Saturation transposon mutagenesis enables genome-wide identification of genes required for growth and fluconazole resistance in the human fungal pathogen *Cryptococcus neoformans*

R. Blake Billmyre[1,2,3,4,5*], Caroline J. Craig[5,#], Joshua Lyon[1,2], Claire Reichardt[1,2,3], Michael T. Eickbush[5], and Sarah E. Zanders[5,6]

[1]Department of Pharmaceutical and Biological Sciences, College of Pharmacy, University of Georgia, GA, United States,[2]Department of Infectious Diseases, College of Veterinary Medicine, University of Georgia, GA, United States, [3]Department of Microbiology, Franklin College of Arts and Sciences, University of Georgia, GA, United States, [4]Department of Genetics, Franklin College of Arts and Sciences, University of Georgia, GA, United States, [5]Stowers Institute for Medical Research, Kansas City, MO, USA, [6]Department of Cell Biology and Physiology, University of Kansas Medical Center, KS, United States [#]Current affiliation: University of Utah Medical School, UT, United States

*Corresponding author:

Email: blake.billmyre@uga.edu
Phone: 706-583-0216

Abstract

Fungi can cause devastating invasive infections, typically in immunocompromised patients. Treatment is complicated both by the evolutionary similarity between humans and fungi and by the frequent emergence of drug resistance. Studies in fungal pathogens have long been slowed by a lack of high-throughput tools and community resources that are common in model organisms. Here we demonstrate a high-throughput transposon mutagenesis and sequencing (TN-seq) system in *Cryptococcus neoformans* that enables genome-wide determination of gene essentiality. We employed a random forest machine learning approach to classify the *Cryptococcus neoformans* genome as essential or nonessential, predicting 1,465 essential genes, including 302 that lack human orthologs. These genes are ideal targets for new antifungal drug development. TN-seq also enables genome-wide measurement of the fitness contribution of genes to phenotypes of interest. As proof of principle, we demonstrate the genome-wide contribution of genes to growth in fluconazole, a clinically used antifungal. We show a novel role for the well-studied *RIM101* pathway in fluconazole susceptibility. We also show that 5' insertions of transposons can drive sensitization of essential genes, enabling screenlike assays of both essential and nonessential components of the genome. Using this approach, we demonstrate a role for mitochondrial function in fluconazole sensitivity, such that tuning down many essential mitochondrial genes via 5' insertions can drive resistance to fluconazole. Our assay system will be valuable in future studies of *C. neoformans*, particularly in examining the consequences of genotypic diversity.

Introduction

Human fungal pathogens are a serious human health risk, causing approximately 3.8 million deaths a year (1). Fungal infections are frequent in immunocompromised patients, particularly in conjunction with the HIV/AIDS epidemic. While the number of severely immunocompromised HIV patients has declined in recent years with the advent of HAART therapy, HIV management remains difficult in resource-limited areas and HIV-related illness remains the leading cause of death in Sub-Saharan Africa. Even in resource-rich areas, HIV is still a significant challenge. In the US, up to 1 in 13 individuals who were at one point virally suppressed experience viral rebound (2). Other sources of immunocompromise are also on the rise as solid organ transplants and the accompanying immunosuppression become more common. In combination, we face continued threats from fungal diseases.

In addition, treating fungal infections is highly challenging. On an evolutionary scale, fungal pathogens are more closely related to their human hosts than almost every other infectious disease, except animal pathogens like parasitic worms. This high relatedness makes fungi strong tractable models for human biology, as many aspects of mammalian biology are conserved in fungi. A drawback of this conservation is that many potential inhibitors of fungal growth are unacceptably toxic to humans because they inhibit shared biological pathways (3). As a result, we have only five classes of antifungal drugs available to treat invasive diseases: azoles, echinocandins, polyenes, the nucleoside analog 5FC, and a relatively new drug ibrexafungerp. Two of these five classes, echinocandins and ibrexafungerp, target the same gene (*FKS1)*. Most of these drugs target biology unique to fungi. For example, azoles, such as fluconazole, inhibit the essential activity of Erg11, which is responsible for part of the biosynthesis of ergosterol, a

fungal analog of cholesterol not found in humans. The paucity of antifungal drugs means that when resistance emerges or when a species is intrinsically resistant to an antifungal drug, there are few therapeutic alternatives.

Among the most dangerous fungal pathogens are species from the *Cryptococcus* pathogen species complex (4). *Cryptococcus neoformans* is the most common source of infections within a pathogenic species complex composed of at least 7 different species of pathogens (5). The vast majority of the approximately 147,000 annual deaths from Cryptococcosis (1) are caused by infections with *C. neoformans* and occur in Sub-Saharan Africa in conjunction with the HIV/AIDS epidemic (1). *C. neoformans* rarely infects immunocompetent individuals. In immunosuppressed individuals, infection typically begins in the lungs as fungal pneumonia and can then disseminate through the bloodstream into the brain to cause fungal meningitis. Cryptococcal meningitis is uniformly fatal if left untreated.

Treating Cryptococcal meningitis is complicated by the fact that *Cryptococcus neoformans* is intrinsically resistant to both echinocandins and ibrexafungerp (6,7). Multidrug resistance is not common because invasive fungal pathogens, including *C. neoformans*, are generally acquired from environmental reservoirs. Infections do not typically transmit from patient to patient or even from a patient back into an environmental reservoir. This means that multidrug resistance has not typically accumulated over time the way it does in many other pathogens. However, agricultural drug exposure may select for resistance within the environmental reservoir. For example, agricultural exposure to azoles may be contributing to the increasingly common azole resistance in the human fungal pathogen *Aspergillus fumigatus* (8,9). It remains to be seen whether *Cryptococcus* will similarly acquire azole resistance, but a strong possibility exists.

Collectively, we have an urgent need to develop new and synergistic therapies to treat fungal diseases (10). In *Cryptococcus*, this effort has been hampered both by a lack of available experimental tools and by a poor understanding of the essential gene set. Essential genes, particularly those not conserved in humans, are the ideal set of genes for targeted design of novel antifungals. Past systematic efforts to define essential genes in *Cryptococcus* by taking advantage of diploid genetics have scaled poorly, resulting in relatively small numbers of defined essential genes (11). In contrast, most inferences of essentiality were drawn from the ability/inability to obtain gene deletions during the construction of two *Cryptococcus* systematic deletion collections (12). However, absence from a deletion collection can be explained by transformation failure rather than essentiality. In addition, many large-scale deletion collections are prone to errors, even in organisms with large communities iteratively fixing them (13). One study using the *Cryptococcus* collection found 14 of 82 tested strains that were hits in a separate CRISPR-based screen failed to produce the correct diagnostic digest pattern that would indicate proper deletion of the relevant gene (14). Using these data to predict essentiality across the genome is thus likely to be error prone.

One alternate approach to determine essentiality is an approach called transposon mutagenesis sequencing or TN-seq. TN-seq was originally developed in bacteria (15–18) and more recently has been applied to fungi, including *S. pombe* (19–22), *S. cerevisiae* (23–25), and *C. albicans* (26) among others. This approach inserts one transposon randomly into the genome in each cell within a pool. Insertions into essential genes kill the cells, which are thus removed from the pool. Insertions into non-essential genes are tolerated and can be identified using a high throughput sequencing approach (Figure 1A). In pools with millions of cells and in organisms with relatively small genomes, TN-seq can generate saturated maps of genome function. TN-seq has

been used broadly to identify essential genes, with some studies using machine learning to predict essentiality of the entire gene set based on an initial subset of confidently assigned genes (26,27).

Further, once a TN-seq insert library has been generated, that library can be selected and resequenced to quantitatively measure the contribution of the remaining nonessential genes to a given phenotype. In eukaryotes, this approach has powered analysis of contributions to heterochromatin formation (20), sexual reproduction (22), general fitness (21), and chemical genetics (23,28). Similar approaches have been explored using CRISPR in *Cryptococcus*, but thus far limitations in effective guide RNAs and resulting low insert densities have limited the statistical power in comparison to TN-seq (14).

Here we have developed a TN-seq system for *Cryptococcus neoformans* using a modified Ds (dissociation) transposon mobilized by the Ac (activator) transposase from maize. This transposon acts using a cut and paste mechanism and was originally identified by Barbara McClintock (29). The Ds transposon has been adapted for multiple fungal TN-seq systems (25,26). We generated a dense insertion library and applied a machine learning model to predict essentiality across the genome. Further, we have selected this library using fluconazole to identify modifiers of drug susceptibility. We demonstrate that inserts into regulatory regions of predicted essential genes can sensitize a strain to drug treatment without altering wildtype fitness. To our knowledge, this is the first use of TN-seq to query essential gene function in fungi. In sum, we demonstrate that TN-seq enables both identification and analysis of essential genes for their role in drug susceptibility in *Cryptococcus neoformans*.

Results

To determine the set of essential genes in *Cryptococcus neoformans*, we developed a TN-seq assay to generate random insertional mutations across a pool of cells (Figure 1A-B). Because many of the existing fungal TN-seq systems utilized plasmid systems that were unavailable in *Cryptococcus*, we based our approach on a previous TN-seq method in *Candida albicans* (26), with some modifications. In our previous work with a Hermes transposon-based TN-seq system in *Schizosaccharomyces pombe* (22), we observed that transposon insertions (~2.6 kb) were not tolerated within the introns of essential genes. The largest described intron in *S. pombe* is only 819 nucleotides (30), suggesting that the spliceosome is unlikely to capable of splicing sequences as large as a transposon insertion. In fact, extension of a functional artificial intron from 36 bases to larger sizes, including both 350 and 252 base introns, showed that increased length introns often fail to splice at all in *S. pombe* (31).

Introns are even slightly smaller in *Cryptococcus neoformans*, with a mean length of 65 bases (32). There is also a bias towards RNAi silencing of genes with longer introns (33). With the assumption that a transposon inserted into an intron would disrupt gene function in *Cryptococcus*, we integrated a Ds transposon carrying a neomycin resistance cassette into the first intron of the *URA5* gene in the H99 *C. neoformans* type strain. This insertion disrupts the gene, but gene function is generally restored when the transposon is excised during transposition, even if excision is imperfect (ie. reading frame does not need to be maintained during repair) (Fig 1C). These features allow us to select both against transposon movement (on media containing 5-FOA) and for transposon excision (on media lacking uracil). The ability to select

against transposon movement enabled us to grow large cultures prior to triggering transposition to avoid jackpot events (i.e. clonal expansion of a limited number of independent mutations).

