding many gene functions and gene-gene relationships in the context of a living system.

**Materials and Methods**

**Yeast Strains**

RY0622/GM512 (Green Monster MAT**a**):

*MAT***a** *adp1Δ snq2Δ ycf1Δ pdr15Δ yor1Δ vmr1Δ pdr11Δ nft1Δ bpt1Δ ybt1Δ pdr18Δ yol075cΔ aus1Δ pdr5Δ pdr10Δ pdr12Δ can1Δ::GMToolkit*-**a** (*CMVpr-rtTA KANMX4 STE2pr-Sp-his5*) *his3Δ1 leu2Δ0 ura3Δ0 met15Δ0*

RY0146 (Toolkit-a strain):

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

RY0148 (Barcoder Strain MAT**α**):

*MAT****α*** *lyp1Δ his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 can1Δ::GMToolkit-α (CMVpr-rtTA NatMX4 STE3pr-LEU2) ho*∆*::LoxP UPtag HpHMX4 DNtag Lox2272*

**Media**

**SC (SC-His, SC-Leu, SC-Ura)**

**YPD (+HygroB, +Clonnat, +G418)**

**Creating the Barcoder Plasmid**

We added a barcoder locus flanked by LoxP and Lox2272 into pSH47 plasmid backbone expressing GAL1pr-CRE. This barcoder locus consisted of a random 25bp DNA sequence (‘UP tag’) in between two common primer regions (US1 and US2), followed by a HphMX4 cassette, and another random 25bp DNA sequence (‘DN tag’) in between two common primer regions (DS1 and DS2).

First, a barcoded HphMX4 construct was created. HphMX4 was amplified from a pIS420 plasmid using the STEP1F and STEP1R primers containing HphMX4 homology and US2/DS1 overhangs. The PCR program used for this step was 98°C for 30sec; 25 cycles of 98°C for 10sec, 59°C for 10sec, 72°C for 60sec; 72°C for 5min; 4°C forever. These PCR products were purified using a Qiagen Qiaspin kit and confirmed using 2% gel electrophoresis. To the resulting purified products, the STEP2F and STEP2R primers were used to add the random barcodes and US1/DS2 regions with the following PCR program: 98°C for 30sec; 25 cycles of 98°C for 10sec, 68°C for 10sec, 72°C for 60sec; 72°C for 5min; 4°C forever. These resulting products were again purified using a Qiagen Qiaspin kit and ~1.5-1.6kb products were confirmed using 2% gel electrophoresis. To add LoxP/Lox2272 sites, PCR was performed with the STEP2 products using the SacI-LoxP-HphMX4-Barcode-F / SacI-Lox2272-HphMX4-Barcode-R primers. The PCR program used for this step was: 98°C for 30sec; 26 cycles of 98°C for 15sec, 64°C for 20sec, 72°C for 65sec; 72°C for 5min; 4°C forever. The resulting PCR products were purified using a Qiagen Qiaspin Kit, and ~1950bp products were confirmed using 2% gel electrophoresis. Two PCR reactions were performed on the resulting products to confirm correct synthesis. The first PCR reaction was performed with the SacI Reamp F/US2 primer pairs, and the second was performed using DS1/SacI Reamp R primer pairs. The PCR program used for both of these reactions was: 98°C for 30sec; 25 cycles of 98°C for 10sec, 59°C for 15sec, 72°C for 30sec; 72°C for 5min; 4°C forever. Expected sizes (~132bp, 137bp) were confirmed using 4% gel electrophoresis. All PCR reactions above were performed using High Fidelity Phusion Master Mix (NEB).

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

**Generating a Barcoder Strain**

A linear URA3 cassette flanked by LoxP and Lox2272 sites and homology to the HO gene was amplified from purified pIS418 with the 5'HO-LoxP-URA and URA-Lox2272-3'HO primers using the following PCR program: 98°C for 30sec; 25 cycles of 98°C for 10sec, 60°C for 10sec, 72°C for 70sec; 72°C for 5min; 4°C forever. This PCR reaction was performed using High Fidelity Phusion Master Mix (NEB) and was purified using Qiagen Qiaspin. This cassette was integrated into the HO locus of the RY0148 strain through transformation to serve as the ‘landing pad’ for barcode integration using an EZ transformation kit. Transformants selected for growth in SC –Ura plates, and were later verified to exhibit no growth in 5-FOA. A transformant was selected to confirm HO locus integration using three PCR reactions with the following primer pairs: 5'HO-URAreamp + midURA-5'; 5'HO-URAreamp + midURA-3'; 5'HO-URAreamp + 3'HO-URAreamp. All PCR reactions were performed using High Fidelity Phusion Master Mix (NEB) with the following program: 98°C for 30sec; 25 cycles of 98°C for 10sec, 50°C for 10sec, 72°C for 70sec; 72°C for 5min; 4°C forever. Expected PCR product size was confirmed using 2% gel electrophoresis.

