**Deciphering complex traits with deep combinatorial genetic analysis**

Albi Celaj1,2,3, Marinella Gebbia1, Louai Musa2, Atina Cote2, Minjeong Ko2,6, Jamie Snider1, Victoria Wong1, Tiffany Fong5, Paul Bansal1,2, Joe Mellor1, Gireesh Seesankar5, Maria Nguyen5, Shijie Zhou1, Igor Stagljar1, Nozomu Yachie4,7, and Frederick Roth1,2,3,6,7,8

[Author list and order is not final]

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

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

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

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

5McMaster University, Hamilton, Ontario, Canada.

6Canadian Institute for Advanced Research, Toronto, Ontario, Canada.

7Department of Computer Science, University of Toronto, Toronto, Ontario, Canada.

8Corresponding authors

**Corresponding Author Information:**

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

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

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

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

**Abstract**

Many traits are controlled by complex biological systems encoded by multiple genes. Understanding complex genotype-to-trait relationships may require perturbing genes in many different combinations and observing the impact. Here we describe a method to efficiently engineer and phenotype many multi-gene variant combinations within a targeted gene set, enabling a deep combinatorial genetic analysis (DCGA). We use this method to generate and profile each of ~6,000 multi-knockouts amongst 16 yeast ABC transporters, representing 8% of 216 (65,536) possible combinations. We measured the genotype of each combinatorial knockout strain and its resistance to each of 16 bioactive compounds (‘drugs’). The resulting genotype-to-resistance landscapes revealed many complex drug-dependent genetic interactions. For example, we identified a quadruple knockout (*snq2*∆ *yor1*∆ *ybt1*∆ *ycf1*∆) which conferred resistance to both fluconazole and ketoconazole, further finding that addition of *pdr5*∆ yielded a quintuple mutant with high sensitivity to azoles. We used a non-linear computational model of underlying genetic relationships to guide further experimental characterization of the azole-resistant quadruple knockout. Together, our results show that DCGA can discover unexpected high-order genotype-to-trait relationships and model them to better understand complex biological systems.

**Introduction**

Extensive functional interdependency and redundancy in many biological systems results in traits which cannot be straightforwardly understood by observing the individual effects of sequence variants1–4. Genes encode gene products which often form interdependent pathways and protein complexes, such that combinatorial genetic perturbations can yield surprising phenotypes. This phenomenon defines genetic interaction. Observing the phenotypic effects of genes varied in combination, i.e., performing a combinatorial genetic analysis (CGA), can uncover functional dependencies between genes, and can be used to reconstruct large-scale maps of gene co-function5. The ability of CGAs to better understand gene function has been amply demonstrated by comprehensive two-gene interaction maps in yeast via growth profiling of >23 million different double knockout combinations5. Large scale-efforts to map two-gene interactions in human cell lines are also underway, similarly demonstrating their power to better understand gene function6,7. In yeast, the number of genes yielding a phenotype under standard growth conditions climbs from ~2,000 to ~5,000 when genes are perturbed in pairs5,8. Furthermore, the resulting genetic interaction map enabled genes to be clustered into functional groups5,9, while smaller-scale CGAs have been used to better predict the genetic basis of drug resistance10, and to inform order-of-action in biological pathways11.

CGA of many biological traits has shown that additional genetic interactions can arise from the simultaneous perturbation of more than two genes. In yeast, triple mutant interactions (for which a triple mutant phenotype cannot be simply explained by the component single and double mutant phenotypes) are likely to vastly outnumber two-gene interactions12. Triple mutant interactions can arise from diverse pathway architectures13, and complex multi-variant effects involving more than three genes have been reported (e.g. a five14, seven15 and >20 variant interaction16). In the simplest cases, such interactions arise from redundancy in large gene families. For example, in *Arabidopsis*, a method has been developed to simultaneously disrupt multiple paralogs with a single artificial miRNA to better study the function of such families17, while the genetic interactions within them have been less explored. Complex interactions may also have medically-relevant phenotypes, with CGA of antibiotic resistance in *E. coli*, for example, suggesting that multi-variant interactions can enable many mutational paths towards resistance18. In vertebrates, complex multi-variant effects mediate disease, e.g., myeloid malignancies19,20. Moreover, discovery of such interactions can be practically useful. For example, the induction of pluripotent stem cells requires the simultaneous exogenous expression of four genes21.

While two-knockout CGA has been used extensively to genetically dissect and order biological pathways11, CGA studies of higher-order genetic interactions have been few and smaller in scope. Performing exhaustive ‘deep’ combinatorial genetic analysis (DCGA) has been limited by the experimental challenge of generating and characterizing the vast number of mutant combinations required for such studies. Genome-scale DCGA of three-gene combinations will likely remain out of reach for years to come. Although DCGA can be targeted towards smaller biological subsystems, the large-scale engineering and profiling of many multi-variant strains is a major bottleneck even in yeast. Exhaustive DCGA for a relatively modest target set of 10 genes would require construction of 1,024 haploid strains to sample all combinations of two alleles per gene, or 1M strains if diploid genotypes were considered. Thus, even targeted DCGA remains a challenge requiring technological improvements. Although there are methods to generate multi-mutant strains that can circumvent the limited number of usable selection markers, these have focused on construction of single multi-mutant strains22. While methods exist to make modifications at multiple loci simultaneously (multiplex automated genome engineering – MAGE)23,24, major challenges remain in isolating and genotyping the large number of strains required to perform a DCGA. Extensions of MAGE have been developed to allow parallel phenotyping of many strains for DCGA in *E. coli*25,26, but exhibit high variance across biological replicates, perhaps due to currently-limited accuracy of large-scale genotyping. Methods have been described for parallel generation and parallel phenotyping of yeast [cite Diaz-Mejia MSB 2018] and human cells27, but the resulting CGA studies have not gone beyond two-gene combinations.

Here we describe an ‘engineered population profiling’ strategy enabling DCGA in yeast. We apply this strategy to a target set of all 16 yeast ABC transporters implicated in multi-drug resistance, carrying out high-order DCGA for each of 16 drug resistance phenotypes. ABC transporters were chosen as the pilot gene set for several reasons: First, ABC transporters are an important and conserved gene family which mediates functions such as multidrug resistance, disease progression, and basic cellular homeostasis28,29. Indeed, ABC transporters are part of a large gene family with over 10,000 members across all three domains of life30. Second, although ABC transporters are generally thought of as imparting drug resistance as one might expect for an efflux pump (and indeed the ABC-16 strain is generally more drug sensitive22), ABC transporter knockouts can mediate either drug sensitivity or resistance, and some two-gene ABC transporter knockouts have exhibited synergistic drug resistance22,31,32. Complex dependence between mammalian ABC transporters has also been observed, e.g. compensatory activation of ABCC3 upon disruption of ABCC2 in rats33 and in humans in the context of Dubin-Johnson syndrome34. In another example, mouse ABCG5 and ABCG8 are activated in response to disruption of ABCG2 (a protein that confers breast cancer xenobiotic resistance)35. Finally, a DCGA study of these 16 transporters is made simpler by the fact that the ABC-16 strain does not show major fitness defects in the absence of drugs. Therefore, we expect progeny bearing a subset of these 16 knockouts will generally be viable, enabling study of the full range of genotypes across a range of different drug exposures.

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

**Results**

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

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

A simple yet powerful way to generate a complex population is to cross two outbred individuals, such that offspring each inherit a random variant at each position of unlinked variation that differs between the parents10. Genotyping and profiling progeny for traits such as gene expression36 or small molecule resistance37 then allows statistical association of the resulting variants to the measured phenotypes. However, such approaches have been traditionally used at a large scale with natural isolates rather than engineered strains37, presenting several limitations. For example, many yeast genes known to be important for drug resistance, such as ABC transporters38, were undetected in such studies due to limited natural variation in parental strains. Although the use of diverse parents differing at hundreds of thousands of positions often results in associations of a single locus to a trait, there are typically many linked variants at each locus which makes it difficult to pinpoint the causal variant(s). A large number of positions varying between parents brings multiple testing issues that may require a prohibitive number of individuals for statistical reconstruction of complex variant-to-phenotype associations. To extend cross-based approaches beyond natural strains, we therefore designed an engineered population strategy in which all variation of interest is engineered into one or a few individuals, and these individuals are then crossed to yield a population of random segregants.