To mobilize the Ds transposon, we integrated a codon-optimized version (designed using Optimizer (34)) of the Ac transposase under the control of the *GAL7* promoter at the neutral safe haven locus (35). We generated a transposon insertion library by selecting transposition events on media containing galactose but lacking uracil so that the transposase could be induced but cells could only grow if they regained the ability to produce their own uracil after transposition (Figure 1C, see methods).

To identify and quantify transposon insertion sites, we amplified the boundaries between transposon insertions and the genome using PCR and sequenced them using Illumina sequencing (modified from (22), see Methods). We integrated random barcodes during the sequencing library preparation, enabling us to eliminate PCR duplicates and more accurately measure transposon insertion frequency within a pool (as in (21)). We mapped only reads containing PCR-amplified Ds fragments to the *Cryptococcus* genome. We identified 18,417,912 reads mapping to 1,750,772 unique sites in the genome including coding sequences of 6,190 of the 6,975 total annotated genes or 92.5 unique insert sites per kb, on average.

We observed some insertional bias in transposon location. Insertions near the original Ds site in *URA5* were much more common than those in distant locations, suggesting that the transposon was more likely to insert into locations spatially near the original excision site (Supplemental Figure 1A). This was previously observed in the SATAY TN-seq system of *S. cerevisiae*, which also uses an Ac/Ds transposon system (23). We also observed some preference for insertion within the upstream regions of genes (Supplemental Figure 1B). These biases were much

stronger in genes we predicted to be essential (see below) than in those we predicted to be nonessential, suggesting that at least a portion of the bias resulted from selection rather than bias in the location of insertion (Supplemental Figure 1B). Despite these biases, even after discounting inserts on the chromosome containing *URA5* (Chromosome 8), we retained 1,322,416 unique inserts or 75.5 inserts per kb on average. Of these remaining inserts, 37% were genic (40 inserts per kb) and 63% were intergenic (159 inserts per kb). Because the mean gene length in this region (including introns) was 1,906 bases, on average we would still expect 76 insert sites per typical gene, giving us confidence that we could score gene function despite these biases. In contrast, the average gene on Chromosome 8 contains 432 insert sites and intergenic sequences on Chromosome 8 have 472.7 inserts per kb on average.

Machine learning helps predict essentiality

TN-seq is commonly used in bacteria to identify essential genes (36). This approach generally relies on using a set of *bona fide* essential genes, which should have few transposon inserts, and *bona fide* non-essential genes, which should have many inserts. In *C. neoformans*, there is a limited set of *bona fide* essential genes. *ERG11* is one such essential gene and the target of the antifungal drug fluconazole. We could very clearly see that Ds insertions within the coding sequence of *ERG11* were not tolerated in our mutant library (Figure 2A). Similarly, transposon insertions were enriched in genes deleted within the *Cryptococcus* deletion collection compared with those not contained within the collection ($p=3.47 * 10^{-15}$, Supplemental Figure 1C), with some clear exceptions.

To assess essentiality across the entire *C. neoformans* genome, we built a random forest machine learning model (Figure 2B). We started by taking the set of genes and essentiality predictions reported in Segal et al. (26) which contained ortholog and essentiality predictions for *C. albicans*, *S. pombe*, and *S. cerevisiae* genes. We then used FungiDB to identify *Cryptococcus neoformans* orthologs of these genes (37). Selecting only genes with identifiable orthologs in all four species left us with 1,653 genes. There are existing essentiality predictions for all species except *C. neoformans* (26). We reasoned that the majority of these genes conserved over the approximately 600 million years of evolution separating basidiomycetes from ascomycetes would likely have conserved essentiality as well, enabling us to use this set of genes to train our model. Specifically, genes essential or nonessential in *S. cerevisiae*, *S. pombe*, and *C. albicans* were similarly assumed to be essential or nonessential in *C. neoformans*.

We built our model by first parameterizing our TN-seq data set based on those used in previous studies (Figure 2C; (26,27)). We captured 11 variables for each gene, 9 of which described the transposon insertions (total insert frequency within a gene, total insert frequency within a gene normalized to gene length, number of unique transposon insert locations within a gene, number of unique transposon insertion locations within a gene normalized to gene length, total insert frequency within the middle 80% of a gene, total insert frequency within the middle 80% of a gene normalized to gene length, length of largest insert-free gap, length of largest insert free gap normalized to gene length, and total transposon frequency in the 100 kb surrounding the gene) and 2 of which described the gene absent TN-seq information (chromosome and length of gene). We then randomly split our set of 1,653 genes into a training set of 1,322 genes and a validation set of 331 genes. This set includes 600 genes we assumed to be essential and 1,053 we assumed to be nonessential. We randomly chose the training and validation subsets 100 times and built a

random forest model from the training subset each time, testing the model against the validation set, and predicting essentiality (Essentiality Scores ranging from 0-1, with 1 being essential and 0 being nonessential) for every gene on every iteration. Across the 100 models, we obtained a maximum combined precision (true positives/(true positives + false positives)) and recall (true positives/ (true positives + false negatives)) at a threshold essentiality score of 0.41, where the mean precision was 0.78 and the mean recall was 0.9 (Figure 2D). For consistency with past models, we used the default cutoff of 0.5, which still had a mean precision of 0.8 and recall of 0.88. This model compares favorably to one used for *C. albicans* (Precision 0.71, Recall = 0.62) (27). Given that some differences are likely biological (ie. a conserved essential gene is not essential in *Cryptococcus*) rather than model defects, these scores are quite strong.

We then used this model to predict the essentiality score for each gene from the 100 models and used the variation across these models to test whether the mean essentiality score (ES) was statistically different from the 0.5 cutoff. As a result, we define three classes of genes: essential genes (ES >0.5 and $p<0.05$ after Bonferroni correction), nonessential genes (ES<0.5 and $p<0.05$ after Bonferroni correction) and unknown genes ($p>0.05$). Using this approach, we predicted 1,412 essential genes, 5,382 nonessential genes, and 181 unknown genes. We then revisited our training data using these classifications. Of 1,653 genes in our original training set, 89 were classified as unknown (5.4%), compared with just 92 in the remaining 5,322 genes (1.7%) ($p<0.0001$, Fisher's exact test). Given that the original assignment of essentiality was based on conservation, we were concerned that our model may have been overfit and was struggling to accurately determine essentiality for genes in the training set whose actual essentiality did not match the original assumption. Indeed, we identified no genes in our training set that were

statistically assigned opposite essentiality from the initial assumption we provided from homology.

We thus retrained our model using a training set without all 89 genes with unknown essentiality in our first model. As before, we ran our training and validation 100 times and determined mean precision and recall values from this data (Figure 2D). Unsurprisingly, precision and recall were improved by removing data that the model had struggled to classify and our mean precision and recall at a threshold value of 0.5 were now 0.94 and 0.87. Our revised model had just 31 unknown genes in the training set (2.0%) compared with 60 in the remaining 5,411 (1.1%). While this difference was still statistically significant (p=0.011, Fisher's exact test), the magnitude of the discrepancy was substantially reduced from the first model. Notably, this revised model was capable of assigning essentiality to many genes that differed from that assumed in the initial conservation-based assumptions (Figure 3B,C, discussed below).

We also extracted the importance of each value to the model's predictions (Figure 2E). Our model was most strongly impacted by length adjusted metrics, including the strongest metric, which was the percentage of the gene that was covered by the largest insert free gap. Interestingly, the frequency of transposon insertions in the local environment (surrounding 100 kb) and the specific chromosome had very little impact on the model, despite the biases in transposon insertion frequencies across the genome that we observed.

We then used our revised model to predict essentiality scores for every *Cryptococcus* gene (Figure 2F, Supplemental Table 1). Using the statistical approach described above, we predicted 1,465 essential genes and 5,419 nonessential genes, with 91 unknown genes. There was reasonably good overlap between our approach and the existing deletion collection. 76.4%

(n=4,139) of the genes we scored as nonessential are deleted in the collection and 92.6% (n=1,357) of the genes we scored as essential are not in the collection. Still, there are numerous discrepancies. 23.6% of the genes (n= 1,280) we predict are non-essential are missing from the deletion collection. This is not inherently surprising as there are known nonessential genes missing from the whole genome deletion collection (eg., *BAP1* and *PAN1*, TFs (38); *PBS2* and *MEC1*, kinases (39); *PTC5* and *HPP2*, phosphatases (40); and *LMP1* (41)) and the deletion collection is available prepublication rather than in a complete state. Similarly, no claims have been made that failure to construct a deletion distinguishes essentiality rather than experimental failure.

More surprising is that we classified 108 genes as essential that were present in the deletion collection. We hypothesized three potential explanations for this discrepancy. First, the construction of our mutant library required growth on SC-URA+GAL media. Any genes required for growth on this medium would score as essential in our assay. Additionally, our assay is performed in a pool. Insertion mutants with severe growth defects may be outcompeted by other strains and not be represented in the final TN-seq library, thus scoring as essential. Finally, some strains in the deletion collection may not have the intended gene deleted.

We tested putative mutants from the deletion collection for five of the genes we classified as essential with a range of essentiality prediction scores (ranging from 0.58 to 0.98) to determine which of the models above may explain the discrepancy. We started by testing all five mutants, using 5' and 3' junction PCRs and in gene PCRs (Supplemental Figure 2A). For two of the mutants (CNAG_06887 (essentiality score (ES) = 0.58) and CNAG_05292 (ES = 0.95)), we were unable to PCR validate any of 5 independent colonies from the deletion collection, including positive PCR products for an in-gene PCR, suggesting that these strains did not contain

deletions of the indicated gene (Supplemental Figure 2B, C). For the other three genes (CNAG_02190 (ES = 0.76), CNAG_04763 (ES = 0.91), CNAG_00996 (ES = 0.98)) we validated their deletion in the deletion collection, indicating our TN-seq approach misclassified the genes (Supplemental Figure 2D).

We tested the growth of the deletion alleles of the three misclassified genes. We found that the CNAG_02190 deletion had a growth defect that was exacerbated on YP+GAL and SC-URA+GAL (Supplemental Fig 3A). The CNAG_04763 deletion had a very mild growth defect on YP+GAL that was exacerbated on SC-URA+GAL and SC-URA, while the CNAG_00996 deletion made normal size colonies on both media but with unusual morphology on SC-URA+GAL (Supplemental Figure 3C). To further query the growth of these mutants we used competition assays where we competed the NAT resistant mutants against an unmarked KN99 wildtype strain. Both CNAG_02190 and CNAG_04663 were outcompeted by the wildtype while CNAG_00996 grew at relatively the same rate as the wildtype (Supplemental Fig 3B). Our data suggest that the discrepancies are likely a mix of at least two of our three models, although we lack a satisfying explanation for why CNAG_00996 scored as confidently essential in our hands (essentiality score of 0.867) but was clearly not essential and lacked a growth defect in our relevant media conditions.