The HO::LoxP-URA3-Lox2272 integrant strain was then transformed with a mixture of digested pSH47 and purified PCR products to enable in-yeast-assembly43. Transformation was carried out using a previously established protocol44, with a ~1:6 mixture of digested pSH47:HphMX4 barcode cassette (~12μg digested pSH47 and 15μg cassette). Transformants were grown at 30°C in YPG +HygroB plates for 3 days, allowing both selection of successful transformants and Gal1p-Cre induction. These cells were then scraped and grown overnight in 5-FOA plates to select against non-recombinant strains, and strains containing the recombined barcoder plasmid.

Twenty colonies were tested for barcode integration using PCR and Sanger sequencing. Lysates were made by mixing a sample of each colony with 2μl Sterile DNA Free Water, 2μl 0.2M pH 7.4 Sodium Phosphate Buffer, 0.5 μl 5U/μl zymoresearch zymolyase and incubated at 37°C for 25min and 95°C for 10 min, and stopped by adding 125μl of sterile DNA-free Water. To each lyzed colonly, two sets of primer pairs to verify the strain barcode-specific UP and DN tag - US2 and a sequence complementary to 5’ of the HO gene (5’HO); DS1 and a sequence complementary to the 3’ of the HO gene (3’HO), using the following program: 98°C for 30sec; 25 cycles of 98°C for 10sec, 59°C for 15sec, 72°C for 30sec; 72°C for 5min; 4°C forever. PCR reactions were performed using High Fidelity Phusion Master Mix (NEB) and analyzed using 4% gel electrophoresis to verify the presence of 263bp and 251bp bands. EXOSAP purification was performed on the PCR products by adding 10μl EXOSAP mix (0.025μl ExoI (0.5U), 0.1μl Antarctic Phosphatase (0.5U), 3.5μl 10X Antarctic Phosphatase Buffer, 6.375μl dH2O) to 25μl of PCR products and incubating at 37°C for 30min; 80°C for 20min, then diluting with 35μl of DNA-free H2O to stop the reaction. Diluted EXOSAP products were Sanger sequenced with the 5’HO seq and 3’HO seq primers to confirm the correct barcode construct.

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

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

To genotype each strain at the 16 ABC transporter loci, two PCR reactions were performed for each locus - one to determine the presence of a GFP integration cassette, and another to determine the presence of the wild type gene, as previously described15. For the cassette confirmation reactions, locus–specific PCR primers from the 5′ flanking sequences of each gene were paired with a common primer complementary to the *GFP* cassette (Data S2). Gene presence confirmation primers were designed individually for each gene (Data S2). PCR reactions were performed with a Platinum HiFi mix using the following program: 94°C for 2min; 34 cycles of 94°C for 30sec, 55°C for 30sec, 68°C for 60sec; 68°C for 10min; 4°C forever. PCR products were analyzed using gel electrophoresis.

**Generating Barcoded Random Knockout Progeny**

Mating, sporulation, and haploid selection was performed between the RY0622 ‘Green Monster’ strain (MAT**a**) and the RY0148 barcoder strain (MAT**α**) as previously described15, selecting for MAT**a** and MAT**α** progeny separately. Using colony plating, sporulation efficiency was estimated at 24% - 1080 colonies grew in SC, 140 colonies grew in SC –His (MAT**a** haploid selection), and 120 colonies grew in SC –Leu (MAT**α** haploid selection). The two pools were then grown in YPD +HygroB to select for barcoded haploids. The SC –Leu pool was further grown in SC –Ura to select against barcoder strain parents that may have escaped diploid selection. Using a QPix colony picker, 5,461 MAT**a**  and 5,461 MAT**α** colonies were picked onto 384 well plates. In addition, 299 known positions in both the MAT**a** and MAT**α** arrayed collections consisted of known strains – either one of 40 ‘Gold Standard’ genotyped strains, RY0148, or RY0622 – to act as genotyping controls (Data S2).

