**Enabling in-depth characterization of complex multi-knockout traits using engineered population profiling**

Albi Celaj1,2,3, Louai Musa2, Marinella Gebbia1, Minjeong Ko2,6, Gireesh Seesankar5, Shijie Zhou1, Atina Cote2, Tiffany Fong5, and Nozomu Yachie4,7, Frederick Roth1,2,3,6,7

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

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

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

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

5McMaster University, Hamilton, Ontario, Canada.

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

7Lead contact

**Lead Contact Information:**

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

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

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

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

**Abstract**

Many biological traits are controlled by an underlying system of multiple genes. Key to the study of such systems is understanding the effects that arise when these genes are perturbed in combination. However, current approaches make the systematic investigation of combinatorial gene knockout effects impractical. Here we develop an ‘engineered population profiling’ strategy to study arbitrarily complex knockout phenotypes and genetic relationships within a targetted gene set, and use this strategy to map multi-knockout phenotypes amongst 16 yeast ATP Binding Cassette (ABC) transporters under the effects of 16 drugs. We engineer and genotype a population of ~7,000 yeast strains containing random knockout combinations, and perform barcode-indexed growth profiling. Numerous surprising multi-knockout phenotypes were revealed, leading to a map of context-dependent transporter-transporter relationships, and new roles for *YBT1* and *YCF1*. Amongst the findings, a quadruple deletion of *SNQ2*, *YBT1*, *YCF1*, and *YOR1* confered resistance to the *PDR5* substrate fluconazole, and further investigation pointed to a new model of complex mutual inhibition amongst these ABC transporters. Thus we illustrate the potential for an engineered population profiling strategy to inform understanding of complex traits and gene function.

**Introduction**

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

A challenge to systematically studying multi-gene variant or knockout effects are the vast number of potential combinations. While single-knockout effects have been widely profiled in many model organisms, a genome-wide profile of genetic interactions in yeast required the development of tools to individually engineer and profile more than 23 million strains containing two-gene knockout combinations5. Large two-gene knockout profiles like these have been used to accurately cluster genes into functional groups5,6, to better predict the genetic basis of drug resistance7, and to inform order-of-action in biological pathways8. Despite the scale of such efforts, approximately 1,000 yeast genes remain with no single-knockout growth phenotypes and with no two-gene interactions under standard growth conditions5. While it is expected that many of these genes would reveal a phenotype in a systematic three-gene knockout screen, such a screen would require the engineering and phenotyping of more strains than is currently feasible.

While multi-knockout phenotypes have been challenging to explore using existing approaches, they are likely to affect many biological traits. A large survey of triple knockout phenotypes in yeast predicts that three-gene interactions would greatly outnumber all discovered two-gene interactions if mapped genome-wide9. Consistent with this, a survey of individual examples shows that they can arise from many different underlying mechanisms, and may thus be prevalent within many biological systems10. Extending beyond three knockouts, individual examples of complex genetic interactions involving 5+, 7, and 20+ simultaneous mutations have been reported in this model organism11–13, but their general prevalence is unknown. A formal approach to capture complex genetic interactions in a collection of published multi-gene perturbation experiments shows that they can have large effects which are comparable to or greater than those seen with single gene knockouts14. Despite their prevalence and potentially large effects, however, general understanding of mechanisms leading to complex genetic interactions and their contribution to genotype-to-phenotype relationships remains far behind the analgous understanding of two-gene knockout effects.

To surmount the combinatorial complexity required to study complex multi-knockout phenotypes, methods are needed to efficiently engineer and genotype the required strains. Many molecular tools used to study two-gene knockouts were developed for the engineering of many strains with a few deletions, and are not easily adaptable for the engineering of many multi-deletion strains due to the limited availability of usable molecular markers15. To make modifications at multiple loci simultaneously, several multiplex automated genome engineering (MAGE) methods have been designed16,17. However, such methods often aim to isolate a handful of engineered strains exhibiting a phenotype of interest, and are not designed to genotype the large number of strains required to create an in-depth genotype to phenotype map. To extend MAGE for the creation of extensive genotype to phenotype maps, methods are under development in *E. coli* to allow the genotyping of many engineered strains at multiple loci18,19. In principle, MAGE-like methods can serve as the basis for creating the genetic variation required to understand phenotypes resulting from multiple genetic modifications, and are likely to do so in the near future.

With the aim of profiling arbitrarily complex genetic relationships within a gene family, we had previously developed a method for knocking out multi-gene redundancies, and had deleted 16 ATP Binding Cassette (ABC) transporters in a single yeast strain (ABC-16)15. Using this strain, we demonstrate an ‘engineered population profiling’ strategy to study arbitrarily complex drug resistance knockout phenotypes within the 65,536 possible knockout combinations within this targetted set of genes. To this end, we first back-crossed the ABC-16 strain to a pool of isogenic wild-type parents, each containing a unique DNA barcode20. We then genotyped 6,709 progeny from this cross *en masse*, linked the genotype to a strain-specific barcode, and performed genotype-indexed growth profiling using high-throughput barcode sequencing under 16 drugs20. We use these profiles to uncover numerous surprising and reproducible multi-knockout relationships, leading to a map of many novel drug-dependent transporter-transporter relationships, and many new roles for *YBT1* and *YCF1*. Further exploration of a multi-knockout trait pointed to a new model of complex mutual inhibition amongst ABC transporters, and demonstrates the potential of targetted profiling strategies to map phenotypes controlled by multiple genes and inform understanding of gene function.