Finally, Ianiri et al. characterized essential *C. neoformans* genes in their 2015 study. They identified 21 genes that were described as essential in previous publications and additionally identified 21 more essential genes by looking for the failure to recover mutant haploid progeny from heterozygous mutants (11). Our analyses predicted 39 of those 42 genes to be essential as well. One possible explanation for the three exceptions is that the assay employed by Ianiri et al required mutant spores to germinate. Mutants defective in germination would score as essential

in their assay but not ours. Alternately, in our pooled assay, essential gene mutants could be complemented by neighboring wildtype cells through production of diffusible factors. This type of complementation would not occur in the assay used by Ianiri et al. None of the three genes we predict as nonessential are found in the deletion collection, perhaps favoring an issue with the TN-seq assay rather than a germination defect.

Conservation of essential genes

One of the primary goals of identifying essential genes in fungal pathogens is to identify potential targets for antifungal drug development. Because fungi are so closely related to humans, many genes are present in both pathogens and their hosts, making these genes poor choices for antifungal development. We used FungiDB (42) to annotate if each gene in the *C. neoformans* genome had a human ortholog. This approach identified 3,817 genes without human orthologs and 3,158 genes with human orthologs (Fig 3A). However, we identified only 302 genes that we scored as essential that also lacked human orthologs, compared to the 1,163 predicted essential genes with human orthologs (Fig 3A). The predicted essential genes lacking human orthologs were enriched for functions in various biosynthetic processes (such as "small molecular biosynthetic process", p=1.87 e-21; "organic substance biosynthetic process", p=4.53 e-10; "biosynthetic process", p=8.67 e-10) and also cell division (such as "cell cycle", p=3.53 e-5; "cell division", p=7.83 e-5; chromosome segregation, p=7.23 e-4) (Supplemental Table 2).

Previous work identified 694 genes whose orthologs were essential in ascomycete yeasts *S. cerevisiae*, *S. pombe* and *C. albicans* (26). We used FungiDB (42) to identify orthologs for these genes and found that 600 of these genes also had orthologs in *C. neoformans,* which is a basidiomycete. Our model predicted that 558 of these genes were also essential in *C.*

neoformans. Once again, we used FungiDB (42) to identify human orthologs and only 34 of these 561 conserved essential genes lacked human orthologs. This compares with 30 genes identified by Segal et al. for a similar comparison (essential in *S. cerevisiae*, *S. pombe*, *C. albicans*, ortholog in *C. neoformans*, no ortholog in humans) (26). Segal et al. used a different, relatively stringent method to identify orthologs and it is likely not surprising that they identified fewer genes, even without considering essentiality in *C. neoformans*. The 34 genes we identified as conserved essential genes lacking human orthologs include genes involved in inositol phosphoceramide metabolism, including an ortholog of *AUR1*, the target of the antifungal drug aureobasidin A (43). Further exploration of these genes, as well as those specific to *C. neoformans*, will likely be fruitful in targeted design of broad or narrow spectrum antifungal agents, respectively.

When we compared our predicted essential genes in *C. neoformans* to those of the other fungi above, we identified 39 genes that had orthologs that scored as essential in the other three fungi, but our model scores as nonessential (Figure 3B). 28 of these 39 (71.8%) genes are represented in the deletion collection, which is similar to the average representation of deletions in the nonessential genome (76.4% of the scored nonessential genes have a deletion in the collection; p=0.570, Fisher's exact test). The differential essentiality of a number of these orthologs may be explained by paralogy (gene duplicates). For example, *C. neoformans* CNAG_01675 is confidently nonessential, unlike the identified orthologs in the other species. But whereas the gene is single copy in the other three fungi, there are three paralogs in *C. neoformans.* Two of these paralogs, CNAG_01675 (ES=0.000003, p=0) and CNAG_02590 (ES=0, p=0), are in our set of predicted nonessential genes while the third paralog CNAG_06770 is confidently predicted to be essential (ES=0.987, p= $1.57 * 10^{-168}$). 17 of the genes that are predicted to be nonessential

in *C. neoformans* but are essential in the other fungi have paralogs in *C. neoformans*. However, the reasons for the non-essentiality specifically in *C. neoformans* for the 22 other genes lacking paralogs are not obvious. For example, the predicted nonessential *C. neoformans* genes CNAG_03878 and CNAG_07951 genes are orthologs of essential genes involved in 60S ribosomal biogenesis, which should be an essential process. Further exploration of many of these discrepant genes will likely be enlightening.

We also queried the essentiality in *C. neoformans* of 1,053 genes that were classified nonessential in all 3 of the other fungi. 50 of these genes scored as essential in *C. neoformans* (Figure 3C). 11 of the 50 are present in the deletion collection, which is statistically enriched compared to the whole set of our predicted essential genes ($p= 0.0012$, Fisher's exact test), suggesting that some of these discrepancies are likely false positives for essentiality in our predictions. Some of these genes may be false positives resulting from the TN-seq approach. For example, *UGE1* and *UGE2* are both homologs of *GAL10* in *S. cerevisiae,* and presumably score as essential in our assay because they are required for growth on galactose, which was used to induce transposition. Both are present in the deletion collection. However, as above, other examples are less easily explained. For example, CNAG_06737 encodes an ortholog of *VPS16*, which is not essential in any of the three ascomycete yeasts, does not appear to have an annotated role in galactose metabolism, and does not appear to have paralogs in any of the three ascomycete species. While *vps16*$^{+}$ is required for high temperature growth in *S. pombe* (44), at normal temperatures mutants appear to grow faster than wildtype (45). Understanding how *VPS16* and other genes have become essential in *C. neoformans* will likely reveal unique elements of adaptation that differentiate basidiomycetes like *C. neoformans* from the ascomycete model yeasts.

TN-seq identifies genes involved in growth in fluconazole

We also used our TN-seq libraries to assay the sensitivity/tolerance of *C. neoformans* to fluconazole on a genome-wide scale (Figure 4A). Fluconazole is one of the primary drugs used to treat many fungal pathogens. While the frontline therapy for cryptococcosis is a combination therapy of 5FC and amphotericin, azoles are often used instead either alone or in combination with amphotericin. Fluconazole is also typically used for long term suppression or prophylactic treatment. Azole drugs function by inhibiting the Erg11 protein, both reducing the synthesis of ergosterol and promoting the production of toxic intermediate sterols. We performed growth curves with serial dilutions of fluconazole to determine the half-maximal inhibitory concentration (IC50) for our starting strain when grown in YPD (Supplemental Figure 4). We then split our TN-seq library to grow in YPD with either 0.1% DMSO or the IC50 level of fluconazole (13.5 ug/mL), which is dissolved in DMSO (final concentration of 0.1%). After two rounds of dilution to an OD of 0.1 followed by growth to saturation (approximately 8 doublings per round, 16 total) in either DMSO or fluconazole, we determined the transposon landscapes in both populations. We compared these landscapes to that of the starting library that we used for essential gene prediction above. To control for differential growth rates and the effect of DMSO, we compared the frequency changes at every site in the genome after growth in fluconazole to the frequency changes after growth in 0.1% DMSO. To test for statistical significance, we compared the distribution of frequency changes across a gene to those of the presumably neutral intergenic sites across the entire genome (n=454,341). Because some individual insert sites can be noisy, especially at low insertion frequencies, we only scored genes with at least 5 unique insertion locations. This filtering removed 1,228 of our predicted essential genes but still allowed us to assay 237 predicted essential genes. The filtering also eliminated 387 predicted

nonessential genes and 64 “unknown” genes. In total, we were able to score the effect of fluconazole on mutants of 5,296 genes (Supplemental Table 1).

We identified 398 genes that were specifically enriched or depleted for inserts after treatment ($p<0.05$ after a Bonferroni correction, n=5296) (Figure 4B). Of these, 185 genes had inserts that on average increased in frequency in fluconazole, meaning the mutants had a relative growth advantage in the presence of the drug, while 213 genes had inserts that on average decreased in frequency, meaning the mutants had enhanced sensitivity to the drug. We recovered known genetic interactions for fluconazole. For example, *AFR1* is well known as a major efflux pump for fluconazole (46). Transposon insertions in *AFR1* were significantly depleted after fluconazole selection (log(fluconazole/DMSO) = -0.971, p=1.52 e-57) (Figure 4B). In contrast, insertions in either of two genes, *FCY1* and *FCY2*, involved in 5FC toxicity but not fluconazole toxicity were unaffected (*FCY1*: log(fluconazole/DMSO)= 0.0474, p=0.171; *FCY2*: log(fluconazole/DMSO)= 0.0276, p=0.162) (Figure 4B) (47,48). Together, these data gave us confidence to explore the hits from our assay.

Interestingly, among the strongest hits for fluconazole sensitivity were mutants in components of the *RIM101* pathway, which had not previously been implicated in fluconazole sensitivity. Of nine described components of the *RIM101* pathway in *C. neoformans* (49), we assayed eight here and six were sensitive to fluconazole (Figure 4C). Some of these *RIM101* pathway genes exhibited similar magnitude insertion losses to the antifungal efflux pump *AFR1* (46). We had limited statistical power for some of these *RIM101* pathway genes (*SNF7* had 0 coding transposon insertions and is not pictured) because insert sites were relatively low in number even prior to selection with fluconazole. To further investigate these hits, we assayed deletion mutants of eight *RIM101* pathway genes (all but *vps25*) in spot dilution assays. All eight mutants exhibit

fluconazole sensitivity compared to wildtype (H99), even those that did not score as statistically significant in our TN-seq assay (Figure 4D).

Assaying Essential Genes using TN-seq

Essential genes are not represented in deletion collections, so these collections are not useful for directly screening essential gene functions. TN-seq mutant libraries can similarly not maintain null alleles of genes with essential functions. However, we hypothesized that insertions in regulatory regions of essential genes, which can sometimes be tolerated, may generate sensitized, or hypomorphic alleles. If so, then any selection affecting a pathway involving a given essential gene could favor or disfavor insertions in the regulatory regions of that gene. To test this hypothesis, we examined inserts near *ERG11* in our datasets. We reasoned that mutants with transposon insertions in the region upstream of *ERG11* should be specifically sensitive to fluconazole and thus should be depleted when a TN-seq library is treated with fluconazole (Figure 5A). This model was supported by our TN-seq data. Mutants with inserts into the 300 base pair region directly upstream of *ERG11* were well tolerated in DMSO but were selected against with the addition of fluconazole ($p= 4.14 * 10^{-8}$; Figure 5B). In contrast, mutants with inserts into the 300 base pair region downstream of *ERG11* were not depleted by the addition of fluconazole (p=0.383).