A DCGA study requires that each individual progeny strain be genotyped and phenotyped. For this purpose, we wished to enable tracking of individual progeny. We therefore designed the process so that one of the haploid parental strains is transformed with a complex pool of random DNA barcodes39, such that each cell of one parental strain bears a single specific random barcode. Each haploid progeny cell resulting from the cross will then be barcoded. If the number of random sequences in barcoded parental cells vastly exceeds the number of progeny, then progeny barcodes will generally be unique. As described below, the unique tracking identifier facilitates large-scale genotyping and phenotyping of progeny. Isolating a strain, sequencing its identifier barcode, and performing PCR-based genotyping, for example, associates the identifer barcode with a genotype, thereafter allowing for a ‘barcode-to-genotype lookup’. An individual barcode identifier also allows for straightforward growth-based phenotyping, in that relative strain abundance measured over time in a competitive pool using high-throughput barcode sequencing can be interpreted as a phenotype40. Thus we developed the concept of combining a cross-based approach with a large trackable engineered population of genotyped multi-variant individuals. This trackable genotyped population can be stored as a pool and aliquots of the pool can be interrogated for various phenotypes by tracking competitive growth of each strain in parallel under multiple conditions.

**Generating a large pool of barcoded parental cells**

A key step in the above population engineering scheme is the introduction of a unique molecular barcode into the genome of each individual. For this, we adopted previously-described methods to create a large pool of uniquely-identifiable clones for one of the parental strains39[ref MSB 2018]. Briefly, to introduce a high diversity of unique barcodes into a ‘wild-type’ parental strain (RY0148), we used a Cre-mediated recombination strategy. First, we introduced a ‘landing pad’ into RY0148 via the addition of Loxp and Lox2272 recombination sites at the HO locus. Then, we transformed this landing-pad strain with a heterogeneous plasmid pool, such that each plasmid bears a random DNA barcode and a hygromycin resistance marker flanked by Loxp/Lox2272 recombination sites, as well as a counterselectable marker outside of the recombination sites (Fig S1). Next, we induced the expression of Cre to allow for recombination and integration of the plasmid-borne barcodes into the HO locus (Fig S1). Finally, we selected for integrants and counter-selected plasmids. The resulting barcoded pool is a generally useful reagent that can be employed in different crosses for different DCGA studies (Fig. S1, see Methods for details).

**Engineering a population of high-order combinatorial ABC transporter knockouts**

After creating a universal barcoded pool with a wild-type parental strain background, we crossed this pool en masse with a previously-generated “ABC-16 strain” bearing knockouts of all 16 ABC transporters implicated in multidrug resistance22. The ABC-16 strain had previously been engineered to contain all markers necessary to perform mating, sporulation, and selection for haploid cells, while the barcoded wild-type parent provided the marker necessary to select for cells carrying a barcoded HO locus (Methods). After selection for barcoded haploid progeny of the cross, automated colony-picking enabled isolation of an arrayed collection of ~5,000 MAT**a** and ~5,000 MAT**α** segregants in 384-well plates. Thus, we generated an engineered population in which each individual haploid strain bears a random subset of knockout alleles for the target set of 16 ABC transporters.

**Genotyping and Barcode Identification for the Engineered Population**

After creating a large collection of barcoded multi-knockout progeny, we genotyped each strain and linked this genotype to an individual DNA barcode identifier *en masse*. For this purpose, we exploited the fact that the ABC-16 strain was derived from crosses between barcoded YKO yeast deletion collection strains8,22, so that each knockout carries at least one identifying barcode that flanks and uniquely identifies the deleted gene. We adapted our previously-described row-column-plate PCR (RCP-PCR) strategy41, which allows amplification of barcodes in each segregant while introducing additional index tags that identify the plate, row, and column of origin for each amplification product (Methods; Fig. 1). Thus, a single next generation sequencing reaction can reveal both the strain-specific tracking barcode at the HO locus and the identity of every gene deleted in the segregant at each plate location (Methods; Fig. 1). To validate and calibrate the genotypes determined by high-throughput-sequencing, multiple replicates of 40 individually genotyped ‘gold standard’ strains, as well as two additional control strains with known genotypes, were added to the collection at defined positions (Methods; Data S2). Using data from calibration strains, we estimated an overall genotyping accuracy of 93.2% (Fig. S2A, Methods). An independent method relying on the distribution of knockouts in the pool estimated a similar overall accuracy of 93.8% (Fig. S2C, Methods). Based on the genotyping data, all genes were either unlinked or weakly linked except for *BPT1* and *YBT1* (Fig. S2B; r = 0.49), which are separated by 70.1kb on chromosome XII. Surprisingly, three gene pairs – *YOR1*-*YCF1*, *YOR1*-*BPT1*, and *SNQ2*-*PDR5* – exhibited weak but significant negative correlation in the appearance of KO genotypes (-0.04 ≥ r ≥ -0.08) (Fig. S2B). This effect may have arisen via a negative genetic interaction conferring lower growth for the corresponding double-knockout genotypes during the sporulation, haploid selection, or automated colony picking steps. Overall, usable genotyping data linked to at least one unique DNA barcode was obtained for 6,709 progeny (XX MAT**a** and XXY MAT**α** strains) with 5,095 unique genotypes. Segregant strains for which both genotyping and tracking barcode identification were successful were combined by mating type to yield a MAT**a** and a MAT**α** pool.

**Phenotyping the Engineered Population for Diverse Drug Resistance Traits**

Knowledge of the tracking barcode for each segregant enabled us to profile each strain’s resistance or sensitivity to particular drugs42. Strain pools were grown competitively in each of 16 different anticancer and antifungal drugs (Data S3), as well as a solvent control. Using high-throughput strain barcode sequencing40, strain frequency was measured at five time points (corresponding to 0, 5, 10, 15, and 20 generations of overall pool growth, Fig. 1) and a drug resistance score was computed for 3,221 MAT**a** and 3,592 MAT**α** strains [XX numbers don’t add up] by comparing estimated growth in each drug to the solvent control (Methods, Fig. 1, Data S5).

**Inferring genotype-phenotype relationships in an engineered population**

By combining genotype data and with the time-course of barcode abundance for each strain, we sought to both infer phenotypes for each segregant and to identify genotypes associated with particular phenotypes.

To identify gene deletions having a drug-independent growth impact (see Methods), we examined barcode abundance in the pre-selection pool, finding that strains bearing deletions in the [XXX, XXY and XXZ] genes were more likely to be absent or weakly detected, suggesting potential drug-independent growth impacts [XXFisher’s exact test?]. We next used post- and pre-selection barcode counts to infer a growth rate under the DMSO condition for each segregant. Genotypic association was assessed by aggregating data from strains in the population with or without particular knockouts (over all genetic backgrounds), using a linear model with only single-gene terms to uncover ‘marginal’ associations of each knockout to the inferred growth rate for each strain under the no-drug condition (Methods). In this way, we found [XXX, XXY and XXZ] genes to have reduced representation after no-drug selection, confirming findings based on absence from the pre-selection pool [XX?]. For all subsequent analysis, strains were excluded from analysis if they were initially absent in the solvent control (XXX strains) or exhibited a substantial drug-independent growth defect (XXY strains; Data S5).

We next sought to identify the subset of ABC transporters which are most relevant for resistance or sensitivity to the drugs we assayed. Using the barcode abundance-based time-course for each segregant under no-drug and drug conditions, we inferred drug resistance (growth rate under the drug condition relative to the no-drug condition) for each segregant for each drug. Applying a linear model, we tested for association between each gene knockout and each drug resistance. While prior literature had reported on the effects of *snq2∆*, *pdr5∆*,and *yor1∆* for many of the drugs we tested, our approach also found many novel associations for deletions of *ycf1∆*, *ybt1∆*, two vacuolar ABC transporters (Fig. S4, Data S6). Combining both the MAT**a** and MAT**α** pools (Data S6), 16 out of 21 previously-reported single-knockout phenotypes were reproduced, including 6 out of 7 of those which had been reported in at least two publications (Fig. S4; Data S7). In addition, we found [XXX] strong gene-drug associations and [XXY] weak gene-drug associations. Taken together, we detected [XXX%] of previous associations while revealing [XXX+XXY] new associations between drug resistance and specific gene deletions.