**Results**

**Creation and Genotyping of an Engineered Population**

In order to engineer a population of strains containing random multiple knockouts within a targetted set of genes, we made use of our previously-created ABC-16 strain15. The ABC-16 strain was mated to a pool of isogenic wild-type strains containing two unique DNA ‘barcode’ tags (UP and DN) to allow molecular identification of each individual in the population. After sporulation and barcoded haploid selection, automated colony picking was used to isolate approximately equal numbers of MATa and MATα progeny into 384-well plates, creating a collection of random knockout strains. To genotype these progeny *en masse* and link this information to their DNA barcode, we adapted our previously-developed row-column-place PCR (RCP-PCR) protocol21 to amplify and sequence the strain-specific barcodes as well as barcodes identifying the deletion of specific genes (deletion-specific barcodes). Here, the RCP-PCR products were tagged with DNA sequences indicating the plate as well as the row and column of origin, allowing high-throughput sequencing to identify which strain-specfic and deletion-specific barcodes came from the same well in the collection (Methods; Fig. 1). To calibrate the genotyping results obtained by high-throughput sequencing analysis, multiple replicates of 40 individually genotyped ‘gold standard’ strains, as well as two additional control strains were added to the collection at known positions (Methods; Data\_S\_Genotypes). Using these strains, we estimate a genotyping accuracy of 95% or higher for 12 of 16 genes, and an overall accuracy of 93.2% (Fig. S1A, Methods). An independent method based on the distribution of knockouts in the genotyping data estimated a similarly high overall accuracy of 93.8% (Fig. S1C, Methods). We obtained usable genotyping data linked to at least one unique DNA barcode for 6,709 progeny with 5,095 unique genotypes, representing approximately 8% of the 65,536 possible knockout combinations within this gene set.

Given the use of a cross-based strategy, the genotyping data was used to verify the lack of strong linkage between ABC transporters. All genes were either unlinked or weakly linked except for *BPT1* and *YBT1* (Fig. S1B; r = 0.49). Surprisingly, three pairs of genes – *YOR1-YCF1*, *YOR1-BPT1*, and *SNQ2-PDR5* – had weak (-0.04 ≥ r ≥ -0.08) but significant apparent negative linkage (Fig. S1B), which reflects the ability of this large population to detect very small synthetic effects which may have affected the sporulation, haploid selection, or automated colony picking steps.

**Phenotyping An Engineered Population**

After validating the genotyping results and determining a unique barcode for many strains, we aimed to profile resistance to a collection of drugs. The arrayed collection was transferred into two liquid pools separated by mating type (MAT**a** and MAT**α**), which were treated as independent populations in subsequent analyses. These pools were then grown in the presence of 16 different anticancer and antifungal drugs (Data\_S\_Drug\_Concentration) as well as a solvent control. Using high-throughput strain barcode sequencing20, strain frequency was estimated at five time points (0, 5, 10, 15, 20 generations of overall pool growth, Fig. 1). Sequencing data from all time points were used to determine an overall growth metric for each strain, and resistance was determined by comparing growth in each drug to the solvent control (Methods, Fig. 1). After excluding individuals with missing barcodes, resistance scores were determined for 3,221 MATa and 3,592 MATα strains. To determine barcode reproducibility, we measured correlation between the resistance scores generated by the UP and DN barcode tags. For strains with two successfully mapped unique DNA barcodes, the resistance scores for 8 drugs were highly correlated (r > 0.9 for both MATa and MATα populations), and moderately correlated (minimum r > 0.7) for 4 other drugs (Fig S2). For the remaining 4 drugs, the relatively low correlation appears to result from a lack of strain to strain variability, suggesting that the 16 ABC transporters tested did not play a major role in mediating resistance to these compounds (Fig S2).

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

We aimed to identify relevant genes in each drug for further exploration, as only a subset of the 16 ABC transporters are expected to mediate resistance to any given compound even when taking into account contextual gene-knockout effects. A linear model was used to perform a marginal association of each knockout to drug resistance. This approach identifies relevant genes by testing whether their knockout is expected to have an effect in the population when averaged over all genetic backgrounds. Based prior literature, we were able to verify XX% of previously-known single-knockout resistance phenotypes in both of our pools (Fig. S3; Data XX). While prior literature had been largely focused on the effects of *snq2∆*, *pdr5∆*,and *yor1∆*, we found many unreported associations with *ycf1∆*, *bpt1∆*, and *ybt1∆* to be common amongst the drugs tested. Notably, the latter three genes had comparably small marginal effects, indicating that either their individual knockouts have minor effects, or that their effects depend on a specific knockout background. Of the significant marginal associations, 62% were reproducible between the MAT**a** and MAT**α** pools (Fig. S3). To verify the generalizability of the marginal linear models fitted, we found that training on one mating type did not diminish the ability to predict knockout phenotypes of the opposite mating type (Fig. S4).