We then extended these analyses to assay regulatory inserts on a genome-wide scale, similar to examining coding sequence insertions. We were able to examine 5' regulatory insertions for 6194 genes (compared with 5296 for coding insertions). 1,251 genes we predicted were essential had at least 5 unique insertion sites within the 300 bp upstream of the start codon compared with only 237 predicted essential genes that had at least 5 unique insertion sites within the far larger

coding sequences. Of those 1,251 predicted essential genes, 126 had inserts significantly enriched (n=79) or significantly depleted (n=47) within 300 base pairs upstream of the start codon after fluconazole treatment ($p<0.05$ after Bonferroni correction, n=1,251).

Interestingly, when we examined predicted essential genes with significant increases in insert frequency at the 5' end in response to fluconazole (ie. mutants were more resistant), we saw a number of genes with mitochondrial functions (30 of 79 genes). These genes encoded components of the electron transport chain, the ATP synthase, mitochondrial ribosomal subunits, and others. In contrast, predicted essential genes with significant decreases in insert frequency at the 5' end had significantly fewer genes with mitochondrial function (4 of 47 genes; $p=0.0003$, Fisher's exact test). There are no obvious differences between the thirty genes we identified where inserts in the 5' end conferred resistance and the four genes where inserts in the 5' end conferred sensitivity (a mitochondrial ribosomal subunit, a malate dehydrogenase, a phosphoglycerate kinase, and an inositol phosphorylsphingolipid-phospholipase C), but there are reasonable models that could explain this pattern. For example, genes could fall into this category if they play a role in both mitochondrial function and cell wall integrity for example. Together, these results suggest that loss of mitochondrial function generally makes cells more resistant to fluconazole. This is consistent with past work showing that rotenone, a specific inhibitor of the electron transport chain complex I, is strongly antagonistic with fluconazole in *C. neoformans* (50). Our genetic data suggests this phenotype is broader than just the electron transport chain and includes general mitochondrial function.

We also found we could use the TN-seq data to guide construction of hypomorphic alleles of essential genes, using *ERG11* as a test case. Inserts across the region upstream of *ERG11* allowed us to define the regions where inserts would or would not disrupt *ERG11* function. We

generated two targeted mutants, with inserts either 296 bases upstream of the start codon that were predicted not to alter fluconazole resistance or 77 bases upstream which were predicted to decrease fluconazole resistance. Consistent with our predictions, the -77 insertion generated strains which are highly sensitive to fluconazole while growing at wildtype levels on rich media (Figure 5D), while the -296 insertions generated strains that grew at wildtype levels on fluconazole and rich media.

Shiny app allows navigation of fluconazole TN-seq data

As a resource for the *C. neoformans* community, we developed an interactive Shiny app to help readers navigate our TN-seq data (Supplemental Figure 5, https://simrcompbio.shinyapps.io/Crypto_TN_seq_viewer/). This app includes data for all genes, however, genes that lack sufficient inserts (and thus did not pass our depth cutoffs) will display only a subset of the plots. The plots enable visualization of both the distribution and frequency of transposon insertions within a gene as well as how they respond to fluconazole selection. This should enable the *C. neoformans* community to more easily navigate this dataset.

Discussion

Here we present a TN-seq system for *Cryptococcus neoformans* than enables genome-wide assays of gene function. We used our TN-seq data to predict gene essentiality and to select and identify genes contributing to fluconazole sensitivity and resistance. We identified numerous essential genes as well as novel modifiers of fluconazole tolerance and resistance. Finally, we showed that transposon insertions can help guide the construction of hypomorphic alleles for essential genes and allow analysis of essential genes in a screen-like fashion.

Essential Gene Identification

Identifying essential genes is important because these genes can serve as targets for antifungal drug discovery. We predicted 1,465 essential genes using a random forest machine learning model trained on our TN-seq data. This represents 21.0% of the genes tested. Approximately 18% of the *S. cerevisiae* genome is essential (51) and 26% of the *S. pombe* genome is essential (52). There are multiple factors that could contribute to these differences in essentiality. *S. pombe* has far fewer genes (n=5,134)(53) than *S. cerevisiae* (n=6,275)(54) and *Cryptococcus neoformans* (n=6,975)(37), suggesting that the elevated proportion of essential genes in *S. pombe* may be the result of differential conservation of nonessential and essential genes in a reduced genome. However, *Cryptococcus neoformans* also has a higher proportion of essential genes than *S. cerevisiae*, despite having more genes. This could be explained by the whole genome duplication in *S. cerevisiae*, which generated many redundancies in gene function (55,56). Alternately, the TN-seq approach may be more liberal in assigning essential status than traditional genetic approaches. Indeed, a TN-seq based approach in *C. albicans* suggests that 27.5% (1,610 of 5,893 genes) of the genome is essential (26). Because of the competitive growth in pooled settings, mutants that are sufficiently sick are difficult to sample, despite being viable, resulting in overestimation of the number of essential genes.

Practically, scoring essentiality as a binary phenotype, where mutants either grow or do not, may not be a totally meaningful distinction, particularly in the context of antifungal drug development. Severely impairing pathogen growth may not be dramatically different from completely impairing pathogen growth in terms of clinical outcome. Further, TN-seq may also score genes required for various growth stages as essential. Our mutant library preparation method required growth on synthetic defined media containing galactose and lacking uracil.

Genes required for growth in these conditions may score as essential using TN-seq but not in traditional approaches. Indeed, we showed that some discrepancies between the existing deletion collection and our TN-seq approach were explained by growth defects under these conditions. Essentiality is frequently context dependent, with a gene's essentiality being dependent both on the environment (eg. plates versus mice versus human hosts) and on the genetic background of the pathogen. Future experiments exploring variation in essentiality in animal models of infection will likely be informative.

Fluconazole resistance and tolerance are highly multigenic

Our TN-seq approach revealed numerous modifiers of fluconazole susceptibility and resistance. One of the strongest phenotypes we observed in our fluconazole TN-seq assay was for sensitivity to fluconazole conferred by loss of components of the *RIM101* pathway. This result was confirmed by analysis of individual mutants. The *RIM101* pathway is responsible for pH sensing and cell wall remodeling in response to elevated pH, as typically occurs upon entry into mammalian lungs. He we find that the *RIM101* pathway function promotes resistance to fluconazole. This is not unprecedented as *RIM101* pathway components are also required for fluconazole tolerance in *Candida albicans* (57). However, the Rim pathway in *Cryptococcus neoformans* is substantially diverged from that found in the Ascomycete lineage, which includes *C. albicans*. In Ascomycetes, pH sensing is carried out by the Rim21 pH sensor acting in conjunction with the arrestin-like Rim8 and the Rim9 chaperone. None of these three proteins are present in Basidiomycetes and are instead replaced by a different upstream pH sensing module including Rra1 and Nap1 (49,58). Given this dramatic upstream reprogramming of the Rim pathway in *Cryptococcus* relative to the ascomycete *C. albicans*, it is interesting that a similar phenotype is present. In *S. cerevisiae*, the Rim21 sensor appears to respond to imbalances

in plasma membrane lipids (59). While Rra1 is not a homolog of Rim21, perhaps it is detecting and responding to the same type of plasma membrane stress in *C. neoformans*, including potentially that triggered by fluconazole, which impairs the production of the membrane sterol ergosterol. In fact, loss of the Cdc50 lipid flippase changes how Rim pathway activation occurs in *Cryptococcus*, further supporting this model (60). *cdc50*Δ mutants have increased fluconazole susceptibility (61) and our assay confirmed this, albeit with a weak p-value resulting from very few coding inserts in *CDC50* (p= 0.035, n=4).

We also saw increased fluconazole resistance when transposon insertions were found upstream of genes involved in the electron transport chain and mitochondrial metabolism. This is consistent with past reports that rotenone, an inhibitor of the electron transport chain complex I, is very strongly antagonistic with fluconazole (50). It is not immediately clear why downregulating electron transport chain function would make cells grow better in the presence of fluconazole. One possible explanation is that slowing growth may allow cells without complete inhibition of *ERG11* to produce enough ergosterol to prevent depletion from the membrane, thus allowing slower but more successful growth to continue (62).

High throughput genomics

TN-seq is one of several modern approaches that can be used to quickly assay gene function on a large scale. Broadly, these approaches fall into either insertional mutagenesis approaches (such as TN-seq) or CRISPR-based approaches. Many of the CRISPR approaches operate similarly to TN-seq except that they use a library of guide RNAs and a Cas9 to generate a single break and mutation per cell. The template sequence expressing the guide RNA can then be sequenced to determine which gene was disrupted. CRISPR approaches typically require synthesis of guide

RNA libraries, which can be expensive, but has advantages in customization relative to TN-seq approaches. As well, in the short term, TN-seq is typically cheaper because of the lower startup costs, but in the long run CRISPR will typically become less expensive because of lower sequencing and library preparation costs for more targeted libraries with less complicated library preparation.

TN-seq does have significant advantages in portability. Our approach only requires the ability to introduce and mobilize a marked transposon. In contrast with CRISPR-based approaches, transposon mutant libraries do not require the synthesis of expensive whole genome guide RNA libraries that may need to be modified for every additional species or even isolate in species with particularly high sequence divergence. Our TN-seq approach could be readily applied to additional isolates or species across the *Cryptococcus* pathogenic species complex. Variation in essential genes is common across various organisms, including fungi (24,63). This variation also extends to drug resistance phenotypes as well (64). Because this type of variation can be driven both by sequence polymorphism and by gene content polymorphism, future exploration of strain variation in *Cryptococcus* and other fungi would likely benefit from a pangenome oriented approach, as is becoming common in bacteria (63). Indeed, initial studies in fungi like *Aspergillus fumigatus* have revealed enormous variation in gene content between individual isolates that TN-seq approaches would be extremely powerful in exploring (65,66).