**Pooled Strain Genotyping**

A previously-developed Row-Column-Plate (RCP)-PCR protocol30 was adapted in order to perform *en-masse* genotyping of the random knockout progeny using high throughput sequencing. This protocol first uniquely tags PCR products originating from the same well on a given plate, by the use of a 5’ tag encoding the well row (R) in forward primers, a 5’ tag encoding the well column (C) in the reverse primers30. Additionally, these primers contain a linker sequence (PS1 or PS2) which primes a second reaction encoding the plate of origin (Data S2).

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

Four ‘Row-Column’ PCR reactions were performed on the lysates with the following primer pairs: PS1+R+U1 and PS2+C+U2 to amplify DNA barcodes encoding the UP tags for each gene deletion; PS1+R+D1 and PS2+C+D2 to amplify the deletion-specific DN tags; PS1+R+US1 and PS2+C+US2 to amplify the strain-specific UP tag; PS1+R+DS1 and PS2+C+DS2 to amplify the strain-specific DN tag (Data S2). PCR reactions were performed with 2μl of lysed colonies using a Hydrocycler with the following program: 95°C for 5min; 23 cycles of 95°C for 60sec, 57°C for 35sec, 72°C for 45sec; 72°C for 2min; 4°C forever. Row-Column PCR products from each plate were pooled and size was verified on a 4% agarose gel. PCR products from each plate were pooled and 260μl was purified using a Qiagen Qiaquik Spin kit. DNA yield was quantified using a Nanoquant. From the resulting products from each plate, Illumina adapters containing plate tags were added using an additional PCR reaction as previously described30. A pair of PXX\_PE1.0 and PYY\_PE2.0 primers (Data S2) were added to 3-6μl pooled products (calibrated to ~150ng) from each plate to encode the plate of origin, and were amplified using the following PCR program: 98°C for 30sec; 15 cycles of 98°C for 10sec, 59°C for 15sec, 72°C for 40sec; 72°C for 2min; 4°C forever. All PCR reactions above were performed using High Fidelity Phusion Master Mix (NEB).

Expected product size from the plate tags was confirmed on 4% agarose gel. PCR products were purified using a Qiagen Qiaquik Spin kit. qPCR was performed on all plate tag PCR products using a light cycler and KAPA Illumina sequencing quantification kit. qPCR results were used to pool approximately equal amounts of all samples, and 100μl of this multiplexed sample was run on multiple lane of a 4% gel. The desired size (260-290bp) was purified from each lane, and purified using a Qiagen gel purify kit and another qPCR was run on the purified sample.

**Analysis of Pooled Strain Genotyping Data**

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

For each sample, a genotype calling pipeline determined the strain-specific tag sequences and genotype from the reads. The parameters of this pipeline were trained based on known reference strains. Cross-validated accuracy for each gene is reported in Fig. S1A.

UP or DN tag identity and a corresponding genotype was successfully determined for 7,195 samples. For 7,030 samples, the UP or DN tag was unique, and for 165 samples, both the UP and DN tag sequences were redundant with another sample where the called genotype was isogenic or highly similar (≤2 differences), indicating the presence of a single strain in multiple wells. When processing the sequencing data, a single strain was randomly chosen to represent each unique UP and DN tag sequence.

**Examining Putative Wild-Type Pool Strains**

For 73 MAT**a** and 131 MAT**α** strains, pooled sequencing analysis had called the genotype as wild-type. Many of these strains were isolated and tested for the presence of one or more gene knockout cassettes by growth in SC –Ura. Out of 96 MAT**α** strains, 74 exhibited no detectable growth in SC –Ura, and likely arose from remaining barcoder parents which had escaped a previous SC –Ura selection step. The genotypes for the these 74 strains were kept as is, while the other 23 strains, as well as 46 untested strains were discarded from the analysis. Out of 45 MAT**a** strains, all exhibited growth in SC-Ura. Individual genotyping was performed for these MAT**a** strains, and was successful for 40 of 45 strains. These strains had their stated genotype was corrected (Data S2). The 5 unsucessfully genotyped strains, as well as 28 additional strains were discarded from analysis. When calculating linkage and distribution of gene knockouts (Fig. S1), the wild-type MAT**α** strains were excluded from analysis as they were likely parental strains rather than progeny arising from mating.