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

After identifying gene knockouts of interest within the drugs tested, we used the population data to explore all combinations of multiple knockout effects amongst *snq2∆*, *pdr5∆*, *yor1∆*, *ycf1∆*, *bpt1∆*, and *ybt1∆*. Strains in the pool were first grouped based on all 64knockout combinations amongst these transporters (allowing variation at the other 10 transporters within each group), and a grouped resistance profile was created and used to cluster drugs by profile similarity (Fig. S5).

We expanded this grouped resistance profile into a ‘complex fitness landscape’ to determine the effects of knocking out a transporter in the context of any other knockout combination within the grouped genes. This fitness landscape representation allowed us to use the population data to find many apparently-surprising and reproducible genetic relationships in different drugs (Fig. 2A, S6). In the most straightforward cases, multiple knockouts appeared to confer surprising sensitivity which was not as apparent when knocking out a smaller set of the underlying genes. Such effects are illustrated, for example, by the *snq2∆pdr5∆* group under camptothecin (Fig. S6), and by the *snq2∆pdr5∆ybt1∆* and *snq2∆pdr5∆yor1∆* groups under mitoxantrone (Fig.2A). Broadly, such complex sensitivity patterns suggest that multiple transporters are able to efflux a given drug redundantly (‘parallel transport’).

In less intuitive fitness landscapes, knockouts appeared to have opposite drug resistance effects in different genetic backgrounds. Under benomyl, for example, the effects of *yor1∆* were highly context dependent - *yor1∆* was shown to have no significant effect in the presence of the other five genes, to increase fitness relative to *pdr5∆* as *yor1∆pdr5∆*, and to decrease fitness relative to *snq2∆* as *yor1∆snq2∆* (Fig. 2A). The *yor1∆pdr5∆* phenotype may have been expected given a dependence of benomyl resistance to SNQ2 and an earlier report of a (non-additive) ‘compensatory activation’ of Snq2 given disruption of PDR5 and YOR122. However,the *yor1∆snq2∆* phenotype is novel and suggests that Yor1 can also serve a secondary role in benomyl resistance, for example by providing some efflux in the absence of Snq2.

In perhaps the most surprising examples, complex knockout combinations appeared to confer considerable drug resistance rather than sensitivity. Under fluconazole and ketoconazole, a quadruple deletion of *snq2∆ybt1∆ycf1∆yor1∆* led to drug resistance, whereas *pdr5∆snq2∆ybt1∆ycf1∆yor1∆* was comparably as sensitive as the *pdr5∆* group (Fig. 2A, S6)*.* Furthermore, combinations of one or two knockouts within these four genes appeared to have only minor effects (Fig. 2A, S6). While these results extend a previously-reported ‘compensatory activation’ phenomenon22, we were surprised by the requirement of multiple knockouts, and the involvement of YCF1 and YBT1.

To further develop ways to explore such multi-knockout population data, we created a ‘radial fitness landscape’. This representation maintains the knockout relationships present in a traditional landscape, but in a top-down view which allows both a broad overview, as well as an explicit visualization of all possible paths towards a 6-knockout state (Fig. 2B-C, S7). This landscape allows quick determination of the relevant transporters in mediating resistance to any drug, and identification of drugs with with similar resistance paths. In many drugs, we were surprised by the prevalance of paths which were apparently more resistant than the wild type state (Fig. 2C).

Because each group represents a population of cells with random background knockouts, we further visualized the underlying distribution of drug resistance (Fig. 2D). This revealed, for example, that variance in fluconazole resistance within the *pdr5∆* groups are much less than groups containing PDR5. The lack of variability in the *pdr5∆* groups may be explained by the unconditionally high fluconazole sensitivity of these strains, whereas the variability in the PDR5 groups may reflect still-unidentified complex ABC transporter knockout effects in the respective backgrounds.

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

Observing many apparently-surprising multi-knockout drug resistance phenotypes, we aimed to formally capture the complex genetic interactions present within the engineered population. To do this, we used a linear model which extends the multiplicative model of combined genetic effects23 to incorporate interaction terms of arbitrary complexity (see Methods). This is similar to the approach used when studying controlled crosses of natural populations7, with a much more straightforward link from association to causality. We treated the search for complex genetic interactions as a feature selection problem, searching for up to to 4-gene interactions within a total space of 2,516 coefficients (See Methods; Fig. 3A). We found 23 2-way, 12 3-way, and 8 4-way interactions which were reproducible between MAT**a** and MAT**α** pools at a stringent Bonferroni corrected p-value cutoff of 0.05 (Data\_S\_lm\_result\_summary).