Assaying essential gene function

Because deletion alleles of essential genes are inviable, screens using traditional deletion collections are unable to assay the contribution of essential genes to a given phenotype. This has long been a challenge for high throughput genetic assays and a number of solutions have been

developed to allow assays of essential genes. These solutions generally rely on partial loss of function and include the construction of collections of repressible alleles (27,67), hypermorphic alleles including the DAmP (decreased abundance by mRNA perturbation) libraries (68), RNAi libraries (69,70), and CRISPRi libraries (71,72). All of these approaches require either laborious one by one construction of targeted alleles (repressible promoters, DAmP alleles) or expensive synthesis of whole genome targeting libraries (RNAi, CRISPRi). In contrast, our TN-seq approach assayed 5' modifiers across the entire genome in a single experiment that required no modifications to the typical TN-seq assay. We were able to identify modifiers of both the known essential gene and target of fluconazole *ERG11* as well as recapitulate previously described fluconazole interactions that were unveiled using chemical screens. TN-seq in combination with chemical inhibitors can help to reveal mechanism of action and predict potential chemical interactions to help guide therapeutic development. While our TN-seq assay was conducted in a pool, it can also provide guidance to construct individual mutants, as we did to generate the fluconazole sensitive *erg11* hypomorph (Figure 5D). Alternately, the pools can be cloned out and decoded using approaches like knockout sudoku to efficiently construct libraries of hypomorph alleles if desired (73). As described above, this approach also benefits from portability, allowed TN-seq to potentially be applied to assay the function of essential genes across multiple isolates or species.

Material and Methods

Media

We routinely grew *Cryptococcus neoformans* strains on YPD media at 30°C. We made YPD using 20 g/L Bacto Agar, 20 g/L Bacto Peptone, 10 g/L Bacto Yeast Extract and 2% glucose. We

made YPD liquid the same way but omitting agar. We supplemented media with G418 (NEO) at 200 ug/mL and Nourseothricin (NAT) at 100 ug/mL. We made SC-URA media using 6.7 g/L of yeast nitrogen base without amino acids, 20 g/L Bacto Agar, 2% glucose and 1.92 g/L SC-URA dropout powder mix. For raffinose or galactose media, we substituted the 2% glucose with 2% raffinose or 2% galactose. Strains were cryopreserved in 20% glycerol.

We routinely grew bacterial stocks at 37°C. We made LB plates with 5 g/L Bacto Yeast Extract, 10 g/L Bacto tryptone, 10 g/L sodium chloride, and 15 g/L bacto agar. We supplemented media with 100 mg/L ampicillin or 25 mg/L kanamycin. Strains were cryopreserved in 20% glycerol.

Strain Construction

We generated strains using an electroporation protocol modified from (74). We grew strains to saturation overnight at 30°C in YPD and diluted back to an OD_{600} of 0.2 in 100 mL of fresh YPD liquid. We then grew with shaking at 30°C until cultures reached an OD_{600} between 0.6 and 1.0. We then centrifuged cells to pellet them and washed them twice with ice-cold water. We resuspended them in 10 mL of electroporation buffer (10 mL Tris-HCl (pH 7.5), 1 mM $MgCl_2$, 270 mM sucrose, 1 mM DTT). We incubated cells on ice for one hour. We then pelleted cells again and resuspended them in 250 uL of EB. We mixed 45 uL of cell suspension with 5 uL of transformation DNA in a pre-cooled 2 mm gap electroporation cuvette. We then transformed using a BioRad Gene Pulse with settings of 0.45 kV and 125 uF. Immediately after electroporation, we added 2 mL of liquid YPD and transferred cells to a round bottom culture tube. We let cells recover overnight at room temperature and then pelleted cells to plate on selective media. After colonies were visible, we streak purified colonies on selective media before preparing DNA to PCR validate transformants.

We generated our TN-seq library strain (yBB119) in two steps. We first commercially synthesized (IDT) a dS minitransposon (75) containing a NEO resistance marker (35) (pBB51). We then used overlap PCR to add homology to the up and downstream regions surrounding an intron of *URA5*. We performed a first round PCR to build upstream and downstream regions with homology to a product containing the transposon (oBB229 & oBB283, 5'; oBB285 & oBB286, transposon; oBB284 & oBB235, 3'). We then used a second round PCR to assemble the pieces (oBB229 & oBB235). We gel extracted the amplified band and topo blunt cloned to make plasmid pBB65. We then digested plasmid pBB65 with KpnI and NotI to release the transposon with targeting homology and used that as template for a transformation into H99 *C. neoformans* to produce strain yBB115. This strain was selected on G418 to select for transposon integration. G418 resistant colonies were checked for growth on SC-URA to detect integrations in the correct location.

We also commercially synthesized (IDT) a codon optimized version of the hyperactive AcTPase4x (Ac) transposase (pBB52) (76). We codon optimized the Ac transposase using the Optimizer webtool and the "random" method which assigns codon usage probabilistically across the gene sequence to roughly match the codon usage of the genome (34). We added a GAL7 promoter using overlap PCR. In our first round PCR we amplified the GAL7 promoter (oBB183 & oBB290) and the Ac transposase (oBB271 & oBB272). We then assembled these pieces in a second round PCR (oBB183 & oBB272). We gel extracted the amplified band and topo blunt cloned this product to make plasmid pBB63. We then subcloned the gal driven Ac into plasmid pSDMA25 (35) by digesting both vectors with XhoI. We additionally digested pBB63 with XmaI to digest the vector backbone to produce bands of a different size from the released insert. We called the resulting plasmid pBB68. We digested this plasmid with PacI to linearize it and

transformed it into yBB115. We selected transformants on YPD+NAT plates. We verified transposase activity by checking transformants for the ability to produce colonies on SC-Ura plates after overnight growth in YPGalactose liquid media and conducted a stability assay to check for stable integration. A successful strain was frozen down as yBB119.

To generate specific transposon inserts at the 5' end of *ERG11*, we used overlap PCR to add homology arms to the dS transposon. To generate 5' inserts that disrupted *ERG11* function (*erg11-77*), we amplified the upstream region from H99 genomic DNA using primers oBB583 and oBB595, the downstream from H99 genomic DNA using primers oBB598 and oBB588, and the transposon from pBB51 using oBB596 and oBB597. We then stitched together the pieces in a second PCR using primers oBB589 and oBB590. We excised the amplified band from a gel, blunt cloned it with a topo blunt kit and sequence verified as pBB102. To generate inserts that did not disrupt *ERG11* function (*erg11-296*) we amplified the upstream region from H99 genomic DNA using primers oBB583 and oBB591, the downstream from H99 genomic DNA using primers oBB594 and oBB588, and the transposon from pBB51 using oBB592 and oBB593. We then stitched the pieces together in a second PCR using primers oBB589 and oBB590. We excised the amplified band from a gel, blunt cloned it with a topo blunt kit and sequence verified it as pBB100. We digested each plasmid with XhoI to release the insert and transformed into H99. We verified transformants using stability tests and PCR validation for junction PCRs (oBB212 & oBB583, 5' junction; oBB171 & oBB588, 3' junction). We called the resulting strains yBB163 and yBB164 (pBB100, *erg11-296*) and yBB165 and yBB281 (pBB102, *erg11-77*).

Strain, plasmid, and primer lists are available as Supplemental Tables 3-5.

Mutant library construction

We constructed our TN-seq mutant library from strain yBB119, which was built as described above. We began library construction by subculturing yBB119 from the freezer onto YPD+5FOA plates. We then started 5 x 10 mL cultures of yBB119 in YPD + 5FOA liquid media from independent colonies. We shook these cultures overnight at 30°C. We then centrifuged the cultures, decanted the supernatant, and resuspended the cell pellet in 20 mL of YP+raffinose. We then shook them overnight again at 30°C. The next day we spun down, decanted, resuspended in water, spun down a second time, decanted, and finally resuspended in 100 mL of YPGal. We then plated 500 uL of culture to 160 SC-URA+GAL plates. With 5 replicate cultures this totaled 800 plates. We incubated the plates for 4 (3 sets of plates) or 5 days (2 sets of plates) before scraping the colonies off of the plates using water and a cell spreader. We then spun them down, washed once with water and then resuspended in 500 mL of water. We then diluted that culture 1:10 into SC-ura media, which we shook overnight at 30C. We froze 1.5 mL aliquots from these cultures in 20% glycerol.

We then revived all five cultures from 1.5 mL aliquots in 50 mL of YPD+G418. We allowed them to grow to saturation with shaking at 30°C and then mixed all five cultures to form a pool. We froze 1.5 mL aliquots of this library in 20% glycerol in cryovials for future experiments.

TN-seq sequencing library preparation

To prepare DNA for sequencing, we used a modified version of the Qiagen Genomic-Tip Yeast DNA protocol. We followed all standard steps except that we extended both the lyticase and proteinase K treatments to 24 hours on a shaking incubator. Prior to applying DNA to the genomic-tip columns, we transferred the digested cells to a 50 mL tube containing lysing matrix

Y (MP Biomedicals). We then freeze cracked twice in liquid nitrogen (alternating between liquid nitrogen and a 42°C water bath). We then bead beat the cells using a Fastprep big prep adapter with 40 seconds of shaking at 4.0 m/sec with 5 minutes pause in between. We then pelleted the cells and applied the supernatant to the Qiagen genomic tip column.

We prepared TN-seq libraries as previously described in *S. pombe*, with some modifications (22). We digested purified DNA using either MseI (10,000 U/mL) or CviAII (10,000 U/mL). Our digests used 10 ug of DNA in 500 uL of volume overnight at 37°C in CutSmart buffer, including 15 uL of each enzyme. We then used SPRIselect beads to clean and size select the digested DNA. We washed with 0.5 volumes of beads, pelleted on a magnet, removed the supernatant and added 0.2 volumes of beads before precipitating again. We then washed and eluted the DNA with 500 uL of water. We then added linkers with unique random barcodes via end ligation (88 uL of 10x ligation buffer, 143.5 uL of water, 153 uL of annealed barcoded linker, 5 uL of T4 DNA ligase, and 490 uL of the cleaned and size-selected DNA). We made the linkers from primers oBB912 (MseI) or oBB913 (CviAII) combined with primer oBB914. We mixed the primers at a concentration of 10 uM each in 1x HF PCR buffer and denatured at 95°C for 1 minute, followed by cycles of 10°C decreases in temperature for 7 minutes until a final cycle at 20°C.

After end ligation, we split the DNA from each digest into two pools. We used PCR to amplify these pools separately with 4 different dS-specific oligos (oBB915, oBB916, oBB917, oBB918), intended to create diversity in starting location. We thus performed 12 PCR reactions per digest per pool, except for oBB918, for which we performed only 9 PCR reactions per digest. In addition, we ran 3 linker-only controls per pool. In sum, each digest thus produced two pools of 45 reactions, or 90 reactions per digest. Each PCR reaction contained 24.5 uL water, 10 uL 5x

HF buffer, 1 uL dNTPs (10 mM), 8 uL linker ligated DNA, 1 uL linker oligo oBB919 (10 uM), 5 uL dS-specific oligo (2uM) or water, and 0.5 uL Phusion polymerase (2,000 U/mL). We ran this PCR using 94°C for 1 minutes, 6 cycles of (94°C for 15 seconds, 65°C for 30 seconds, 72°C for 30 seconds), 24 cycles of (94°C for 15 seconds, 60°C for 30 seconds, 72°C for 30 seconds), and finally 72°C for 10 minutes.