**Estimating Genotyping Accuracy by Knockout Distribution**

To lend independent support to the genotyping accuracy determined by gold standard strains, an alternate method based on the distribution of knockouts in the population was used. Since *en masse* genotyping associates barcode sequences with ABC transporter knockouts, the absence of a given barcode implies either a wild-type genotype at that locus or a failure in amplification, sequencing, or calling. Conversely, cases where a wild-type is called as a mutant are expected to be comparably negligible. Excess wild-type calls lead to a reduction in the average number of knockouts in the pool, and can be used to estimate genotyping accuracy. The average number of knockouts in the pool was 7.0, lower than the 8 expected with perfect genotyping. This number was most consistent with an asymmetric genotyping accuracy of 93.8%, compared to the 93.2% derived by comparison to gold standards (Fig. S1C).

**Indiviual Liquid Growth Profiling**

To measure individual strain growth, the OD600 nm of a 0.0625 OD600 nm starting culture was measured in the appropriate medium every 15 mins using a GENios microplate reader (Tecan).

**Drug Testing for Growth Inhibition**

The effects of 16 different drugs on strain growth were tested to find a concentration which inhibits wild type growth by approximately 20% (Data S3). All drugs used were dissolved in 2% DMSO, which was used as a solvent control. Growth was determined by the Average\_G metric45, which represents the average generation time.

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

Progeny with at least one mapped strain-specific barcode (Data S2) were combined into two separate liquid YPD + glycerol pools separated by mating type, and kept at −80°C. Samples from the original YPD + glycerol pool were thawn and added to the appropriate drug or solvent containing medium at a final concentration of 0.0625 OD600 nm in 10ml. For the solvent control, a 0 generation sample was immediately harvested for sequencing. After growth to approximately 2 OD600 nm, a sample was taken from each drug for sequencing and cells were resuspended in fresh medium to a final concentration of 0.0625 OD600 nm. This process was repeated until 4 generations of samples were collected. Collected samples corresponded approximately to 5, 10, 15, and 20 generations of growth. Harvested samples were subject to genomic DNA extraction using a YeaStar™ Genomic DNA Kit, quantified using a Qubit® 2.0 fluorometer, and diluted to a final concentration of 20ng/μl. Approximately 350ng of isolated DNA was extracted from each sample and added to 20μL of 2x Platinum PCR SuperMix High Fidelity, 1μL of 10μM F primer, and 1μl of 10μM R primer. F and R primer pairs were PXX+US1/ PYY+US2 and PXX+DS1/PYY+DS2 for the strain-specific UP and DN tag, respectively. PXX and PYY correspond to sequences containing plate-specific Illumina sequencing adapters, as well as tags which were used to demultiplex the samples (See Data S2). PCR products amplified using the following PCR program: 98°C for 30sec; 24 cycles of 98°C for 10sec, 60°C for 10sec, 72°C for 1min; 72°C for 5min; 4°C forever.

PCR products were subject to gel electrophoresis, and ~210bp bands were isolated, subject to gel purification, and eluted in 60μl tris buffer. DNA yield was quantified in duplicate using a KAPA qPCR assay kit, at 1,000-fold, 10,000-fold, and 100,000-fold dilutions to find a concentration within standard curve range. Samples were pooled to yield approximately equal amounts of DNA, and subject to sequencing using an Illumina NextSeq 500 Mid Output kit.

**Sequence Data Processing**

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

All samples with less than 200,000 reads were discarded from the analysis. Additionally, if a sample was discarded for one mating type, the corresponding sample for the opposite mating type was also discarded (e.g. if ‘miconazole t=15 MAT**a**’ was discarded due to lack of coverage, ‘miconazole t=15 MAT**α**’ would also be discarded regardless of coverage).

**Defining a Resistance Metric**

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

The frequency of each strain was then converted into 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 over all measured timepoints to (the total number of generations measured). Frequencies between measured timepoints were linearly interpolated.

The drug resistance score of a strain is defined 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 interactions31 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 this model in an additive rather than multiplicative form, the log of the resistance metric was taken , so that:

We defined a two-gene interaction term as the deviation of the observed double mutant log-fitness from this expectation, rather than the traditional linear difference from a multiplicative estimate.

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, and signficant terms denote interactions between the corresponding knockouts. The genotype of each strain is encoded using a binary variable, so that 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 as a way to account for a potential batch effect.

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 value of , where is the total number of genetic background contexts tested for that gene. A linear model including genes passing this test and all of their way interactions was then fit.

To speed up feature elimination, 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’). The λ which minimizes the mean cross validated error value was then chosen, and all terms with a nonzero coefficient at that λ value were selected for further elimination. These terms were subjected to stepwise elimination (again eliminating the gene with the highest value at each step), at a significance threshold of where is the number of terms in the initial linear model.