Taking advantage of the fact that our engineered population contained diverse combinations of knockouts, we sought to tease our combinatorial effects. Because 87% of the single-gene associations involved only five ABC transporters, *snq2∆*, *pdr5∆*, *yor1∆*, *ycf1∆*, and *ybt1∆*, we restricted our attention to these transporters. Considering only the knockout status at each of these five loci, there are 25 = 32 possible genotypes. By aggregating the phenotypic values within each group, we can provide a phenotypic profile for each drug for each of 32 genotypes (averaging over genetic backgrounds at the 11 excluded loci). These profiles were initially calculated separately for MAT**a** and MAT**α** strains (Fig S5). Strikingly similar (r ≥ .99) profiles were observed between MAT**a** and MAT**α** populations for camptothecin and tamoxifen (Fig. 2A). Indeed, high reproducibility (r ≥ 0.95) between these independent biological replicate pools was observed for 13 of 16 drugs (Fig. 2B).

To visualize phenotypes arising from complex combinatorial knockout combinations, we developed a radial representation in which the drug resistance consequences of knocking out increasingly many ABC transporters in a specific order can be explored by tracing different paths leading outward from the central wild-type genotype (Fig. 2C). Reflecting the quantitative reproducibility of our profiles, graphs were visually similar between independent biological replicate MAT**a** and MAT**α** populations for many drugs, while showing substantial differences only for colchicine (Fig. S7). Given the high reproducibility, we merged the MAT**a** and MAT**α** data for all subsequent analyses, except where noted (Methods).

**DCGA Reveals High-Order Combinatorial Drug Resistance Effects**

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

Visualizing the knockout profiles for each genotype in a fitness landscape representation (Fig. 3A), we first verified that these fitness landscapes could capture previously-reported relationships between ABC transporters and benomyl resistance. Our knockout profiles clearly captured the sensitivity of *snq2*∆ deletions to benomyl (Fig. 3A left panel; *p* < XX; Wilcoxon rank sum test), which was expected given that *SNQ2* is known to be the primary efflux pump for benomyl43. We also observed several previously-reported phenomena, including: increased benomyl resistance in *pdr5∆* knockouts ([effect size]; p < XX), further increased benomyl resistance of the *pdr5∆ yor1∆* double-mutant ([effect size]; p < XX), and dependence of these increases on the presence of *SNQ2* 31,32([compare pdr5 effect in the presence and absence of SNQ2 AND compare pdr5 yor1 effect in the presence and absence of SNQ2][effect size]; p < XX; Fig. 3A left panel). We did not observe *yor1∆* to confer benomyl resistance (p = XX), a phenomenon that was previously reported to be relatively weak32. Thus, engineered population profiling could largely recapitulate previously-reported ABC transporter knockout relationships to benomyl resistance, including two- and three-gene combinatorial effects.

In other cases, these fitness landscapes interactions suggested efflux action in parallel. Specifically, we saw gene sets where each individual knockout shows sensitivity to a drug, and each higher-order knockout combination exhibits drug sensitivity that is higher than any of the individual component knockouts. Examples of this include the set *snq2∆* and *pdr5∆* under camptothecin (Fig. S8), and the *snq2∆*, *pdr5∆*, *ybt1∆* and *yor1∆* set under mitoxantrone (Fig. 3A middle, S8). These sensitivity patterns are consistent with a simple scenario in which each transporter in the set is able efflux a given drug.

In perhaps the most surprising cases, the fitness landscapes involved complex multi-knockout patterns mediating both drug resistance and sensitivity. For many compounds, multiple paths of successive introduction of deletions led to greater resistance, yielding ‘peak’ strains that were considerably more resistant than a wild-type cell (Fig. S8). For example, under fluconazole, ketoconazole, and itraconazole, a quadruple deletion – *snq2∆ybt1∆ycf1∆yor1∆* – led to resistance. Interestingly, the quintuple mutant *pdr5∆snq2∆ybt1∆ycf1∆yor1∆* (differing from the resistant quadruple mutant in that it bears an additional *pdr5∆* deletion) showed sensitivity that was comparable to the single-knockout *pdr5∆* genotype (Fig. 3A right, S8). The deletions of *snq2∆*, *ybt1∆*, *ycf1∆*, *yor1∆* showed synergistic resistance in that combinations of one or two knockouts within these four genes resulted in subtle or absent effects. These findings extend previous findings that deletions of *SNQ2* or *YOR1* (either alone or together) increase resistance to fluconazole and ketoconazole, and that *snq2∆* and *yor1∆* deletions result in increased *PDR5* expression and activity31. Specifically, our results (Fig. 3A, right) suggest that: 1) resistance is increased further by *ybt1∆* and *ycf1∆* knockouts in addition to *snq2∆* and *yor1∆*; 2) the effects *ybt1∆* and *ycf1∆* are synergistic with *snq2∆* and *yor1∆*; and that 3) resistance of the *snq2∆ybt1∆ycf1∆yor1∆* knockout depends on *PDR5*. Because each genotype group contains a heterogeneous population of individuals with additional background knockouts, we visualized the distribution in fluconazole resistance in each group (Fig. 3B). The high variability within many groups suggests the presence of more complex multi-knockout effects not captured by these five genes.

We expanded our search for multi-gene effects using an extension (see Methods) of the linear model described above in the context of single-gene effects. All single and multi-gene interactions that passed the significance test (*p* < 0.05 after adjusting for multiple testing) are shown in Figure 3C. This analysis yielded genetic interactions involving two or more genes for fifteen out of sixteen (94%) of the drugs examined (Fig. 3C). For beauvericin, the only exception, all strains with *YOR1+* genotypes were completely insensitive as previously described44, while all *yor1∆* genotypes lacked any detectable growth. We have not explored lower concentrations for beauvericin at which higher-order genetic interactions might yet be found. Higher-order genetic interactions (involving three or more genes) were observed for fourteen of sixteen (88%) of drugs tested (Fig. 3C). Here the exception (beyond beauvericin) was cycloheximide. For cycloheximide, we observed the previously-known strong single-gene *pdr5∆* effect32,43,45,46, many weak single-knockout effects, and only one weak two-gene interaction between *pdr5∆* and *snq2∆* (Fig. 3C). Thus, engineered population profiling revealed higher-order genetic interaction involving three or more genes for nearly all drug resistance phenotypes studied.

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

Formal complex genetic interaction analysis allowed finer parsing of the relationship between genes involved in a higher-order interaction. For example, the striking mitoxantrone sensitivity of the *snq2∆pdr5∆ybt1∆yor1∆* quadruple mutant was modelled as the combination of small marginal effects of *snq2∆* and *pdr5∆* alone, a two-gene negative interaction between *snq2∆* and *pdr5∆*, two three-gene negative interactions (between *snq2∆pdr5∆* and each of *ybt1∆* and *yor1∆*), and a four-gene *snq2∆pdr5∆ybt1∆yor1∆* negative interaction (reflecting the fact that the quadruple mutant is more sensitive than would be expected from the resistance of any of the three-deletion subset genotypes; Fig. 3C, Data S6). Together, these complex negative genetic interaction patterns suggest that the four genes enable mitoxantrone efflux in parallel. A similar ‘parallel action’ genetic interaction pattern was observed for *pdr5∆snq2∆yor1∆* in cisplatin (Fig. 3C, Data S6).

[Fritz stopped here]

Interestingly, the multi-knockout resistance phenotype in fluconazole was modelled (in addition to one- and two- gene effects), as the combination of three positive three-gene interactions involving *yor1∆*, *snq2∆, ybt1∆* and *ycf1∆* (Fig. 3C). The dependence of these multi-knockout effects on *PDR5*, however, was modelled as three two-way negative interactions involving *PDR5* and *ycf1∆*, *ybt1∆*, *snq2∆*. Complex positive interactions in several drugs were also found involving *pdr15∆*, *bpt1∆*, *adp1∆*, and *vmr1∆*, and in each case the knockout of these genes helped confer some resistance in a highly-sensitive multi-knockout background (Fig. 3C).