We then combined the reactions with shared primers from each pool (12 or 9 reactions) to form a subpool. We ran a subset of these subpools on a gel for validation. Successful amplification of a TN-seq library produces a smear of DNA rather than discrete bands. Defined bands suggest low complexity libraries, potentially resulting from jackpot events.

We then added the remaining Illumina adapters and barcodes with a second round of PCR on the subpools. This second reaction contained 28.5 µL water, 10 µL 5x HF buffer, 1 µL dNTPs (10 mM), 2.5 µL oBB920 (10 µM), 2.5µL barcode oligo (oBB921-oBB932; 10µM), 5 µL of insert PCR pool, 0.5 µL Phusion polymerase (2,000 U/mL) and was incubated at 94°C for 2 min, 5x (94°C for 30 sec, 54°C for 30 sec, 72°C for 40 sec). We gave each individual pool a unique barcode (1 barcode per enzyme digest mix). For each subpool, we ran 4 PCRs or 3 for those derived from oBB918. We then pooled reactions these reactions and cleaned the DNA using SPRIselect beads. We used 600 uL of pooled PCR product from the reactions above and added 0.75 volumed of SPRI beads. We precipitated the beads on a magnet, removed the supernatant and washed the pellet with 85% ethanol. We then eluted with 200 uL of TE. We repeated this wash a second time.

Finally, we confirmed this library by using a PCR with oligos that matched the end of the Illumina adapters (oBB933 and oBB934). This reaction was 15.75 µL water, 5 µL 5x HF buffer,

0.5 µL dNTPs (10 mM), 1.25 µL oBB933 (10 µM), 1.25 µL oBB934 (10 µM), 1 µL library DNA, 0.25 µL Phusion polymerase (2,000 U/mL). We ran this PCR at 98°C for 30 sec, 35x (98°C for 10 sec, 62°C for 30 sec, 72°C for 40 sec), 72°C for 10 min. We then quantified the library using a Qubit and determined fragment sizes using a Bioanalyzer. Sequencing was performed on a NextSeq instrument at the Stowers Molecular Biology Core. Raw sequencing reads are available at PRJNA1134100 on the Sequence Read Archive.

TN-seq Fluconazole Assay

To perform our fluconazole assay, we revived a 1.5 mL aliquot of our TN-seq library in 99 mL of YPD. After it grew to saturation, we inoculated two 100 mL YPD cultures at OD_{600} = 0.1. We then saved the remaining culture to prepare DNA for a T_0 timepoint. We added 0.1 mL of DMSO to one culture and 0.1 mL of DMSO containing 13.5 mg/mL of fluconazole to the other. We grew these cultures with shaking at 30 C overnight. The next day we diluted back to OD_{600} = 0.1 in 100 mLs of fresh YPD with 0.1% DMSO and with or without fluconazole. We grew to saturation overnight with shaking at 30°C again. The next day we collected cell pellets via centrifugation to prepare DNA.

Data analysis

We processed data by first identifying reads with exact matches to the dS transposon sequence with a custom perl script (available at https://github.com/bbillmyre/Crypto_TNseq). We trimmed dS and linker sequences, saving the barcode sequences trimmed from the linker. We then mapped the reads to the H99 genome (v44 from FungiDB (41,42)) using bwa mem (77). We used samtools (78), picard, and bedtools (79) for further processing before identifying the start of each transposon derived read and outputting a .bed file using a custom perl script (available at

https://github.com/bbillmyre/Crypto_TNseq). We reduced inserts that shared a common linker barcode to a single representative. We then used R to combine bed files by site from the various runs (T0, DMSO, and IC50 fluconazole). We converted insert counts to insertion frequencies and set 0 insert sites (where there were inserts at T0) from the DMSO and IC50 fluconazole libraries to the minimum frequency >0 otherwise detected in that library. Resetting 0 values to minimum frequency artificially reduces the size of the phenotypes detected but allows log adjustment of our data. We also removed any sites with less than 6 inserts in the original T0 pool to reduce potential sampling error in the fluconazole experiments. We then labeled each insert site either with the gene body it was part of or as intergenic if not between an ATG and STOP codon of a gene.

For our machine learning model, we then used custom R scripts to calculate parameters describing the landscape of transposon insertions across each gene (available at https://github.com/bbillmyre/Crypto_TNseq). For each gene we determined the following: Chromosome, Gene Length, Total Number of inserts, Total number of inserts divided by gene length, sum of insert frequency, sum of insert frequency divided by gene length, largest insert free gap in bases, largest insert free gap as percentage of gene length, sum of insert frequency in the middle 80% of the gene body, sum of insert frequency in the middle 80% of the gene body divided by length, and sum of insert frequency over the surrounding 100 kb sequence surrounding the gene. We obtained a set of gene essentiality and orthology predictions for *C. albicans*, *S. cerevisiae*, and *S. pombe* from a previous study (26). We used FungiDB to find orthologs in *C. neoformans* of the *C. albicans* genes in this dataset. We then selected a set of genes with orthologs in all four species that shared essentiality status in *C. albicans*, *S. cerevisiae*, and *S. pombe*. We used this set as a training set for a machine learning model. Using

R, we randomly split it into 80% training data and 20% validation data (random seed=100). We then trained using the "randomForest" package with parameters (mtry=5, ntree=10000, importance =TRUE)(80). We repeated this 100 times, saving importance parameters and precision/recall estimates from the validation set for each run. We also used each model to predict the genome each time. We then used a Student's t-test to test whether the distribution of our essentiality scores for each gene were different from the essentiality threshold of 0.5 ($p<0.05$ after a Bonferroni correction).

Upon reviewing the data, we discovered that genes in the training set were enriched for poor p values. We manually reviewed transposon insertions in many of these genes, which suggested that the model was struggling to predict essentiality for these genes because the TN-seq data disagreed with the conservation-based assumptions that were part of the training set. We dropped all of the genes with $p>0.05$ that were part of the original training set and repeated our model construction as above to produce our V2 model. As before, we predicted essentiality from this model 100 times.

For the fluconazole experiments we used R to perform statistical tests. We calculated a $\log_{10}$ fold change for IC50 to DMSO growth (mathematically the same as the ratio of IC50/T0 to DMSO/T0). We used a Mann-Whitney U test to compare the distribution of fold-changes in a gene region to that of the intergenic inserts. We also performed multiple test correction using a Bonferroni correction. For 5' and 3' modifiers we performed the statistical tests in exactly the same manner except that we compared the distributions of fold changes of inserts within the 300 bp immediately upstream of the start codon (3') or downstream of the stop codon (5') to the intergenic inserts. We again performed a Bonferroni correction.

Competition Assay

To conduct competition assays, we grew cultures overnight in YPD at 30°C. We then added enough KN99 and enough competitor strain to reach an OD_{600} of 0.05 of each strain to 5 mL of SC-URA+GAL broth. We plated a concentration series of these initial mixed cultures to YPD plates. We then grew the cultures overnight with shaking at 30°C. The next day we plated a concentration series to YPD plates. After colonies grew from each concentration series, we replica plated to YPD+NEO and to new YPD plates. We counted colonies the next day and recorded the number of colonies that grew on YPD+NEO (just marked strain) versus YPD (both strains).

Spot Dilution Assays

To perform spot dilution assays, we grew cultures overnight in YPD at 30°C until they reached saturation. We then measured the OD_{600} using a spectrophotometer. We diluted cells to an OD_{600} of 20. We then performed 10-fold serial dilutions and plated 5 uL spots of each dilution to the appropriate media. Plates were incubated at 30°C until we took pictures as detailed in each figure.

Growth Curves

To measure growth curves we grew cultures overnight in YPD at 30°C until they reached saturation. We diluted cells to an OD_{600} of 0.00166 in fresh YPD and added 150 uL to each well of a 96 well plate. We then made a drug stock containing (for each sample) 50 uL of YPD with 0.2 uL of DMSO containing 1000X fluconazole. This stock was thus 0.4% DMSO and 4X fluconazole for each condition. We then added 50 uL of drug stock YPD mix to each well to

achieve a final mix in each well of cells at an OD_{600} of 0.00125, DMSO at 0.1%, and fluconazole (if present) at 1X concentration. We grew the plates on a TECAN Infinite M200 Pro at 30°C with shaking. We ran 8 technical replicates for each sample and report the OD_{600} at 24 hours normalized to the no drug control at 24 hours.

Data Availability

Sequences have been deposited into the Sequence Read Archive under project accession no. PRJNA1134100. Original data underlying this manuscript can be accessed from the Stowers Original Data Repository at http://www.stowers.org/research/publications/libpb-2480

Acknowledgements

We thank members of the Billmyre and Zanders labs for helpful comments on the manuscript. We thank the Heitman and Alspaugh labs for supplying strains. We thank both the University of Georgia and the Stowers Institute for Medical Research for financial support of this project. We also acknowledge support from NIH grants DP2GM132936 and R35GM151982 to SEZ. We thank the Madhani Laboratory for prepublication access to the *C. neoformans* knockout collection and acknowledge the NIH funding that has helped fund its construction (R01AI100272).

Figure Legends

Figure 1. TN-seq in *C. neoformans.* A) Transposon insertions (orange arrow) into nonessential genes results in viable cells. In contrast, insertions into essential genes will result in dead and nonrecoverable cells. B) TN-seq works by generating a library of cells where each cell has a single independent transposon insertion in a random location. As in A, those insertions into essential regions cause the cells to die and are nonrecoverable. As a result, the total library (bottom) is depleted in insertions in essential regions. C) The Ac/Ds transposon was split into an Ac transposase and a Ds transposon containing a neomycin resistance marker. This Ds transposon was integrated into an intron of *URA5* and the Ac transposase was integrated into the safe haven locus. The resulting stain is *ura-* and neomycin resistant. Upon initiating transposition via growth on galactose, the strain becomes *URA+* and mutant at another locus (depicted here as *YFG1*).

Figure 2. TN-seq enabled prediction of gene essentiality. A) 176 unique transposon insertions (orange vertical lines) are plotted along a region of Chromosome 1 centered on the known essential gene *ERG11* and showing two flanking predicted nonessential genes. For nine of the displayed sites, we recovered transposon insertions in both orientations. B) Flow chart depicting the random forest approach to classifying gene essentiality. C) Schematic illustrating parameters that describe each gene within the TN-seq data for machine learning. D) Precision recall curve describing tradeoff between precision and recall for both of the random forest models. Each point is the mean of 100 replicates where the training data was randomly split into training and validation sets. The threshold was then varied by 0.01 from 0.01 to 0.99 for each set. E) The importance of each feature for the V2 model is plotted. Each importance value is also calculated

from the same 100 replicates of the training data. Error bars indicate standard deviation. F) Histogram of the essentiality prediction score for the entire gene set of *C. neoformans* based on the mean of 100 replicates.