This formal approach captured the surprising genetic relationships found by observing the fitness landscapes. For example, *yor1∆* was found to have no main effect under benomyl, to have a positive genetic interaction with *pdr5∆*, and to have a negative genetic interaction with *snq2∆* (Fig 3A, Data\_S\_lm\_result\_summary). In camptothecin, *pdr5∆* and *snq2∆* each had a minor main effect, as well as a strong negative interaction between them (Fig 3A, Data\_S\_lm\_result\_summary). Similarly, the surprising *pdr5∆snq2∆ybt1∆* phenotype in mitoxantrone was modelled as a combination of a small marginal effect of *snq2∆*, a negative interaction between *pdr5∆* and *snq2∆*, and a further triple negative interaction upon *YBT1* deletion as *pdr5∆snq2∆ybt1∆* (Fig 3A, Data Data\_S\_lm\_result\_summary). This pattern of genetic interactions suggests not only that these three genes efflux mitoxantrone in parallel, but that Snq2 has the highest efflux capability, followed by Pdr5, and finally Ybt1. A similar ‘parallel transport’ phenotypic pattern was observed for *pdr5∆snq2∆yor1∆* in cisplatin (Fig 3A, Data\_S\_lm\_result\_summary). Under fluconazole, the resistance phenotype was modelled (in addition to one- and two- gene effects involving *pdr5∆*, *yor1∆*, *ybt1∆*, *pdr5∆yor1∆*) as the combination of three positive three-way interactions (*ybt1∆ycf1∆yor1∆*, *snq2∆ybt1∆yor1∆*, *snq2∆ycf1∆yor1∆*). The *PDR5* dependence of these effect were then further modelled as three additional negative four-way interactions(*pdr5∆ybt1∆ycf1∆yor1∆*, *pdr5∆snq2∆ybt1∆yor1∆*, *pdr5∆snq2∆ycf1∆yor1)*.

**Drug-Dependent Transporter-Transporter Relationships from Multi-Knockout Phenotypes**

We sought to intepret the drug-dependent transporter-transporter relationships suggested by the multi-knockout data, using a combination of the grouped fitness landscapes and formally-determined genetic interactions. The interpreted relationships were highly drug-dependent, and were derived from either two, three, or four knockout phenotypes (Fig 3B, Data\_S\_GI\_interpretation). All interpreted relationships involved either parallel clearance of a drug by two or more transporters, and/or the antagonism of an ABC transporter by the presence of another (Fig 3B). The relationships suggest an extensive role for YBT1 and YCF1 in many drugs – YBT1 is hypothesized to work in parallel with SNQ2, PDR5, or YOR1 in the context of mitoxantrone, itraconazole, and methotrexate resistance, and to antagonize *PDR5* in the context of fluconazole resistance (Fig 3B). *YCF1* is modelled to antagonize *SNQ2* in bisantrene resistance, to antagonize *PDR5* in fluconazole and ketoconazole, and to work in parallel with *SNQ2* in the context of benomyl resistance (Fig 3B).

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

In order to validate the complex genetic interactions present in the population, we used targetted crosses to generate individual strains containing all knockout combinations of *pdr5∆, snq2∆, yor1∆, ybt1∆,* and *ycf1∆* (Methods). With this population, we tested the multi-knockout resistance patterns for XX drugs under non-competitive liquid growth environments. It was found that XX. Fluconazole resistance of these individual strains correlated well with the pooled resistance (r=XX, Fig XX). As predicted by the pool data, the *snq2∆yor1∆ybt1∆ycf1∆* strain was the fastest growing, and only minor resistance effects were seen for strains with one and two knockout combinations amongst these four genes.

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

**Pdr5 is inhibited by XX through XX**

Given the genetic evidence of repression of Pdr5 by the presence of *SNQ2*, *YOR1*, *YBT1*, and *YCF1*, we sought to understand the underlying mechanisms. Inhibition of Pdr5p by *SNQ2* and *YOR1* has been previously reported, and thought to be mediated by transcriptional activation by Pdr122, although the mechanism underlying this compensatory activation is unknown. To investigate whether *PDR5* mRNA abundance was elevated in the *snq2∆yor1∆ybt1∆ycf1∆* compared to the wild type and *snq2∆yor1∆*, we performed qRT-PCR in both the solvent control and fluconazole. In both the solvent control and fluconazole, *PDR5* transcript abundance was elevated in the *snq2∆yor1∆ybt1∆ycf1∆* strain relative to the wild type strain (p = 0.039 in DMSO; p = 1.6 x 10-3 in fluconazole; Fig XX). However, we did not find evidence of increased *PDR5* transcript levels in the *snq2∆yor1∆* strain (p = 0.366 in DMSO; p = 0.15 in fluconazole; Fig. XX). This is consistent with the lack of improved fluconazole resistance of this strain (p = XX, Fig. XX) and suggests strain-specific differences in the induction phenomenon. Thus, while we observed transcriptional activation of *PDR5*, here we required additional knockouts in *ybt1∆* and *ycf1∆*.

As an alternative mechanism of *PDR5* repression by Snq2 and Yor1, we investigated the possibility of a protein-protein interaction based mechanism. A previous study investigating an analogous compensatory activation of Snq2 by deletion of *PDR5* and *YOR1* found unchanged protein abundance and localization upon knockout of these two genes, and suggested that physical interaction between these two proteins may lead to their mutual repression24. We also noted that homodimers of Pdr5p, Snq2p, and Yor1p have been detected using the mDHFR protein complementation assay (PCA)25,26, and that the Pdr5p homodimer has also been detected using the membrane yeast-two-hybrid (MYTH)24 assay. Furthermore, an interaction has been reported between Pdr5p and Snq2p with both assays24,25. If these proteins require homodimerization to efflux their given drugs, or the heterodimer forms are otherwise unable to perform this function, then it would be predicted that deletion of any of these genes may shift ABC transporter complex levels towards a greater abundance of their active form. However, this model also predicts the presence of a previously-unreported Pdr5p-Yor1p interaction. We re-tested a potential Pdr5p-Yor1p interaction, and found it using MYTH but not PCA (Fig. ?? XX), lending consistency to the PPI-based repression model.