**Learning a non-linear ABC Transporter System Model**

While many ABC transporter knockout combinations affected drug resistance in a highly non-linear manner, they nevertheless suggested a relatively straightforward schematic of ABC transporter function. That is, many of the multi-knockout phenotypes and complex genetic interaction patterns were consistent with a combination of compensatory activation between ABC transporters, combined with their potential to impart drug resistance in parallel. To test the explanatory power of this schematic, and to determine which compensatory activation and drug efflux relationships were supported by the data in a rigorous manner, we defined a system model of ABC transporters which captures these relationships (Methods, Fig 4A). This system model was expressed as a constrained two-layer neural network, where the first layer consists of a set of negative weights connecting the ABC transporters to each other to model compensatory activation (i.e. ‘inhibition’ between transporters), and the second layer is a set of positive weights connecting each ABC transporter to each compound to model drug clearance. Notably, this model constrains the compensatory activation relationships to be fixed across the compounds tested, and varies only the clearance relationships between environments. The training procedure parameterizes this model by finding a set of weights which best predicts the phenotypes observed in our data (i.e. minimizes squared error), when the corresponding genotype is encoded as the set of transporters (‘neurons’) which are present in this network. The network is furthermore regularized to avoid the addition of non-predictive parameters (e.g. proposing the compensatory activation of a transporter which does not efflux any drugs), and each remaining non-zero weight is afterwards tested for statistical significance (Methods).

This non-linear system model could be fit to achieve high overall correlation with the observed phenotypes (*r* = 0.96, Fig. 4C). This was also true when the model was trained on one population and tested on another (*r* = 0.95, Fig XX). Ultimately, training on either population separately resulted in strong agreement in parameterization (*r* = 0.99, Fig XX). The fit parameters recapitulated the hypothesized compensatory activation of *PDR5* by knockout of *snq2∆, yor1∆, ybt1∆,* and *ycf1∆* (Fig. 4B)*.* The hypothesized redundant efflux of mitoxantrone byPdr5*,* Snq2*,* Yor1*,* and Ybt1 was also supported (Fig. 4B), and furthermore the varying weights between these transporters (Fig. 4B, Data XX) suggest a differential mitoxantrone clearance ability. For example, Snq2 is modelled as having the most clearance capability, and this is reflected in the fitness landscapes - *pdr5∆ybt1∆yor1∆* is similarly as resistant as the wildtype, whereas triple knockouts involving *snq2∆* and two of these three genes show considerable growth defects (Fig XX). The model also pointed to differential inhibitory effects between transporters – for example, Snq2 is predicted to receive a much greater inhibitory effect from *PDR5* than from *YOR1* (Fig XX). This is evident, for example, in the benomyl phenotypes, as *pdr5∆* leads to a clearer resistance effect than *yor1∆* (Fig XX).

Systematic differences in predicted and observed phenotype from this model were further analyzed, and suggested straightforward extensions of this schematic. For example, while *snq∆ybt1∆ycf1∆yor1∆* was indeed modeled as more resistant to fluconazole and ketoconazole than any subset of these knockouts, the four-knockout phenotype observed was more fit than would be predicted by an additive inhibition model (Fig. XX). The addition of a single hidden neuron mediating the inhibitory input from these four genes allowed modelling of non-additive inhibition, and more accurately captured the four-knockout phenotype (p < XX, Fig XX). Non-additive inhibition may be expected, for example, if inhibition is mediated by a hidden factor. In another example, observed landscape in valinomycin was overall poorly recapitulated by this schematic (XX). As it has been observed that the 16-deletion strain is more resistant to valinomycin than the wildtype, it is likely that the compensatory activation of other drug resistance genes beyond the ABC transporters mediate resistance to this compound and others22,47. To incorporate the compensatory activation of other resistance factors in predicting valinomycin resistance, we added an ‘always-present’ extra clearance neuron to the schematic which could be inhibited by the ABC transporters. This extension resulted in more accurate modeling of the observed valinomycin resistances (r = XX, p = XX, Fig. XX), and predicts the involvement of other clearance factor(s) which are activated in the response to *snq2∆*, *pdr5∆*, and *ybt1∆*. Compensatory activation of other clearance factors may also underlie the complex positive interactions involving *pdr15∆*, *bpt1∆*, *adp1∆*, and *vmr1∆* (Fig. 3C).

**Detailed Validation of a Complex Genetic Inhibition Model**

To validate the complex knockout-to-phenotype association mediating fluconazole resistance, we generated individual strains containing all knockout combinations of *pdr5∆, snq2∆, yor1∆, ybt1∆,* and *ycf1∆* in a wildtype background. These strains were generated by mating a chosen strain in the collection with a wildtype and performing a series of genotyping and selection steps (Fig 4B, Methods). Fluconazole resistance of these individual strains correlated well with competitive pool resistance, both when measured as the expected concentration to cause 50% inhibition (*r* = 0.92, Fig 4D), and as the individual growth at a concentration identical to the pool (*r* = 0.9, Fig. S10). Consistent with the pool data, *snq2∆yor1∆ybt1∆ycf1∆* had the highest resistance, and only minor resistance effects were seen for strains with one and two knockout combinations amongst these four genes. Given the ‘non-additive parallel repression’ model of Pdr5 by the presence of *SNQ2*, *YOR1*, *YBT1*, and *YCF1*, we explored two potential mechanisms of inhibition. Inhibition of Pdr5p by *SNQ2* and *YOR1* has been previously reported, and is thought to be indirectly mediated by transcriptional activation of Pdr1 through an unknown mechanism31 To investigate whether *PDR5* mRNA abundance was indeed elevated in *snq2∆yor1∆ybt1∆ycf1∆* compared to the wild type, we performed qRT-PCR. *PDR5* transcript abundance was elevated 2.1 fold in *snq2∆yor1∆ybt1∆ycf1∆* relative to the wild type (*p* = 0.032; Fig 4D). While *snq2∆yor1∆* was expected to result in weaker (~1.5 fold) *PDR5* induction as previously reported31, here the observed ~1.3 fold difference was not statistically significant (p = 0.27, Fig 4E), and no evidence of mRNA induction in the *ybt1∆ycf1∆* strainwas found(0.9 fold expression, p = 0.69 , Fig 4E).

To further investigate the predicted repression of Pdr5 by the membrane transporters *SNQ2* and *YOR1*, we explored the possibility of an alternate mechanism based on physical interactions. Consistent with the lack of transcriptional induction in the lower-order knockouts observed here, a previous study investigating an increase in Snq2-mediated benomyl resistance by *pdr5∆* and *yor1∆* found evidence for mild *SNQ2* mRNA induction only in *pdr5∆yor1∆*, but not in either resistant single mutant32. *PDR5*-mediated repression of Snq2 was instead hypothesized to result from their direct protein-protein interaction32. Consistent with a physical repression model, homodimers of Pdr5, Snq2, and Yor1, as well as a Pdr5-Snq2 heterodimer have been detected either using the mDHFR protein complementation assay (PCA) 48,49 and/or using the membrane yeast-two-hybrid (MYTH)32 assay. While the functional significance of these homodimer interactions have been poorly explored, the Pdr5 homodimer has been characterized using single-particle electron microscopy, which shows a four-lobed asymmetric conformation suggesting a structural dependence between the two units50. Given the support of this model by a previously-known Pdr5-Snq2 heterodimer32,48, we further explored if Yor1 could cause similar repression through an unreported interaction with Pdr5. We tested for such a Pdr5-Yor1 interaction, finding evidence for it using MYTH (Fig. 4F, S12), but not PCA (Fig. S11). All previously-known MYTH and PCA interactions amongst Pdr5, Snq2, and Yor1 were also found by re-testing (Fig. 4F, S11, S12). Given a much-higher baseline abundance of Pdr5 than Snq251, such a ‘heterodimer repression’ model is consistent with the prediction that repression on Snq2 by Pdr5 should be greater than repression on Pdr5by Snq2 (Fig. 4B)*.* That is, each mutually-repressive interaction would affect a greater proportion of Snq2 than Pdr5. Taken together, these experiments suggest an inhibition model where all four genes appear tonon-linearly suppress *PDR5* expression, while milder effects with *snq2∆* and *yor1∆* may be a result of physically direct repression, for example by interfering with homodimerization.