Figure 3. Evolution and conservation of essentiality. A) Histogram of essentiality score as in Figure 2F, split by conservation status in humans. Genes without human orthologs are shown at top and those with human orthologs are shown below. B) Genes with orthologs in *S. cerevisiae*, *C. albicans*, *S. pombe* and *C. neoformans* where the orthologs are essential in all three ascomycetes but predicted dispensable in *C. neoformans*. C) Genes with orthologs in *S. cerevisiae*, *C. albicans*, *S. pombe* and *C. neoformans* where the orthologs are dispensable in all three ascomycetes but predicted essential in *C. neoformans*.

Figure 4. Using TN-seq to assay genetic contribution to fluconazole resistance. A) TN-seq libraries were selected with IC50 levels of fluconazole dissolved in DMSO or with just the equivalent amount of DMSO as a control. Libraries were sequenced at time 0, and after two days growth in DMSO or fluconazole to identify genes with differential transposon insertion frequencies after selection. B) Volcano plot of 5296 genes with 5 or more insert sites (out of 6975 *Cryptococcus neoformans* genes) displaying the mean $\log_{10}$ (Fluconazole/DMSO) value on the x-axis and the $-\log_{10}$ (Bonferroni corrected p-value) on the y-axis. Individual genes are shaded orange if the distribution of inserts is statistically different from the distribution of inserts into noncoding regions ($p<0.05$ via Mann-Whitney U test after Bonferroni correction). Genes that are not statistically different are shaded blue. C) Boxplot displaying distribution of $\log_{10}$-adjusted fold changes in insert density (ie. frequency in fluconazole/frequency in DMSO). Boxplots show first quartile, median, third quartile. The whiskers show the range to a maximum of 1.5 times the interquartile range above and below the first and third quartile, respectively.

Outlier data points (outside the whiskers) are not displayed. Not displaying outliers results in 15,902 of 454,341 intergenic sites, 1 of 107 sites from *afr1*, 1 of 4 sites from *nap1*, 1 of 34 sites from *rim23*, 2 of 127 sites from *rra1*, 12 of 74 sites from *rim20*, 0 of 5 sites from *vps25*, 5 of 77 sites from *rim101*, 5 of 151 sites from *rim 13*, and 1 of 11 sites from *vps23* not being displayed although those data were considered in the statistical analyses. Snf7 is not shown because there were 0 inserts between the start and stop codons. Inserts in intergenic regions are indicated in grey, genes where inserts were significantly depleted after fluconazole treatment are shown in orange (*afr1, rim23, rra1, rim20, rim101, rim13, vps23*) and genes where inserts were not significantly depleted after fluconazole are shown in blue (*nap1 and vps25*). Notably, both *nap1* and *vps25* had very low numbers of inserts that limited statistical power. D) Spot dilution assays with 5 μL spots plated. The initial leftmost spot is of OD_{600} = 20 culture and each successive spot is a 10-fold dilution, so that the final spot should be 10^7 less concentrated than the first. Both plates were spotted on the same day with the same dilution series. YPD plates were imaged after 48 hours at 30°C and fluconazole plates were imaged after 72 hours at 30°C. Mutants were spotted on two separate plates for YPD and fluconazole media, each with a wildtype H99 control present.

Figure 5. Regulatory inserts enable assays of essential gene function. A) Schematic of model for 5' transposon insertions. Wildtype cells (top row) should grow well on YPD and be moderately impaired (~50%) by an IC50 level of fluconazole. Cells with a transposon insertion in the 5' regulatory region of *ERG11* should grow well on YPD but be highly sensitive to IC50 levels of fluconazole. B) Boxplot displaying distribution of $\log_{10}$-adjusted fold changes in insert density (ie. frequency in fluconazole/frequency in DMSO). Each column shows inserts only within a 300 base pair region either immediately upstream of the start codon or downstream of

the stop codon. Boxplots show first quartile, median, third quartile. The whiskers show the range to a maximum of 1.5 times the interquartile range above and below the first and third quartile, respectively. Outliers are displayed as individual datapoints. C) Volcano plot of 1,251 predicted essential genes with 5 or more insert sites in the 300 base pairs upstream of the start codon (out of 6975 *Cryptococcus neoformans* genes) displaying the mean $\log_{10}$ (Fluconazole/DMSO) value on the x-axis and the $-\log_{10}$ (Bonferroni corrected p-value) on the y-axis. Individual genes are shaded orange if the distribution of inserts is statistically different from the distribution of inserts into noncoding regions genome-wide ($p<0.05$ via Mann-Whitney U test after Bonferroni correction). Genes that are not statistically different are shaded blue. D) Spot dilution assays with 5 μL spots plated. The initial leftmost spot is of $OD_{600} = 20$ culture and each successive spot is a 10-fold dilution, so that the final spot should be 10^5 less concentrated than the first. Both plates were spotted on the same day with the same dilution series. YPD plates were imaged after 48 hours at 30°C and fluconazole plates were imaged after 72 hours at 30°C.

Supplemental Figure 1. Transposon insertions are biased regionally and by selection. A) Transposon insertion count is plotted across chromosomes. There are clear peaks at the rDNA array on Chromosome 2 and the *URA5* locus on Chromosome 8, as well as the surrounding chromosome. B) Transposon insertion count is plotted across the 2000 bases immediately upstream and downstream of the start codon. Rows are split based on predicted gene essentiality. C) Insertions per base are plotted for genes not present vs present in the deletion collection. Boxplots show first quartile, median, third quartile. The whiskers show the range to a maximum of 1.5 times the interquartile range above and below the first and third quartile, respectively. Outliers are displayed as individual datapoints.

Supplemental Figure 2. PCR validation of predicted essential genes with strains in deletion collection. A) Schematic of PCR validation strategy for deletions. WT alleles should be detectable with an in-gene PCR that amplifies wildtype DNA sequence (left). Mutant alleles should have the wildtype allele replaced with a drug resistance marker. The junctions between the genomic sequence and the resistance marker should be detectable via PCR on both the 5' and 3' ends (right). A successful mutant should produce bands from the 5' junction and 3' junction but not from the in-gene PCR. A wildtype strain should produce a band only from the in-gene PCR. B) PCR validations for 5 independent colonies from the deletion collection strain for CNAG_05292. All five produce wildtype in-gene bands and no junction bands. C) PCR validations for 5 independent colonies from the deletion collection strain for CNAG_06887. All five produce wildtype in-gene bands and no junction bands. D) PCR validations for deletion collection strains for CNAG_04763, CNAG_00996, CNAG_02190. All three strains produce negative in-gene PCRs and successful junction PCRs for both 5' and 3' ends.

Supplemental Figure 3. Some incorrectly predicted essential genes have growth defects in transposase inducing conditions. A) Spot dilution assays with 5 μL spots plated. The initial leftmost spot is of OD_{600} = 20 culture and each successive spot is a 10-fold dilution, so that the final spot should be 10^5 less concentrated than the first. All four plates were spotted on the same day with the same dilution series. B) Competition assay with percentage of mutant plotted on the y-axis. Each mutant was competed against the same unmarked wildtype KN99 parental strain. Strains were competed in the SC-URA+Gal media used in the original assay. Strains were originally mixed at a 50:50 ratio based on OD_{600}. Inconsistency with mixing of CNAG_00996 suggests an altered OD_{600} to CFU ratio. C) Picture of colonies on YPD plates from the competition assay for the *cnag_00996*Δ mutant. Colonies were distinctly different in appearance

and replica plating to YPD+NAT media confirmed that rough colonies were the mutant strain. Panel on right is zoomed in from inset box on left and is adjusted to help visualize difference between colonies more clearly.

Supplemental Figure 4. Concentration dependent inhibition of growth by fluconazole in YPD liquid. Growth at 24 hours plotted relative to a no drug control. X axis displays concentration of fluconazole added in DMSO. Y axis shows OD_{600} normalized to OD_{600} of no drug control.

Supplemental Figure 5. Shiny app allows visualization of data on publicly available website. Screenshot of a publicly available interactive Shiny app (https://bbillmyre.shinyapps.io/Crypto_TN_seq_viewer/) that visualizes data from the *Cryptococcus neoformans* TN-seq assay. There are four plots, displaying the distribution of fold changes within a gene compared with intergenic inserts (as in figure 4C,) the distribution of insert frequencies across a gene at three different experimental stages, transposon insertion frequencies across a gene with an additional 300 bases before the ATG and after the stop codon, and finally a volcano plot (as in figure 4B) with the current gene highlighted in black. The app only accepts *C. neoformans* systematic names (ie., CNAG_0####).

Supplemental Table 1. Table containing essentiality predictions and fluconazole response data.

Supplemental Table 2. Table containing GO predictions for predicted essential genes lacking predicted human orthologs

Supplemental Table 3. Strain table.

Supplemental Table 4. Plasmid table.

Supplemental Table 5. Primer table.

References

1. Denning DW. Global incidence and mortality of severe fungal disease. Lancet Infect Dis [Internet]. 2024 Jan 12 [cited 2024 Feb 8];0(0). Available from: https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(23)00692-8/fulltext

2. Craw JA, Beer L, Tie Y, Jaenicke T, Shouse RL, Prejean J. Viral Rebound Among Persons With Diagnosed HIV Who Achieved Viral Suppression, United States. JAIDS J Acquir Immune Defic Syndr. 2020 Jun 1;84(2):133.

3. Roemer T, Krysan DJ. Antifungal Drug Development: Challenges, Unmet Clinical Needs, and New Approaches. Cold Spring Harb Perspect Med. 2014 May 1;4(5):a019703.

4. Perfect JR. Cryptococcus neoformans: A sugar-coated killer with designer genes. FEMS Immunol Med Microbiol. 2005;45(3):395–404.

5. Hagen F, Khayhan K, Theelen B, Kolecka A, Polacheck I, Sionov E, et al. Recognition of seven species in the *Cryptococcus gattii*/*Cryptococcus neoformans* species complex. Fungal Genet Biol. 2015 May 1;78:16–48.

6. Jallow S, Govender NP. Ibrexafungerp: A First-in-Class Oral Triterpenoid Glucan Synthase Inhibitor. J Fungi. 2021 Feb 25;7(3):163.