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

that the repression appears to be bi-directional for Pdr5p and Snq2p, both in our data (XX Fig ?? XX) and from previous reports22, we hypothesize that the homodimeric form of Pdr5p may be responsible for the efflux of its respective drugs, and that the formation of Pdr5p::Snq2p and Pdr5p::Yor1p heterodimers may deplete the concentration of Pdr5p::Pdr5p homodimers available for efflux.

**Discussion**

Here we engineered and profiled a large population of strains containing combinatorial gene deletions, and discovered many unexpected multi-gene knockout effects leading to both completey nove and extended roles for the 16 ABC transporters studied. Many knockout phenotypes were highly dependent on the combination of drug stress and knockout background, and we envision that studies of other gene families or pathways would reveal similarly-complex patterns. While principles guiding mechanistic understanding of complex genetic interactions are currently limited, we found that with some prior knowledge of gene function, complex multi-knockout effects can be functionally interpreted and used to construct clear follow-up experiments and models. Here we hypothesized all complex genetic patterns to result from a combination of parallel clearance and antagonism between transporters. Amongst these relationships, novel roles were found for *YBT1* and *YCF1* in mediating both resistance and sensitivity to the tested compounds, which may have been missed due to their subtle or absent single-knockout effects in a wild-type background. For example, the complex drug resistance phenomenon observed in fluconazole extends previous reports of compensatory activation between ABC transporters22,24 to include *YBT1* and *YCF1*. We note that the human ABC transporter ABCC3 has also been shown to exhibit compensatory activation upon knockout or disruption of ABCC2, both in rat knockout models33, and in the context of Dubin-Johnson syndrome34, motivating similar studies of this gene family in other model organisms. Several drugs also exhibited multi-knockout sensitivity patterns, suggesting overlapping substrate specificity, and revealing the differential contribution of each transporter towards resistance. We demonstrate that even within a highly-characterized gene family, an engineered population profiling strategy has the potential to reveal many new functions.

Here, a previously-engineered ABC-16 strain allowed the use of a cross-based strategy to efficiently introduce mutations into a population of cells. In previous studies, profiles of individuals from controlled crosses have been used to study complex traits such as gene expression27 and resistance to small molecules28. Even in more complex model organisms such as mice, initiatives like the collaborative cross project have made use of controlled crosses between inbred strains to create numerous defined heterozygotes that can be used to study both the variability and genetics of mouse traits29. The engineered population design extends these cross-based methods by limiting the genetic variation to include only targetted modifications within a defined set of genes, creating a clear link from association to causality. Efforts to introduce targetted gene knockouts30 or loss-of-function mutations31 into model organisms such as *C. elegans* may enable the use of this strategy with future engineered multi-knockout parental strains. The initial parental strains for such crosses may be generated by a series of targetted matings, and now more straightforwardly with the use of modern genome engineering tools such as CRISPR. A cross-based strategy allows flexibility in the distribution of knockouts between the two parents, and may be useful in cases where introducing all knockouts into a single parental strain would cause considerable defects or lethality. In such cases, the use of conditional alleles or similar tools to suppress complex synthetic lethality when creating the population may also be useful32.

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

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

In addition to the context-dependent growth measured in this study, many traits can be measured and linked to a specific genotype *en masse*. For example, to understand drug efflux, cells may be exposed to a fluorescence-conjugated drug, and fluorescence-activated cell sorting at different time points coupled with barcode sequencing may recreate efflux dynamics for many strains. In a similar fashion, barcode sequencing coupled with cell sorting may be used to reconstruct many phenotypes, for example the effects of knockouts on the activation of signalling pathways, phosphorylation state, epigenetic modifications, or protein abundance35. Furthermore, methods are being developed to genotype large populations of cells after imaging, which would allow relating many multi-knockout genotypes to complex phenotypes obtained by high content screens36. Thus, with the appropriate design, multiple knockouts obtained by population engineering can be related to a wide variety of phenotypes, both simple and complex.

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

**Materials and Methods**

**Yeast Strains**

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

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

RY0146 (Toolkit-a strain):

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

RY0148 (Barcoder Strain MATα):

*MATα lyp1Δ his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 can1Δ::GMToolkit-α (CMVpr-rtTA NatMX4 STE3pr-LEU2) ho*∆*::Barcoder locus* (UPtag HpHMX4 DNtag )

**Media**

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

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

**DNA Oligomers Used**

US1: ATGCCTTTGACGACCCGAGATG

US2: AGTGTGAGCCGAGGTTCTGAGG

DS1: GAATCCCGACTGCCGTGGTTAG

DS2: GCGTATGGTGCGAGTATCAGCG

U1:  
U2:

D1:

D2:

5’HO:

3’HO:

Gene deletion primers

Gene confirmation primers set1

Gene confirmation primers set2

**Creating the Barcoder Plasmid**

A pSH47 plasmid backbone expressing GAL1pr-CRE was combined with a ‘barcoder locus’ flanked by LoxP and Lox2272 sites using in-yeast assembly37. This barcoder locus consists of a random 25bp DNA sequence (‘UPTAG’) in between two common primer regions (US1 and US2), followed by a HphMX4 cassette, and then by another random 25bp DNA sequence (‘DNTAG’) in between two common primer regions (DS1 and DS2). The barcoder locus was generated using fusion PCR of two synthesized fragments.