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

The TWAS21230902 strain (*pdr10∆ pdr18∆ pdr5∆ snq2∆ ybt1∆ ycf1∆ yor1∆;* Data S2) was subject to individual strain genotyping, confirming the genotype generated using the RCP-PCR based method. This strain (MAT**α**) was mated with RY0146 (MAT**a**), and subject to sporulation and MAT**a** haploid selection15. Individuals from this cross were arrayed onto a 384 well plate, and individually genotyped at *PDR10* and *PDR18*. Strains with no deletions at these genes were further genotyped at *PDR5, SNQ2, YBT1, YCF1,* and *YOR1.* PCR reactions for individual genotyping of these progeny used the Qiagen Mix with the following program: 95°C for 5min; 34 cycles of 95°C for 30sec, 57°C for 30sec, 72°C for 30sec; 68°C for 10min; 4°C forever. After analysis of genotyping results, one strain of each genotype combination was chosen to create the 32-strain collection. These chosen 32 strains were again individually genotyped at these 5 loci for validation.

**Analysis of Liquid Growth Data**

**MYTH Testing of Protein-Protein Interactions**

PDR5, YOR1, and SNQ2 were cloned into the L2 AMBV MYTH bait vector to add a Cub-LexA-VP16 MYTH tag as previously described25. A previously-cloned artificial MYTH-tagged bait plasmid was retrieved, and acted as a negative interaction control. NubG-PDR5 (PDR5 prey) and NubI-PDR5 (PDR5 positive interaction control) strains were retrieved from a previously constructed genomic prey library25. Previously-constructed Ost1p-NubG (negative interaction control) and Ost1p-NubI (positive interaction control) strains were also retrieved. All prey-bait combinations were obtained using individual transformations and selected for growth in SD –Trp (SD –W)46. Colonies of transformed strains were grown in solid medium for 5 days in SD –W, SD –Trp–Ade–His (SD –WAH), SD –WAH +25μM fluconazole + 2% DMSO, and SD –WAH + 2% DMSO.

**PCA Testing of Protein-Protein Interactions**

PDR5, YOR1, and SNQ2 MAT**a** (mDHFR-F[1,2]-NatMX fusions) and MAT**α** (mDHFR-F[3]-HphMX fusions) PCA strains were obtained from a previous genome-wide screen33. Strains were individually mated and diploids were selected on solid YPD supplemented with Hygromycin B and Nourseothricin (YPD +Hyg +Nat). Diploid strains were grown in solid YPD +Hyg +Nat with 200 μg/mL methotrexate for 96 hours at 30°C.

**Quantitative RT-PCR**

RNA extraction was performed using the QIAGEN RNeasy® kit. 1μg of isolate was treated with DNAse and analyzed usign an Agilent Bioanalyzer to quantify and verify purity. cDNA synthesis was perfomed using a combination of oligo-DT and random hexamer primers. qPCR on these samples was then performed using a SensiFAST™ Real-Time PCR Kit and Ct values were quantified using a CFX machine.

**Availability of Data and Materials**

**Competing Interests**

The authors declare that they have no competing interests.

**Acknowledgements**

This work was supported by XX.

**Author Contributions**

N.Y. & F.R. conceived the experiments. N.Y., L.M, M.G., S.Z. & T.F. performed experiments. A.C. and N.Y. analyzed the data. A.C, F.P.R., & N.Y. wrote the paper.

**Additional Data Files**

**Additional Data S1.** List of primers used in this study.

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

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

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

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

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

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

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

**References**

1. Benfey, P. N. & Mitchell-Olds, T. From Genotype to Phenotype: Systems Biology Meets Natural Variation. *Science (80-. ).* **320,** (2008).

2. Hartwell, L. Robust Interactions. *Science (80-. ).* **303,** (2004).

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

4. Civelek, M. & Lusis, A. J. Systems genetics approaches to understand complex traits. *Nat. Rev. Genet.* **15,** 34–48 (2014).

5. Costanzo, M. *et al.* A global genetic interaction network maps a wiring diagram of cellular function. *Science (80-. ).* **353,** (2016).

6. Tong, A. H. Y. *et al.* Global mapping of the yeast genetic interaction network. *Science* **303,** 808–13 (2004).

7. Braberg, H. *et al.* Quantitative analysis of triple-mutant genetic interactions. *Nat. Protoc.* **9,** 1867–81 (2014).

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

9. Beh, C. T., Cool, L., Phillips, J. & Rine, J. Overlapping functions of the yeast oxysterol-binding protein homologues. *Genetics* **157,** 1117–40 (2001).