**Discussion**

The lack of tools to efficiently engineer many combinations of gene variants and characterize their impact limits the use of in-depth genetic profiling to obtain a systems-level understanding of many complex traits. Here we illustrate a method to straightforwardly generate and profile a large population of multi-mutant individuals, permitting a DCGA. We demonstrated that a DCGA of ABC transporters revealed many informative complex multi-variant phenotypes, and that these could be used to computationally reconstruct a non-linear system model of the underlying genetic relationships. Ultimately, many novel gene functions and gene-gene relationships were revealed, even within this highly-studied gene family. These results motivate similar DCGAs of other gene groups, and the continued use of this engineered population to study the mechanisms of ABC-transporter-mediated drug clearance of other compounds over existing single-knockout approaches.

The proposed cross-based method enables the efficient introduction of complex combinations of mutations into a population of uniquely identifiable cells. This method is straightforwardly adaptable for use with other multi-gene deletions strains, such as a 16-deletion mutant in GPCR pathway-related genes52. In other model organisms such as *C. elegans*, methods to introduce targeted gene knockouts53 or loss-of-function mutations54 may enable analogous strategies. The cross-based approach allows mutations to be distributed between the two parents, permitting a similar strategy even if introduction of all mutations into a single individual is not possible. To further avoid such potential multi-variant ‘dead ends’, it is also possible to extend the cross-based strategy using a more complex series of matings between several multi-mutant parents.

To extend engineered population profiling beyond cross-based methods, further development of molecular tools may allow for the direct editing an isogenic population to introduce trackable multi-allele diversity. Such direct population engineering can allow similar studies of non-mating model systems, such as human cell lines. The ability to work with cell lines would enable, for example, an analogous DCGA of human ABC transporters, and a much better understanding of the sensitivity and compensatory activation relationships mediating chemotherapeutic compound resistance35. As direct population engineering presents more challenges than the engineering of one or a few parental strains, this is therefore an area of future development26,27.

The engineered ABC-16 population is by far the largest collection of genotyped multi-mutants amongst these genes. However, the majority of knockout combinations amongst the 16 ABC transporters (~92%) remain to be generated and characterized, and will likely reveal further complex multi-knockout effects. A major bottleneck to achieving the scalability required to generate the remaining strains was the requirement that a separate PCR reaction must be performed on each isolated individual in order to link genotype to a DNA barcode. A method to perform genotyping and DNA barcode mapping in a pooled (i.e. ‘population-wide’) format will likely permit the engineering of populations which are orders of magnitude larger than what was demonstrated here, allowing a better representation of all possible multi-knockout strains, and the exploration of complex interactions involving multiple alleles at multiple genes4.

A major element in the demonstrated engineered population profiling approach was the use of a strain-specific molecular barcode. Use of the appropriate selection conditions coupled with sequencing of molecular barcodes allows for many traits which can be measured and linked to a specific genotype *en masse*. Fluorescence sorting strategies can convert many fluorescence-based assays into a barcode sequencing output, such as those using a reporter to study the effects of knockouts on the activation of signaling pathways, phosphorylation state, epigenetic modifications, or protein abundance55. For example, this approach might be expanded such that drug efflux dynamics may be directly studied by incubating cells with fluorescence-conjugated drugs and measuring efflux-mediated reduction in fluoresence31 using a combination of cell sorting and barcode sequencing over time. In addition to DNA-based molecular barcodes, methods to genotype large populations of cells after imaging are being developed, which would allow high-content characterization of multi-knockout strains56. Thus, with the appropriate design, multi-knockout strains obtained by population engineering could be characterized for multiple complex phenotypes at a large scale.

We envision that the profiling of engineered populations will permit a DCGA of many biological systems in multiple organisms. The expanse of currently-available molecular tools will allow an extension of the demonstrated approach in scale, allelic variety, and complexity of phenotypic characterization. With ubiquitous complex genotype-to-phenotype relationships in biological systems, DCGAs may provide a much-needed possibility to dissect, reconstruct, and understand multi-gene systems in many living organisms.

**Materials and Methods**

**Yeast Strains**

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

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

RY0146 (Toolkit-a strain):

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

RY0148 (Barcoder Strain MAT**α**):

*MAT****α*** *lyp1Δ his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 can1Δ::GMToolkit-α (CMVpr-rtTA NatMX4 STE3pr-LEU2) ho*∆*::LoxP UP-tag HphMX4 DN-tag Lox2272*

**Media**

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

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

**Creating the Barcoder Plasmid**

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

First, a barcoded HphMX4 construct was created. HphMX4 was amplified from a pIS420 plasmid using the STEP1F and STEP1R primers containing HphMX4 homology and US2/DS1 overhangs (Data S1). The PCR program used for this step was 98°C for 30sec; 25 cycles of 98°C for 10sec, 59°C for 10sec, 72°C for 60sec; 72°C for 5min; 4°C forever. These PCR products were purified using a Qiagen Qiaspin kit and confirmed using 2% gel electrophoresis. To the resulting purified products, the STEP2F and STEP2R primers were used to add the random barcodes and US1/DS2 regions with the following PCR program: 98°C for 30sec; 25 cycles of 98°C for 10sec, 68°C for 10sec, 72°C for 60sec; 72°C for 5min; 4°C forever. These resulting products were again purified using a Qiagen Qiaspin kit and ~1.5-1.6kb products were confirmed using 2% gel electrophoresis. To add LoxP/Lox2272 sites, PCR was performed with the STEP2 products using the SacI-LoxP-HphMX4-Barcode-F / SacI-Lox2272-HphMX4-Barcode-R primers. The PCR program used for this step was: 98°C for 30sec; 26 cycles of 98°C for 15sec, 64°C for 20sec, 72°C for 65sec; 72°C for 5min; 4°C forever. The resulting PCR products were purified using a Qiagen Qiaspin Kit, and ~1950bp products were confirmed using 2% gel electrophoresis. Two PCR reactions were performed on the resulting products to confirm correct synthesis. The first PCR reaction was performed with the SacI Reamp F/US2 primer pairs, and the second was performed using DS1/SacI Reamp R primer pairs. The PCR program used for both of these reactions was: 98°C for 30sec; 25 cycles of 98°C for 10sec, 59°C for 15sec, 72°C for 30sec; 72°C for 5min; 4°C forever. Expected sizes (~132bp, 137bp) were confirmed using 4% gel electrophoresis. All above PCR reactions were performed using High Fidelity Phusion Master Mix (NEB).

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

**Generating a Barcoder Strain**

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

The HO::LoxP-URA3-Lox2272 integrant strain was then transformed with a mixture of digested pSH47 and purified PCR products to enable in-yeast-assembly57. Transformation was carried out using a previously established protocol58, with a ~1:6 mixture of digested pSH47:HphMX4 barcode cassette (~12μg digested pSH47 and 15μg cassette). Transformants were grown at 30°C in YPG +HygroB plates for 3 days, allowing both selection of successful transformants and Gal1p-Cre induction. These cells were then scraped and grown overnight in 5-FOA plates to select against non-recombinant strains, and strains containing the recombined barcoder plasmid.

Twenty colonies were tested for barcode integration using PCR and Sanger sequencing. Lysates were made by mixing a sample of each colony with 2μl Sterile DNA Free Water, 2μl 0.2M pH 7.4 Sodium Phosphate Buffer, 0.5 μl 5U/μl Zymoresearch zymolyase and incubated at 37°C for 25min and 95°C for 10 min, and stopped by adding 125μl of sterile DNA-free Water. To each lysed colony, two sets of primer pairs to verify the strain barcode-specific UP and DN tag - US2 and a sequence complementary to 5’ of the HO gene (5’HO); DS1 and a sequence complementary to the 3’ of the HO gene (3’HO), using the following program: 98°C for 30sec; 25 cycles of 98°C for 10sec, 59°C for 15sec, 72°C for 30sec; 72°C for 5min; 4°C forever. PCR reactions were performed using High Fidelity Phusion Master Mix (NEB) and analyzed using 4% gel electrophoresis to verify the presence of 263bp and 251bp bands. EXOSAP purification was performed on the PCR products by adding 10μl EXOSAP mix (0.025μl ExoI (0.5U), 0.1μl Antarctic Phosphatase (0.5U), 3.5μl 10X Antarctic Phosphatase Buffer, 6.375μl 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 described22. 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 described22, selecting for MAT**a** and MAT**α** progeny separately. Using colony plating, sporulation efficiency was estimated at 24% - 1080 colonies grew in SC, 140 colonies grew in SC –His (MAT**a** haploid selection), and 120 colonies grew in SC –Leu (MAT**α** haploid selection). The two pools were then grown in YPD +HygroB to select for barcoded haploids. The SC –Leu pool was further grown in SC –Ura to select against barcoder strain parents that may have escaped diploid selection. Using a QPix colony picker, 5,461 MAT**a**  and 5,461 MAT**α** colonies were picked onto 384 well plates. In addition, 299 known positions in both the MAT**a** and MAT**α** arrayed collections consisted of known strains – either one of 40 ‘Gold Standard’ genotyped strains, RY0148, or RY0622 – to act as genotyping controls (Data S2).