**Generating a Barcoder Strain**A linear URA3 cassette flanked by LoxP and Lox2272 sites was integrated into the HO gene of the RY0148 strain through transformation to serve as the locus for barcode integration. This strain was further transformed with the barcoder plasmid, and transformants were selected using HygroB containing media. These transformants were then subjected to a 4 day galactose induction to allow for Cre expression. All strains which had either lost the plasmid and/or were unsuccessful in Cre-mediated recombination were counter-selected by growth in the presence of 5-FOA. Strains surviving the counterselection were then grown in YPD in the presence of HygroB in order to select for successful integration of the plasmid barcode locus. Ten colonies were tested for barcode integration using two sets of primers to verify the strain barcode-specific UPTAG and DNTAG - US2 and a sequence complementary to 5’ of the HO gene (5’HO); DS1 and a sequence complementary to the 3’ of the HO gene (3’HO). PCR products were analyzed using gel electrophoresis.

**Generating Barcoded Random Knockout Progeny**

Mating, sporulation, and haploid selection were performed between the RY0622 ‘Green Monster’ strain (MAT**a**) and the RY0148 barcoder strain (MATα) as previously described15. Using colony plating, sporulation efficiency was estimated at 24% - 1080 colonies grew in SC, 140 colonies grew in SC-His (MAT**a** haploid selection), and 120 colonies grew in SC-Leu (MATα haploid selection). Samples from the haploid pools were plated as a form of quality control. From the SC-His pool, 440 colonies grew in SC-His, 0 grew in SC-Leu, 468 grew in YPD+G418, 13 (~2%) grew in YPD+Nat; from the SC-Leu pool, 0 colonies grew in SC-His, 1600 grew in SC-Leu, 50 (~3%) grew in YPD+G418, 1650 grew in YPD+Nat. The two pools were then grown in YPD+HygroB to select for barcoded haploids. The SC-Leu pool was further grown in SC-Ura to select against barcoder strain parents that may have escaped diploid selection. Using a QPix colony picker, 5,078 colonies were picked from the MAT**a** pool and 5,844 colonies were picked from the MATα pool onto 384 well plates.

**Individual Strain Genotyping**

To genotype each strain at the 16 ABC transporter loci, two PCR reactions were performed for each locus - one to determine the presence of a GFP integration cassette, and another to determine the presence of the wild type gene. Genotyping was performed as previously described15. For the first set of reactions, locus–specific PCR primers from the 5′ flanking sequences of each gene were paired with a common primer complementary to the *GFP* cassette (See Additional Data SX). For the second set of reactions, the 5’ locus-specific primers were paired with primers which were complementary to an internal sequence of their corresponding gene (One of two such pair sets was used per gene, see Additional Data SX). PCR products were analyzed using gel electrophoresis.

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

To create a ‘Gold Standard’ genotyped set, 40 progeny strains (19 MAT**a** and 21 MATα) were subject to individual strain genotyping. For these 40 strains, and for an RY0148 isolate, the strain-specific UPTAG and DNTAG were PCR amplified using two sets of primers – US2 and a sequence complementary to 5’ of the HO gene (5’HO); DS1 and a sequence complementary to the 3’ of the HO gene (3’HO). The PCR products were subject to Sanger sequencing. These 40 ‘Gold Standard’ genotyped strains, as well RY0148 and RY0622 were arrayed into the strain collection at defined positions to act as genotyping controls (Data\_S\_Genotypes).

**Pooled Strain Genotyping**

A previously-developed Row-Column-Plate (RCP)-PCR protocol21 was adapted in order to perform *en-masse* genotyping of the random knockout progeny using high throughput sequencing. In order to uniquely tag PCR products originating from the same well, all forward primers had a 5’ tag encoding the well row (R), and all reverse primers had a 5’ tag encoding the well column (C)21 (Additional File XX).

For each well in the collection, four PCR reactions were performed with the following primers: R+U1 and C+U2 to amplify DNA barcodes encoding the UPTAGs for each gene deletion; R+D1 and C+D2 to amplify the deletion-specific DNTAGs; R+US1 and C+US2 to amplify the strain-specific UPTAG; R+DS1 and C+DS2 to amplify the strain-specific DNTAG. PCR reactions were performed using a Hydrocycler. A plate-specific Illumina adapter (plate tag) were added using an additional PCR reaction as previously described21.

**Analysis of Pooled Strain Genotyping Data**

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

For each sample, a pipeline determined the strain-specific TAG sequences and genotype from the reads. The parameters of the genotype caller from this pipeline were trained to best match the genotyping calls of strains known to be in the control set, and the cross-validated accuracy for each gene is reported in Fig. S1A.