10. Wieczorke, R. *et al.* Concurrent knock-out of at least 20 transporter genes is required to block uptake of hexoses in Saccharomyces cerevisiae. *FEBS Lett.* **464,** 123–8 (1999).

11. Weinreich, D. M., Lan, Y., Wylie, C. S. & Heckendorn, R. B. Should evolutionary geneticists worry about higher-order epistasis? *Curr. Opin. Genet. Dev.* **23,** 700–707 (2013).

12. Costanzo, M. *et al.* The genetic landscape of a cell. *Science* **327,** 425–31 (2010).

13. Bloom, J. S., Ehrenreich, I. M., Loo, W. T., Lite, T.-L. V. & Kruglyak, L. Finding the sources of missing heritability in a yeast cross. *Nature* **494,** 234–7 (2013).

14. St Onge, R. P. *et al.* Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. *Nat. Genet.* **39,** 199–206 (2007).

15. Suzuki, Y. *et al.* Knocking out multigene redundancies via cycles of sexual assortment and fluorescence selection. *Nat. Methods* **8,** 159–64 (2011).

16. Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460,** 894–8 (2009).

17. DiCarlo, J. E. *et al.* Yeast oligo-mediated genome engineering (YOGE). *ACS Synth. Biol.* **2,** 741–9 (2013).

18. Zeitoun, R. I. *et al.* Multiplexed tracking of combinatorial genomic mutations in engineered cell populations. *Nat. Biotechnol.* **33,** 631–637 (2015).

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

20. Yan, Z. *et al.* Yeast Barcoders: a chemogenomic application of a universal donor-strain collection carrying bar-code identifiers. *Nat. Methods* **5,** 719–725 (2008).

21. Smith, A. M. *et al.* Quantitative phenotyping via deep barcode sequencing. *Genome Res.* **19,** 1836–42 (2009).

22. Kovalchuk, A. & Driessen, A. J. M. Phylogenetic analysis of fungal ABC transporters. *BMC Genomics* **11,** 177 (2010).

23. Jungwirth, H. & Kuchler, K. Yeast ABC transporters – a tale of sex, stress, drugs and aging. *FEBS Lett.* **580,** 1131–8 (2006).

24. Dean, M., Rzhetsky, A. & Allikmets, R. The human ATP-binding cassette (ABC) transporter superfamily. *Genome Res.* **11,** 1156–66 (2001).

25. Nakamura, K. *et al.* Functional Expression of Candida albicans Drug Efflux Pump Cdr1p in a Saccharomyces cerevisiae Strain Deficient in Membrane Transporters. *Antimicrob. Agents Chemother.* **45,** 3366–3374 (2001).

26. Kolaczkowska, A., Kolaczkowski, M., Goffeau, A. & Moye-Rowley, W. S. Compensatory activation of the multidrug transporters Pdr5p, Snq2p, and Yor1p by Pdr1p in Saccharomyces cerevisiae. *FEBS Lett.* **582,** 977–83 (2008).

27. Snider, J. *et al.* Mapping the functional yeast ABC transporter interactome. *Nat. Chem. Biol.* **9,** 565–72 (2013).

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

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

30. Yachie, N. *et al.* Pooled-matrix protein interaction screens using Barcode Fusion Genetics. *Mol. Syst. Biol.* **12,** 863 (2016).

31. Mani, R., St Onge, R. P., Hartman, J. L., Giaever, G. & Roth, F. P. Defining genetic interaction. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 3461–6 (2008).

32. Khakhina, S. *et al.* Control of Plasma Membrane Permeability by ABC Transporters. *Eukaryot. Cell* **14,** 442–453 (2015).

33. Tarassov, K. *et al.* An in vivo map of the yeast protein interactome. *Science* **320,** 1465–70 (2008).

34. Schlecht, U. *et al.* Multiplex assay for condition-dependent changes in protein-protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 9213–8 (2012).

35. Brem, R. B. & Kruglyak, L. The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 1572–7 (2005).

36. Perlstein, E. O., Ruderfer, D. M., Roberts, D. C., Schreiber, S. L. & Kruglyak, L. Genetic basis of individual differences in the response to small-molecule drugs in yeast. *Nat. Genet.* **39,** 496–502 (2007).

37. Threadgill, D. W. & Churchill, G. A. Ten Years of the Collaborative Cross. *Genetics* **190,** (2012).

38. C. elegans Deletion Mutant Consortium. Large-Scale Screening for Targeted Knockouts in the Caenorhabditis elegans Genome. *G3 Genes,Genomes,Genetics* **2,** 1415–1425 (2012).