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

**Pooled Strain Genotyping**

A previously-developed Row-Column-Plate (RCP)-PCR protocol41 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 primers41. 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 described41. A pair of PXX\_PE1.0 and PYY\_PE2.0 primers (Data S2) were added to 3-6μl pooled products (calibrated to ~150ng) from each plate to encode the plate of origin, and were amplified using the following PCR program: 98°C for 30sec; 15 cycles of 98°C for 10sec, 59°C for 15sec, 72°C for 40sec; 72°C for 2min; 4°C forever. All PCR reactions above were performed using High Fidelity Phusion Master Mix (NEB).

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

**Analysis of Pooled Strain Genotyping Data**

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

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

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

**Examining Putative Wild-Type Pool Strains**

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

**Estimating Genotyping Accuracy by Knockout Distribution**

To lend independent support to the genotyping accuracy determined by gold standard strains, an alternate method based on the distribution of knockouts in the population was used. Since *en masse* genotyping associates barcode sequences with ABC transporter knockouts, the absence of a given barcode implies either a wild-type genotype at that locus or a failure in amplification, sequencing, or calling. Conversely, cases where a wild-type is called as a mutant are expected to be comparably rare. Excess wild-type calls lead to a reduction in the average number of knockouts in the pool, and can be used to estimate genotyping accuracy. The average number of knockouts in the pool was 7.0, lower than the 8 expected with perfect genotyping. If there are no wild-type to mutant miscalls, this number is most likely with an asymmetric genotyping accuracy of 93.8%, compared to the 93.2% estimated by comparison to gold standards (Fig. S2C).

**Indiviual Liquid Growth Profiling**

To measure individual strain growth, the 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 metric59, 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**

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

If both an UP and DN tag for a given strain were successfully linked to a genotype, is estimated from the UP and DN counts and averaged, otherwise the available tag is used. The frequency of each strain was then converted into a ‘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) to estimate the individual abundance over time of each strain, then taking the integral over all measured timepoints to (the total number of pool generations measured). Frequencies between measured timepoints were linearly interpolated.

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

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

We substitute for and for . We then numerically solve for the which satisfies this relationship using the optimize() function in R, setting a minimum of 0 and maximum of 10 for the interval. To obtain the resistance for each strain in each drug (, is divided by growth in the DMSO control (:

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

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

The multiplicative model of genetic interactions60 was applied to the metric. In this model, the expected resistance of a double knockout strain in a given drug () is the product of the resistances of the corresponding single knockout strains:

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

We defined a two-gene interaction term as the log-ratio of the observed fitness to the fitness expected by single-gene effects, rather than the traditional linear difference from a multiplicative estimate.



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

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

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

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

To train this model, is encoded as a set of 16 binary variables, where 0 represents a wildtype and 1 represents a knockout at a given gene. Therefore, to predict phenotype from , the relevant coefficients are added only if the corresponding gene is knocked out, and the coefficients are added only if all the genes in subset are knocked out. For each drug, we fit this model using the glm() function in R, with terms to a chosen level of complexity.

To perform the marginal association in Fig. SXX, we fit a model with only terms, and performed stepwise feature elimination (eliminating the gene with the highest p-value at each step) until all included terms had a significance level of . Linear model term significance was tested using the Type III Sums of Squares ANOVA implementation given in the car package in R.

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

**Defining a Non-Linear System Model**

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

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

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

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

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

We then allow to capture compensatory activation. That is, can negatively affected by additive inhibitory input

Where are the inhibitory inputs from other ABC transporters. While the form of may depend on the inhibition mechanism, here we also modelled it as having a sigmoidal form for simplicity:

Because the coefficients capture potential inhibition in this model, they are constrained to be negative.

**Learning a Non-Linear System Model as a Neural Network**

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

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

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

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

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

This model was compiled with the mean-squared error (‘mse’) loss function, using the adam optimizer with a learning rate of 0.1. Training was performed for 1,000 epochs, using a batch size of 1,000 and 10% split between training and validation (validation\_split = 0.1). Model initialization and training was repeated 10 times, and the weights to the final model were set to the median weights learned from these 10 iterations.

We searched for an appropriate regularization rate by performing the above training and averaging procedure using a range of rates from to and observing the mean-squared error of the resulting predictions and the number of model parameters. We chose a rate of , as rates considerably higher than this appeared to negatively impact performance in the MAT**a** and MAT**α** pools (Fig. SXX), while considerably lower rates sharply increased the number of non-zero parameters (Fig. SXX).

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

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

We then compute the Mann-Whitney U statistic between and to derive a p-value for , and keep all features with in the final 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 was subject to sporulation and MAT**a** haploid selection22. 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)61. Colonies of transformed strains were grown in solid medium for 5 days in SD –W, SD –Trp–Ade–His (SD –WAH), SD –WAH +25μM fluconazole + 2% DMSO, SD –WAH +50μM fluconazole + 2% DMSO, and SD –WAH + 2% DMSO.

**PCA Testing of Protein-Protein Interactions**

*PDR5*, *YOR1*, and *SNQ2* MAT**a** (mDHFR-F[1,2]-NatMX fusions) and MAT**α** (mDHFR-F[3]-HphMX fusions) PCA strains were obtained from a previous genome-wide screen48. Additional strains acting required to recreated positive and negative controls were also obtained from this screen (Fig. S11). Strains were individually mated and diploids were selected on solid YPD supplemented with Hygromycin B and Nourseothricin (YPD +Hyg +Nat). Diploid strains were spotted on solid YPD +Hyg +Nat supplemented with either 2% DMSO, 2% DMSO + 200 μg/mL methotrexate, or 2% DMSO + 200 μg/mL methotrexate + 46.8μM fluconazole. Strains were grown for 72 hours at 30°C.

**Quantitative RT-PCR**

RNA was extracted from cultures growing exponentially in 23.43μM fluconazole using the QIAGEN RNeasy® kit. 1μg of isolate was treated with DNAse and analyzed using an Agilent Bioanalyzer to quantify nucleic acid concentration and verify purity. cDNA synthesis was performed using a combination of oligo-DT and random hexamer primers. qPCR on these samples was then performed using a SensiFAST™ Real-Time PCR Kit and Ct values were quantified using a CFX machine. cDNA synthesis and qPCR was performed for *PDR5* and *UBC6* (acting loading control).

**Availability of Data and Materials**

**Competing Interests**

The authors declare that they have no competing interests.

**Acknowledgements**

This work was supported by XX.

**Author Contributions**

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

**Additional Data Files**

**Additional Data S1.** List of primers used in this study. Includes the primers used to construct the barcoder strain, perform genotyping, RCP-PCR overhangs, and pool multiplexing primers.

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

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

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

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

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

**Additional Data S7.** Previously-known drug knockout 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. Shen, J. P. & Ideker, T. Synthetic Lethal Networks for Precision Oncology: Promises and Pitfalls. *J. Mol. Biol.* **430,** 2900–2912 (2018).

7. Horlbeck, M. A. *et al.* Mapping the Genetic Landscape of Human Cells. *Cell* **174,** 953–967.e22 (2018).

8. Giaever, G. *et al.* Functional profiling of the Saccharomyces cerevisiae genome. *Nature* **418,** 387–391 (2002).

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

10. 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).

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

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

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

14. 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).

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

16. 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).

17. Zhang, Y. *et al.* A transportome-scale amiRNA-based screen identifies redundant roles of Arabidopsis ABCB6 and ABCB20 in auxin transport. *Nat. Commun.* **9,** 4204 (2018).