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

**Examining Putative Wild-Type Pool Strains**

For 73 unique MAT**a** strains and 131 MATα strains, pooled genotyping had called the strains as wild-type. 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 MATα strains, 74 exhibited no detectable growth in SC-Ura, and likely arose from remaining barcoder parents which had escaped a previous SC-Ura selection step. The genotypes for the these 74 strains were kept as is, while the other 23 strains, as well as 46 untested strains were discarded from the analysis. Out of 45 MATastrains, all exhibited growth in SC-Ura. Individual genotyping was performed for these MATa strains, and was successful for 40 of 45 strains. These strains had their stated genotype was corrected (Data\_S\_Genotypes XX). The 5 unsucessfully genotyped strains, as well as 28 additional strains were discarded from analysis. When calculating linkage and distribution of gene knockouts, the wild-type MATα strains were excluded from analysis as they were likely parental strains rather than progeny arising from mating.

**Estimating Genotyping Accuracy by Knockout Distribution**

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

**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% (Additional File XX). All drugs used were dissolved in 2% DMSO, which was used as a solvent control. Growth was determined by the Average\_G metric38, which represents the average generation time.

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

Progeny with at least one mapped strain-specific barcode (See Additional Data XX) were combined into two separate liquid YPD + glycerol pools, separated by mating type. 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 1uL of 10μM R primer. F and R primer pairs were PXX+US1/ PYY+US2 and PXX+DS1/PYY+DS2 for the strain-specific UPTAG and DNTAG, respectively. PXX and PYY correspond to sequences containing plate-specific Illumina sequencing adapters, as well as tags which were used to demultiplex the samples (See Data\_S\_twas\_phenotyping\_primers\_used XX). PCR products amplified using the following PCR program: 98C for 30sec; 98C for 10sec, 60C for 10sec, 72C for 1min (24 times); 72C for 5min; 4C forever.

PCR products were subject to gel electrophoresis, and 210-220bp 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 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 reads where the strain identity call differed between tags were discarded. A version of this script was also run which does not restrict the mating type of the potential barcode match to verify that samples were not mislabelled or contaminated.

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

**Defining a Resistance Metric**

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

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

The drug resistance score of a strain is defined as the ratio of its total growth in a drug compared to its growth in the solvent control , per generation:

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

The multiplicative model of genetic interactions23 was applied to the metric. In this model, the expected resistance of a double knockout strain in a given drug () is the product of the resistances of the corresponding single knockout strains:

To use this model in an additive rather than multiplicative form, the log of the resistance metric was taken , so that:

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

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

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

This definition can be analogously extended for interactions of arbitrary complexity, and signficant terms denote interactions between the corresponding knockouts. The genotype of each strain is encoded using a binary variable, so that if and only if all the corresponding genes in the term are knocked out.

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

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

To train models containing terms of up to -way complexity, additional steps were performed. First, an initial set of a genes was chosen by testing whether a knockout of the 16 ABC transporters resulted in a significant population-level difference in log-resistance, first in the context of the overall population, then in the context of all additional gene knockouts (where is the desired level of gene-gene interaction complexity to be fit in the overall model). Significant differences were evaluated using a t-test, requiring a significance level of , where is the total number of genetic background contexts tested for that gene. A linear model including genes passing this test and all of their way interactions was then fit.

To speed up feature elimination, many terms were first heuristically eliminated using cross-validated LASSO regression implmented using the cv.glmnet function in R (using the default parameters of alpha = 1, nfolds = 10, type.measure = ‘deviance’). A λ value was then chosen which minimizes the mean cross validated error, and all terms with a nonzero coefficient at that λ value were selected for further elimination. These terms were subjected to stepwise elimination (again eliminating the gene with the highest value at each step), at a significance threshold of where is the number of terms in the initial linear model.

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

The TWAS21230902 strain (*pdr10∆ pdr18∆ pdr5∆ snq2∆ ybt1∆ ycf1∆ yor1∆;* Data\_S\_Genotypes) was subject to individual strain genotyping, confirming the pooled strain genotyping results. This strain (MAT**α**) was mated with RY0146 (MAT**a**), and subject to sporulation and MAT**a** haploid selection15. Individuals from this cross were arrayed onto a 384 well plate, and individually genotyped at *PDR10* and *PDR18*. Strains with no deletions at these genes were further genotyped at *PDR5, SNQ2, YBT1, YCF1,* and *YOR1.* After analysis of genotyping results, one strain of each genotype combination was chosen to create the 32-strain collection. These chosen 32 strains were re-genotyped at these 5 loci for validation.

**Analysis of Liquid Growth Data**

**MYTH and PCA Assays**

The MYTH and PCA assay were performed as previously described24,25.

**Quantitative RT-PCR**

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

**Availability of Data and Materials**

**Competing Interests**

The authors declare that they have no competing interests.

**Acknowledgements**

This work was supported by XX.

**Author Contributions**

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

**Additional Data Files**

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

7. 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).

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

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

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

11. 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).

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

13. 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).

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

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

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

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

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

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

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

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

22. 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).

23. 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).

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

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

26. 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).

27. 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).