39. Thompson, O. *et al.* The million mutation project: A new approach to genetics in Caenorhabditis elegans. *Genome Res.* **23,** 1749–1762 (2013).

40. Morawska, M. & Ulrich, H. D. An expanded tool kit for the auxin-inducible degron system in budding yeast. *Yeast* **30,** 341–51 (2013).

41. Brockmann, M. *et al.* Genetic wiring maps of single-cell protein states reveal an off-switch for GPCR signalling. *Nature* **546,** 307–311 (2017).

42. Emanuel, G., Moffitt, J. R. & Zhuang, X. High-throughput, image-based screening of genetic variant libraries. *bioRxiv* (2017).

43. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6,** 343–5 (2009).

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

45. Proctor, M. *et al.* in 239–269 (Humana Press, 2011). doi:10.1007/978-1-61779-173-4\_15

46. Snider, J. *et al.* Detecting interactions with membrane proteins using a membrane two-hybrid assay in yeast. *Nat. Protoc.* **5,** 1281–1293 (2010).

**Figures**

**Figure 1.** Overview of the engineered population profiling process.

A strain with 16 ABC transporter deletions was crossed with a pool of wild-type cells, each containing a DNA barcode. Genotype is schematized by the grid drawn in each cell (black = knockout, white = wild type). Diploids from this cross were subject to sporulation and barcoded haploids were then selected. Single colonies were picked and arrayed onto a series of 384-well plates. *En masse* genotyping was performed on this collection using an adapted RCP-PCR30 strategy, and genotypes were linked to strain-specific DNA barcodes to allow pooling and strain identification. High throughput pool barcode sequencing at different time points allowed determination of a drug by strain resistance profile, creating a profiled engineered population.

**Figure 2.** Exploring multi-knockout fitness landscapes within groups of 6-gene knockouts.

**A** Strains were grouped on knockout genotypes at *pdr5∆, snq2∆, ybt1∆, ycf1∆,* and *yor1∆*. Each point represents a group of strains containing the 6-locus genotype indicated by the legend. Strains in each group vary at the remaining 10 loci. Each point represents the median log2(resistance) of each group in the MAT**a** (x-axis) and MAT**α** (y-axis) pools.

**B** Distribution of MAT**a**–MAT**α** correlations of the grouped log2(resistance) profiles in A) amongst all drugs tested.

**C** A radial fitness landscape in benomyl showing all multi-knockout paths. Each graph is centred by the 6-locus wildtype group, and concentric extensions represent further knockouts. Each section is coloured by the mean resistance of the indicated 6-locus group. Genotypes of outwardly connected concentric sections differ by the addition of the labelled knockout. Sections are coloured by the mean resistance of each group relative to the 6-gene wildtype (black for reference), and the colour scale extends in both directions by the largest observed difference in resistance between the 6-gene wildtype and any other group (blue in the positive direction, orange in the negative direction). All paths where any further multi-knockouts do not result in significant changes in drug resistance (Bonferroni-adjusted *p* < 0.05, Mann-Whitney U test) are pruned.

**D** As in B, showing radial fitness landscapes for 10 additional drugs.

**Figure 3.**  Formalizing and interpreting surprising multi-gene knockout phenotypes.

**A** A linear landscape of resistance to mitoxantrone, benomyl, and fluconazole in the MAT**a** pools. Each point represents a group of strains containing the 6-locus genotype indicated by the legend. Strains in each group vary at the remaining 10 loci. Groups are arranged horizontally by the number of knockouts (with jitter added for visual clarity), and vertically by median drug resistance. Groups separated by single knockouts are connected – those with significant differences in resistance (Bonferroni-adjusted *p* < 0.05, Mann-Whitney U test) by solid lines, and otherwise by dashed lines. Representations for all pools are available in Fig. S7.

**B** Distribution of fluconazole resistance amongst all *ybt1∆,* *yor1∆*, *snq2∆, ycf1∆,* and *pdr5∆* knockout groups in the MATa pool. Group genotype is indicated on the sides using the same representation as in A), and *pdr5∆* groups are paired with their corresponding *PDR5* equivalent in each line. The density of the resistance distribution in each group is drawn on each line.

**C** A linear model was used to formally determine significant gene knockout and genetic interaction effects mediating resistance to the tested drugs. Linear model terms which were significant (Bonferroni adjusted *p* < 0.05) in both MAT**a** and MAT**α** pools for their given drug are coloured according the legend on the left. Other terms are coloured in grey. Coefficents are sorted by the number of genes involved, as indicated by the colour code on the left.