18. Palmer, A. C. *et al.* Delayed commitment to evolutionary fate in antibiotic resistance fitness landscapes. *Nat. Commun.* **6,** 7385 (2015).

19. Cancer Genome Atlas Research Network *et al.* Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. *N. Engl. J. Med.* **368,** 2059–2074 (2013).

20. Heckl, D. *et al.* Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing. *Nat. Biotechnol.* **32,** 941–6 (2014).

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

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

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

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

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

26. 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).

27. Wong, A. S. L. *et al.* Multiplexed barcoded CRISPR-Cas9 screening enabled by CombiGEM. *Proc. Natl. Acad. Sci. U. S. A.* **113,** 2544–9 (2016).

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

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

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

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

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

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

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

35. Huls, M. *et al.* The breast cancer resistance protein transporter ABCG2 is expressed in the human kidney proximal tubule apical membrane. *Kidney Int.* **73,** 220–225 (2008).

36. 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).

37. 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).

38. Lee, A. Y. *et al.* Mapping the cellular response to small molecules using chemogenomic fitness signatures. *Science* **344,** 208–11 (2014).

39. 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).

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

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

42. Smith, A. M. *et al.* Highly-multiplexed barcode sequencing: an efficient method for parallel analysis of pooled samples. *Nucleic Acids Res.* **38,** e142 (2010).

43. KOLACZKOWSKI, M., KOLACZKOWSKA, A., LUCZYNSKI, J., WITEK, S. & GOFFEAU, A. *In Vivo* Characterization of the Drug Resistance Profile of the Major ABC Transporters and Other Components of the Yeast Pleiotropic Drug Resistance Network. *Microb. Drug Resist.* **4,** 143–158 (1998).

44. Shekhar-Guturja, T. *et al.* Beauvericin Potentiates Azole Activity via Inhibition of Multidrug Efflux, Blocks *C. albicans* Morphogenesis, and is Effluxed via Yor1 and Circuitry Controlled by Zcf29. *Antimicrob. Agents Chemother.* **60,** AAC.01959-16 (2016).

45. Katzmann, D. J., Burnett, P. E., Golin, J., Mahé, Y. & Moye-Rowley, W. S. Transcriptional control of the yeast PDR5 gene by the PDR3 gene product. *Mol. Cell. Biol.* **14,** 4653–61 (1994).

46. Ernst, R. *et al.* A mutation of the H-loop selectively affects rhodamine transport by the yeast multidrug ABC transporter Pdr5. *Proc. Natl. Acad. Sci.* **105,** 5069–5074 (2008).

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

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

49. 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).

50. Ferreira-Pereira, A. *et al.* Three-dimensional reconstruction of the Saccharomyces cerevisiae multidrug resistance protein Pdr5p. *J. Biol. Chem.* **278,** 11995–9 (2003).

51. Newman, J. R. S. *et al.* Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* **441,** 840–846 (2006).

52. Shaw, W. M. *et al.* Engineering a model cell for rational tuning of GPCR signaling. *bioRxiv* 390559 (2018). doi:10.1101/390559

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

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

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

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

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

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

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

60. 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).

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

**Figures**

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

We created a barcoded wild-type pool (Fig. S1) to enable construction of an engineered population using any multi-mutant strain. In this study, this pool was mated with a 16 ABC transporter knockout strain (ABC-16). The genotype at 16 ABC transporters is indicated by the squares drawn in each cell (black = knockout, white = wild type). Diploids from this cross were subject to sporulation and barcoded haploids were then selected. Each haploid inherits either a wildtype or knockout allele at these 16 loci. Single colonies were picked and arrayed onto a series of 384-well plates. *En masse* genotyping was performed on this collection using an RCP-PCR41 strategy, which uses a combination of row and column tags to allow identification of PCR products arising from the same well in each plate. An additional PCR reaction adds a plate tag (not shown). High throughput sequencing of pooled RCP-PCR products allows large scale genotyping and identification of a strain-specific DNA barcode for many strains. Strains with a sucessfully determined barcode and genotype are transferred into two liquid pools based on mating type (MAT**a** or MAT**α**), and grown under each of 16 drugs. High throughput sequencing of strain-specific DNA barcodes at t = 0, 5, 10, 15, and 20 generations of growth in each drug reconstructs a drug by strain resistance profile, profiling the engineered population.

**Figure 2.** An exploration and assessment of multi-knockout fitness landscapes within a 6-gene group.

**A** Comparison of MAT**a** and MAT**α** group resistance profiles in camptothecin and tamoxifen. Individuals were grouped by their genotype at *pdr5∆, snq2∆, ybt1∆, bpt1∆, ycf1∆,* and *yor1∆*. The 6-locus genotype of each group is indicated by the legend. Individuals in each group vary at the remaining 10 loci. Each point represents the average log2(resistance) of the indicated group in the MAT**a** pools (x-axis) and MAT**α** pools (y-axis). Profiles for the remaining drugs are shown in Fig. S6.

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

**C** A radial landscape of benomyl resistance. The graph is centered by the 6-gene wildtype group, with outward extensions adding cumulative knockouts. Each section is coloured by the average resistance of the corresponding 6-gene group relative to the 6-gene wild type. Extensions to 1, 2, and 6 total knockouts are shown. Sections are coloured by the mean resistance of each group relative to the 6-gene wildtype. The colour scale extends equally in both directions by the largest observed difference in resistance between the 6-gene wildtype and any other group (blue for increased resistance, orange for decreased resistance). All paths where any additional knockouts do not result in significant changes (Bonferroni-adjusted *p* < 0.05, Mann-Whitney U test) are pruned.

**D** As in B, showing radial fitness landscapes for 10 additional drugs. The remaining 5 drugs are shown in Fig. S7.

**Figure 3.**  Exploration and formalization of surprising multi-gene knockout phenotypes.

**A** A linear landscape of resistance to mitoxantrone, benomyl, and fluconazole in the MAT**a** pools amongst 6-gene groups. The 6-gene genotype of each group is indicated by the legend. Groups are arranged on the x-axis by the number of knockouts (with jitter added to improve clarity), and the y-axis by average drug resistance. Groups separated by single knockouts are connected by lines. Solid lines indicate significant differences in resistance (Bonferroni-adjusted *p* < 0.05, Mann-Whitney U test), otherwise dashed lines are used. Linear landscapes for all pools are drawn in Fig. S8.

**B** Distribution of fluconazole resistance amongst all *ybt1∆,* *yor1∆*, *snq2∆, ycf1∆,* and *pdr5∆* knockout groups in the MAT**a** pool. Group genotype is indicated for each line using the same legend as in A). The resistance distribution as a kernel density function is drawn.The distribution of *pdr5∆* groups (magenta) are paired with their corresponding *PDR5+* equivalent (grey).

**C** A linear model was used to formally determine significant gene knockout and genetic interaction effects mediating resistance to the tested drugs (see Methods). Linear model terms which were significant (Bonferroni adjusted *p* < 0.05) in both MAT**a** and MAT**α** pools of their given drug are coloured according the legend on the left. Other terms are coloured in grey. ε terms represent n-way interactions (see Methods). Coefficents are sorted by term complexity. Term complexity is also indicated by the grey colour scale to the immediate left of the heatmap.

**Figure 4.** Interpreting a complex genetic landscape in fluconazole.

**A** A model of drug-dependent transporter-transporter relationships interpreted from the engineered population profiles. Antagonism indicates genetic evidence that the presence of one transporter negatively affects the ability of another to impart drug resistance. Parallelism indicates genetic evidence that multiple transporters are redundantly responsible for resistance to a drug. Data S8 lists the genetic evidence for each relationship.

**B** The TWAS21230902 MAT**α** strain (*pdr10∆ pdr18∆ pdr5∆ snq2∆ ybt1∆ ycf1∆ yor1∆;* Data S2) was mated with RY0146 (MAT**a**), and subject to sporulation and MAT**a** haploid selection. Individuals from this cross were genotyped, and individual strains containing each of 32 knockout combinations at *pdr5∆*, *snq2∆*, *ybt1∆*, *ycf1∆*, and *yor1∆* were identified and isolated.