28. 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).

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

30. C. elegans Deletion Mutant Consortium. Large-Scale Screening for Targeted Knockouts in the *Caenorhabditis elegans* Genome. *G3&amp;#58; Genes|Genomes|Genetics* **2,** 1415–1425 (2012).

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

32. Morawska, M. & Ulrich, H. D. An expanded tool kit for the auxin-inducible degron system in budding yeast. *Yeast* **30,** 341–51 (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. Brockmann, M. *et al.* Genetic wiring maps of single-cell protein states reveal an off-switch for GPCR signalling. *Nature* **546,** 307–311 (2017).

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

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

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

**Figures**

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

A strain with 16 ABC transporter deletions is crossed with a pool of wild-type cells, each containing a DNA barcode. Diploids from this cross are subject to sporulation and barcoded haploid selection. Single colonies are picked and arrayed onto a series of 384-well plates. High-throughput genotyping using an adapted RCP-PCR21 strategy is performed on this collection and genotypes are linked to strain-specific DNA barcodes, allowing pooling and strain identification using high throughput sequencing. High throughput barcode sequencing at different time points allows the creation of a drug by strain resistance profile.

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

**A** A linear landscape of drug resistance to fluconazole, benomyl, and mitoxantrone in the MATa pools. Each point represents a group of strains, and genotype is indicated by the colour code on the right. Within each group, variation is allowed at the 10 other ABC loci. Strain groups are arranged horizontally by number of gene knockouts, with jitter added for visual clarity. Vertical position indicates average drug resistance of a group, and lines connect groups separated by a single knockout. Solid lines indicate a significant difference in fitness between groups (Bonferroni-adjusted p < 0.05, Mann-Whitney U test), dashed lines indicate non-significant fitness differences.

**B** A radial fitness landscape in fluconazole. The centre represents the fitness of the group with 6 wild type genes, and single knockouts extend radially in a fixed order. The mean resistance of each group is indicated by the given colour scale, which is centered by the 6-wild-type resistance (black), and extremes represent plus (blue) or minus (orange) the largest observed difference in resistance between this and any other genotype group. Additional knockouts are indicated by continous radial extensions in a fixed order, and all paths where any further knockouts do not result in significant changes in drug resistance (Bonferroni-adjusted p < 0.05, Mann-Whitney U test) are pruned. The inset highlights the resistance path of sequential *ybt1∆, yor1∆*, *snq2∆*, *ycf1∆*,and *pdr5∆* knockouts in the MATa fluconazole pool.

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

**D** Distribution of fluconazole resistance amongst all *ybt1∆,* *yor1∆*, *snq2∆, ycf1∆,* and *pdr5∆* knockout groups in the MATa pool. Group genotype is indicated on the sides as in A, and *pdr5∆* groups are paired with their corresponding *PDR5* equivalent in each line. Distributions for each group are given by their kernel density.

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

**A** A linear model was used to formally determine significant and reproducible gene knockouts and genetic interactions conferring resistance to the tested drugs. The magnitude of reproducibly significant (Bonferroni adjusted p < 0.05) single-knockout or genetic interactions effects are indicated in the heatmap, with grey boxes indicating effects which were not significant in both pools for a given drug. Coefficents are sorted by the number of genes involved, as indicated by the colour code on the left.

**B** Schematic interpretation of the multi-knockout fitness landscape and surprising interactions in the genetic interaction profiling data.

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

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

**A** Per-gene genotyping accuracy of the 16 ABC transporters surveyed. Genotyping accuracy was estimated through cross-validated evaluation of the high-throughput genotyping procedure by comparison to known reference strains.

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

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

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

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

**B** As in A), analysing the MAT**α** pool.

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

A linear model was used to formally determine significant and reproducible marginal gene knockout effects towards resistance the tested drugs. The magnitude of reproducibly significant (Bonferroni adjusted p < 0.05) single-knockout are indicated in the heatmap, with grey boxes indicating effects which were not significant in one or both pools for a given drug.

**Figure S4.** Reproducibility of marginal linear models.

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

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

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

**Figure S6.** A linear landscape of resistance to 16 drugs.

**A** Each point represents a group of strains, and genotype is indicated by the colour code on the right. Within each group, variation is allowed at the 10 other ABC loci. Strain groups are arranged horizontally by number of gene knockouts, with jitter added for visual clarity. Vertical position indicates average drug resistance of a group, and lines connect groups separated by a single knockout. Solid lines indicate a significant difference in fitness between groups (Bonferroni-adjusted p < 0.05, Mann-Whitney U test), dashed lines indicate non-significant fitness differences. Values are shown for the MAT**a** pool.

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

**Figure S7.** A radial fitness landscape in 6 additional drugs.

The centre represents the group with 6 wild type genes, and single knockouts extend radially in a fixed order. The resistance of each group relative to the wild type is indicated by the given colour scale, which is centered by the wild type resistance (black), and extremes denote +/- the largest difference in resistance between the wild type and any other genotype group. Additional knockouts are indicated by continous radial extensions in a fixed order, and all paths where any further knockouts do not result in significant changes in drug resistance (Bonferroni-adjusted p < 0.05, Mann-Whitney U test) are pruned.