**Figure 4.** XX Under construction. (Meant to show validation for the fluconazole experiment).

**A** Linear models and multi-knockout fitness landscapes were interpreted into a drug-dependent model of transporter-transporter relationships. Antagonism indicates evidence that the presence of one transporter negatively affects the ability of another to mediate drug resistance, and parallel clearance indicates evidence that multiple transporters are responsible for resistance to a drug. Data S8 lists the genetic evidence for each relationship.

**Figure S1.** Analysis of genotyping data.

**A** Estimates of per-gene genotyping accuracy at the 16 ABC transporters surveyed. Accuracy was estimated by evaluating the performance of the *en masse* genotyping procedure on a set of known reference strains (Methods, Data S2).

**B** Tests of gene linkage within the MAT**a** (upper triangle) and MAT**α** pools (lower triangle). The Pearson correlation coefficient of the corresponding genotype pairs are indicated on the right. Pairs without significant correlation (Bonferroni-corrected *p* value > 0.05) are indicated in grey.

**C** Distribution of knockouts in the combined MAT**a** and MAT**α** pools. Observed number of strains with a given number of knockouts are indicated in grey, and the expected number of strains with a given number of knockouts at 93.8% genotyping accuracy under a random assortment model are indicated in black.

**Figure S2.** Correlation of fitness estimates from UP and DN tag counts.

**A** Correlation of resistance estimates in the MAT**a** pool under each drug obtained using UP tag counts (x-axis) compared to those obtained by DN tag counts (y-axis). Strains missing either a mapped UP or DN tag were excluded from the analysis.

**B** As in A), using data from the MAT**α** pool.

**C** Comparing UP and DN tag correlation with pool variability in resistance (as ). Minimum from UP and DN tag estimates is plotted for each pool.

**Figure S3.** Reproducible marginal gene knockout resistance effects in the pool.

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

**Figure S4.** Profiles of grouped genotype data.

Strains were grouped on knockout genotypes at *pdr5∆, snq2∆, ybt1∆, ycf1∆,* and *yor1∆*. Each section in the heatmap represents the average resistance of strains with the genotype indicated on the y-axis, grown in the pool indicated on the x-axis. Pools are arranged by hierarchical clustering of the grouped resistance data using complete linkage.

**Figure S5.** Reproducibility of grouped genotype resistance.

Strains were grouped on knockout genotypes at *pdr5∆, snq2∆, ybt1∆, ycf1∆,* and *yor1∆*. Each point represents a group of strains containing the 6-locus genotype indicated by the legend. Strains in each group vary at the remaining 10 loci. Each point represents the median log2(resistance) of each group in the MAT**a** (x-axis) and MAT**α** (y-axis) pools.

**Figure S6.** A radial fitness landscape in six additional drugs.

A radial fitness landscape in six drugs showing all multi-knockout paths. Each graph is centred by the 6-locus wildtype group, and concentric extensions represent further knockouts. Each section represents the mean resistance of a 6-locus group. Genotypes of outwardly connected concentric sections differ by the addition of the labelled knockout. Sections are coloured by the mean resistance of each group relative to the 6-gene wildtype (black for reference), and the colour scale extends in both directions by the largest observed difference in resistance between the 6-gene wildtype and any other group (blue in the positive direction, orange in the negative

**Figure S7.** A linear landscape of resistance to 16 drugs.

**A** A linear landscape of resistance to all tested drugs in the MAT**a** pools. Each point represents a group of strains containing the 6-locus genotype indicated by the colour code on the right. Strains in each group vary at the remaining 10 loci. Groups are arranged horizontally by the number of knockouts (with jitter added for visual clarity), and vertically by mean drug resistance. Connections are drawn between groups separated by single – those with significant differences in resistance (Bonferroni-adjusted *p* < 0.05, Mann-Whitney U test) by solid lines, and otherwise by dashed lines.

**B** As in A), with values shown for the MAT**α** pool.

direction). All paths where any further knockouts do not result in significant changes in drug resistance (Bonferroni-adjusted *p* < 0.05, Mann-Whitney U test) are pruned.

**Figure S8.** Reproducibility of linear models.

For each drug, linear model predictions of drug resistance are indicated on the y-axis and observed values are indicated on the x-axis. The axis in each plot is labelled by the pool mating type which was used to train the linear model (x-axis) or validate its predictions (y-axis).