**C** The resistance of 32 single strains to various concentrations of fluconazole was measured as growth compared to the solvent control (DMSO). Resistance was measured at the concentrations labelled on the x-axis, and linearly interpolated in between. Resistance of individual strains at concentrations lower than 15.6μM are shown in Fig. S10.

**D** Comparing the IC50 of fluconazole derived from single-strain growth experiments to the resistance expected by in the grouped pool data (MAT**a** and MAT**α** log2-resistance were averaged). Strain genotype is indicated by the legend.

**E** Measuring the mRNA expression of *PDR5* in wild-type (RY0566), *snq2∆yor1∆*, and *snq2∆yor1∆ybt1∆ycf1∆* strains. *PDR5* mRNA expression was measured using qRT-PCR and normalized relative to *UBC6*. Values represent the ratio of *PDR5* expression compared to the average in RY0566. Error bars indicate standard deviation. Three replicates were used in each experiment. P-values were calculated using a t-test.

**F** Comparing the modeled *PDR5* repression by *YOR1* and *SNQ2* with with protein-protein interactions found using MYTH and PCA. Interactions were measured in both this study (Fig. S11, S12) and previous studies32,48. Spotting assays show a re-testing of Pdr5-Snq2, Pdr5-Pdr5, and Pdr5-Yor1 interactions using MYTH. NubG-PDR5, NubI-PDR5, Ost1-NubG, and Ost1-NubI strains were each transformed with plasmids containing clones of *PDR5*, *YOR1*, *SNQ2*, or an artificial bait fused to Cub (YOR1-L2, PDR5-L2, SNQ2-L2, Artificial L2 bait). NubI fusions are expected to spontaneously reconstitute ubiquitin with Cub, while NubG fusions are expected to require a protein-protein interaction for reconstitution. Ost1 is a component of the oligosaccharyltransferase complex localized to the endoplasmic reticulum membrane and is not expected to interact with any baits tested. Colonies of transformed strains were spotted on SD –Trp (SD –W) and SD –Trp–Ade–His (SD –WAH). SD –WAH selects for reconstitution of ubiquitin. Spotting assays were also performed in the presence of DMSO and fluconazole (Fig. S12).

**Figure S1.** Creation of a parent barcoder pool.

**A** Engineering of a barcoder pool cassette. An HphMX4 cassette was amplified from pIS420, with overhangs adding the US2 and DS1 sites. A second PCR reaction was performed to add 25 random base pairs for use as UP and DN tags, as well as two constant US1 and DS2 regions. A third PCR reaction then adds LoxP/Lox2272 sites, and homology to the pSH47 SacI site.

**B** Transforming a pool of barcoder parents. RY0148 was modified to add a LoxP-URA3-Lox2272 site and was co-transformed with the barcoder pool cassette and SacI-digested pSH47 to enable reconstitution of a pSH47-based barcoder plasmid construct through in-yeast assembly. Transformants were selected by growth in YPG +Hyg for 3 days to allow for both selection of successful in-yeast assembly products, as well as induction of Cre to enable recombination and replacement of URA3 with the barcoder pool cassette. Loss of URA3 through Cre-enabled recombination is selected by subsequent growth in 5-FOA.

**Figure S2.** Analysis of pool genotyping quality.

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

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

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

**D** Distribution of DNA content in the MAT**a** and MAT**α** pools compared to a diploid and haploid reference. DNA content was measured using flow cytometry.

**Figure S3.** 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.

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

**C** Distribution of UPtag-DNtag correlation in the pools. The minimum correlation in the MAT**a** and MAT**α** a pool for each drug is taken.

**D** Comparing UP and DN tag correlation with pool variability in resistance (as ). is taken as the minimum from from UP and DN tag estimates.

**Figure S4.** Reproducible marginal gene knockout resistance effects in the pool.

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

**Figure S5.** Profiles of grouped genotype data.

Strains were grouped by knockout genotypes at *pdr5∆, snq2∆, ybt1∆, ycf1∆,* and *yor1∆*. Each section in the heatmap represents the average log2-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 S6.** 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 S7.** A radial fitness landscape in six additional drugs.

A radial fitness landscape in six drugs showing all multi-knockout paths. Each graph is centered by the 6-gene wildtype group, with outward extensions adding cumulative knockouts. Each section is coloured by the average resistance of its corresponding knockout group relative to the 6-gene wild type. Extensions to 1, 2, and 6 total knockouts are shown. Sections are coloured by the mean resistance of each group relative to the 6-gene wildtype. The colour scale extends equally in both directions by the largest observed difference in log2-resistance between the 6-gene wildtype and any other group (blue for increased resistance, orange for decreased resistance). All paths where any additional knockouts do not result in significant changes (Bonferroni-adjusted *p* < 0.05, Mann-Whitney U test) are pruned.

**Figure S8.** A linear landscape of resistance to 16 drugs.

**A** A linear landscape of resistance to all tested drugs in the MAT**a** pools amongst 6-gene groups. The 6-gene genotype of each group is indicated by the legend. Groups are arranged on the x-axis by the number of knockouts (with jitter added to improve clarity), and the y-axis by average drug resistance. Groups separated by single knockouts are connected by lines. Solid lines indicate significant differences in resistance (Bonferroni-adjusted *p* < 0.05, Mann-Whitney U test), otherwise dashed lines are used.

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

**Figure S9.** Reproducibility of linear models.

For each drug, a linear model was trained to identify significant terms mediating drug resistance (see Methods). Linear model predictions of log2-resistance are indicated on the y-axis and values observed in the pool are indicated on the x-axis. Each plot is labelled by the pool mating type which was used to train the linear model on the y-axis and to assess its predictions on the x-axis.

**Figure S10.** Comparing drug resistance measured from single-strain experiments to the grouped pool data.

Resistance of individual strains containing each of 32 knockout combinations at *pdr5∆*, *snq2∆*, *ybt1∆*, *ycf1∆*, and *yor1∆* was measured and compared to the resistance to the pool data. Pool strains were grouped based on genotype at these 5 loci, median log2-resistance was determined for each group in MAT**a** and MAT**α** pools, and these values were averaged to obtain a single pool value. Strain genotype is indicated by the legend. Growth of individual strains was measured at 1.9, 3.9, 7.8, 15.6, 23.4, 31.2, 35, and 40μm of fluconazole.

**Figure S11.** Measuring all protein-protein interactions between Pdr5, Snq2, and Yor1 using mDHFR PCA.

*PDR5*, *YOR1*, and *SNQ2* MAT**a** (mDHFR-F[1,2]-NatMX fusions) and MAT**α** (mDHFR-F[3]-HphMX fusions) PCA strains were obtained from a previous genome-wide screen48. Strains were individually mated to obtain the indicated diploids. Diploid strains were spotted on YPD containing either DMSO, DMSO + methotrexate (MTX), or DMSO + MTX + 46.8μM fluconazole. MTX selects for successful reconstruction of mDHFR from the F[1,2] and F[3] fragments via a protein-protein interaction. Link-F[1,2]/ Link-F[3] is a diploid strain which tests against interaction of the universal linker regions when fused to the mDHFR fragments. Zip-F[1,2]/ Zip-F[3] is a diploid strain which tests for interaction between two leucine Zipper sequences fused to the mDHFR fragments. Strains were grown for 3 days at 30°C.

**Figure S12.** Measuring protein-protein interactions of Pdr5 with Snq2 and Yor1 using MYTH.

NubG-PDR5, NubI-PDR5, Ost1-NubG, and Ost1-NubI strains were retrieved from a previously constructed genomic prey library25 and were each transformed with plasmids containing clones of *PDR5*, *YOR1*, *SNQ2*, or an artificial bait fused to Cub (YOR1-L2, PDR5-L2, SNQ2-L2, Artificial L2 bait). NubI fusions are expected to spontaneously reconstitute ubiquitin with Cub, while NubG fusions are expected to require a protein-protein interaction for reconstitution. Ost1 is a component of the oligosaccharyltransferase complex localized to the endoplasmic reticulum membrane and is not expected to interact with any baits tested. Colonies of transformed strains were spotted on SD –Trp (SD –W), SD –Trp–Ade–His (SD –WAH), SD –WAH +25μM fluconazole + 2% DMSO, SD –WAH +50μM fluconazole + 2% DMSO, and SD –WAH + 2% DMSO. SD –WAH conditions select for reconstitution of ubiquitin.