7. Feldmesser M, Kress Y, Mednick A, Casadevall A. The Effect of the Echinocandin Analogue Caspofungin on Cell Wall Glucan Synthesis by Cryptococcus neoformans. J Infect Dis. 2000 Dec 1;182(6):1791–5.

8. Burks C, Darby A, Gómez Londoño L, Momany M, Brewer MT. Azole-resistant Aspergillus fumigatus in the environment: Identifying key reservoirs and hotspots of antifungal resistance. PLoS Pathog. 2021 Jul;17(7):e1009711.

9. Kang SE, Sumabat LG, Melie T, Mangum B, Momany M, Brewer MT. Evidence for the agricultural origin of resistance to multiple antimicrobials in Aspergillus fumigatus, a fungal pathogen of humans. G3 Bethesda Md. 2022 Feb 4;12(2):jkab427.

10. Denning DW, Bromley MJ. How to bolster the antifungal pipeline. Science. 2015 Mar 27;347(6229):1414–6.

11. Ianiri G, Idnurm A. Essential Gene Discovery in the Basidiomycete Cryptococcus neoformans for Antifungal Drug Target Prioritization. mBio. 2015 Mar 31;6(2):10.1128/mbio.02334-14.

12. Liu OW, Chun CD, Chow ED, Chen C, Madhani HD, Noble SM. Systematic genetic analysis of virulence in the human fungal pathogen Cryptococcus neoformans. Cell. 2008 Oct 3;135(1):174–88.

13. Lehner KR, Stone MM, Farber RA, Petes TD. Ninety-Six Haploid Yeast Strains With Individual Disruptions of Open Reading Frames Between YOR097C and YOR192C, Constructed for the Saccharomyces Genome Deletion Project, Have an Additional Mutation in the Mismatch Repair Gene MSH3. Genetics. 2007 Nov;177(3):1951–3.

14. Li Z, Kim KS. RELATe enables genome-scale engineering in fungal genomics. Sci Adv. 2020 Sep 18;6(38):eabb8783.

15. Goodman AL, McNulty NP, Zhao Y, Leip D, Mitra RD, Lozupone CA, et al. Identifying genetic determinants needed to establish a human gut symbiont in its habitat. Cell Host Microbe. 2009 Sep 17;6(3):279–89.

16. Gawronski JD, Wong SMS, Giannoukos G, Ward DV, Akerley BJ. Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for Haemophilus genes required in the lung. Proc Natl Acad Sci U S A. 2009 Sep 22;106(38):16422–7.

17. van Opijnen T, Bodi KL, Camilli A. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. Nat Methods. 2009 Oct;6(10):767–72.

18. Langridge GC, Phan MD, Turner DJ, Perkins TT, Parts L, Haase J, et al. Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. Genome Res. 2009 Dec;19(12):2308–16.

19. Guo Y, Park JM, Cui B, Humes E, Gangadharan S, Hung S, et al. Integration Profiling of Gene Function With Dense Maps of Transposon Integration. Genetics. 2013 Oct;195(2):599–609.

20. Lee SY, Hung S, Esnault C, Pathak R, Johnson KR, Bankole O, et al. Dense Transposon Integration Reveals Essential Cleavage and Polyadenylation Factors Promote Heterochromatin Formation. Cell Rep. 2020 Feb 25;30(8):2686-2698.e8.

21. Grech L, Jeffares DC, Sadée CY, Rodríguez-López M, Bitton DA, Hoti M, et al. Fitness Landscape of the Fission Yeast Genome. Mol Biol Evol. 2019 Aug 1;36(8):1612–23.

22. Billmyre RB, Eickbush MT, Craig CJ, Lange JJ, Wood C, Helston RM, et al. Genome-wide quantification of contributions to sexual fitness identifies genes required for spore viability and health in fission yeast. Cullen PJ, editor. PLOS Genet. 2022 Oct 27;18(10):e1010462.

23. Michel AH, Hatakeyama R, Kimmig P, Arter M, Peter M, Matos J, et al. Functional mapping of yeast genomes by saturated transposition. Deshaies RJ, editor. eLife. 2017 May 8;6:e23570.

24. Chen P, Michel AH, Zhang J. Transposon insertional mutagenesis of diverse yeast strains suggests coordinated gene essentiality polymorphisms. Nat Commun. 2022 Mar 21;13(1):1490.

25. Michel AH, Schie S van, Mosbach A, Scalliet G, Kornmann B. Exploiting homologous recombination increases SATAY efficiency for loss- and gain-of-function screening [Internet]. bioRxiv; 2019 [cited 2024 Jan 19]. p. 866483. Available from: https://www.biorxiv.org/content/10.1101/866483v1

26. Segal ES, Gritsenko V, Levitan A, Yadav B, Dror N, Steenwyk JL, et al. Gene Essentiality Analyzed by In Vivo Transposon Mutagenesis and Machine Learning in a Stable Haploid Isolate of Candida albicans. mBio. 2018 Oct 30;9(5):10.1128/mbio.02048-18.

27. Fu C, Zhang X, Veri AO, Iyer KR, Lash E, Xue A, et al. Leveraging machine learning essentiality predictions and chemogenomic interactions to identify antifungal targets. Nat Commun. 2021 Nov 11;12(1):6497.

28. Gale AN, Sakhawala RM, Levitan A, Sharan R, Berman J, Timp W, et al. Identification of Essential Genes and Fluconazole Susceptibility Genes in Candida glabrata by Profiling Hermes Transposon Insertions. G3 Bethesda Md. 2020 Oct 5;10(10):3859–70.

29. McClintock B. Chromosome Organization and Genic Expression. Cold Spring Harb Symp Quant Biol. 1951 Jan 1;16:13–47.

30. Wood V, Gwilliam R, Rajandream MA, Lyne M, Lyne R, Stewart A, et al. The genome sequence of Schizosaccharomyces pombe. Nature. 2002 Feb;415(6874):871–80.

31. Gatermann KB, Hoffmann A, Rosenberg GH, Käufer NF. Introduction of Functional Artificial Introns into the Naturally Intronless ura4 Gene of Schizosaccharomyces pombe. Mol Cell Biol. 1989 Apr 1;9(4):1526–35.

32. Janbon G. Introns in Cryptococcus. Mem Inst Oswaldo Cruz. 2018 Feb 19;113(7):e170519.

33. Dumesic PA, Natarajan P, Chen C, Drinnenberg IA, Schiller BJ, Thompson J, et al. Stalled Spliceosomes Are a Signal for RNAi-Mediated Genome Defense. Cell. 2013 Feb 28;152(5):957–68.

34. Puigbò P, Guzmán E, Romeu A, Garcia-Vallvé S. OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. Nucleic Acids Res. 2007 Jul;35(Web Server issue):W126–31.

35. Arras SDM, Chitty JL, Blake KL, Schulz BL, Fraser JA. A Genomic Safe Haven for Mutant Complementation in Cryptococcus neoformans. PLOS ONE. 2015 Apr 9;10(4):e0122916.

36. Cain AK, Barquist L, Goodman AL, Paulsen IT, Parkhill J, van Opijnen T. A decade of advances in transposon-insertion sequencing. Nat Rev Genet. 2020 Sep;21(9):526–40.

37. Basenko EY, Pulman JA, Shanmugasundram A, Harb OS, Crouch K, Starns D, et al. FungiDB: An Integrated Bioinformatic Resource for Fungi and Oomycetes. J Fungi. 2018 Mar 20;4(1):39.

38. Jung KW, Yang DH, Maeng S, Lee KT, So YS, Hong J, et al. Systematic functional profiling of transcription factor networks in Cryptococcus neoformans. Nat Commun. 2015 Apr 7;6(1):6757.

39. Lee KT, So YS, Yang DH, Jung KW, Choi J, Lee DG, et al. Systematic functional analysis of kinases in the fungal pathogen Cryptococcus neoformans. Nat Commun. 2016 Sep 28;7(1):12766.

40. Jin JH, Lee KT, Hong J, Lee D, Jang EH, Kim JY, et al. Genome-wide functional analysis of phosphatases in the pathogenic fungus Cryptococcus neoformans. Nat Commun. 2020 Aug 24;11(1):4212.

41. Janbon G, Ormerod KL, Paulet D, Iii EJB, Yadav V, Chatterjee G, et al. Analysis of the Genome and Transcriptome of Cryptococcus neoformans var. grubii Reveals Complex RNA Expression and Microevolution Leading to Virulence Attenuation. PLOS Genet. 2014 Apr 17;10(4):e1004261.

42. Stajich JE, Harris T, Brunk BP, Brestelli J, Fischer S, Harb OS, et al. FungiDB: an integrated functional genomics database for fungi. Nucleic Acids Res. 2012 Jan;40(Database issue):D675-681.

43. Hashida-Okado T, Ogawa A, Endo M, Yasumoto R, Takesako K, Kato I. AUR1, a novel gene conferring aureobasidin resistance on Saccharomyces cerevisiae: a study of defective morphologies in Aur1p-depleted cells. Mol Gen Genet MGG. 1996 May 23;251(2):236–44.

44. Hayles J, Wood V, Jeffery L, Hoe KL, Kim DU, Park HO, et al. A genome-wide resource of cell cycle and cell shape genes of fission yeast. Open Biol. 2013 May;3(5):130053.

45. Sideri T, Rallis C, Bitton DA, Lages BM, Suo F, Rodríguez-López M, et al. Parallel Profiling of Fission Yeast Deletion Mutants for Proliferation and for Lifespan During Long-Term Quiescence. G3 GenesGenomesGenetics. 2014 Dec 1;5(1):145–55.

46. Sanguinetti M, Posteraro B, La Sorda M, Torelli R, Fiori B, Santangelo R, et al. Role of AFR1, an ABC Transporter-Encoding Gene, in the In Vivo Response to Fluconazole and Virulence of Cryptococcus neoformans. Infect Immun. 2006 Feb;74(2):1352–9.

47. Billmyre RB, Applen Clancey S, Li LX, Doering TL, Heitman J. 5-fluorocytosine resistance is associated with hypermutation and alterations in capsule biosynthesis in Cryptococcus. Nat Commun. 2020 Jan 8;11(1):127.

48. Chang YC, Lamichhane AK, Cai H, Walter PJ, Bennett JE, Kwon-Chung KJ. Moderate levels of 5-fluorocytosine cause the emergence of high frequency resistance in cryptococci. Nat Commun. 2021 Jun 8;12(1):3418.

49. Pianalto KM, Ost KS, Brown HE, Alspaugh JA. Characterization of additional components of the environmental pH-sensing complex in the pathogenic fungus Cryptococcus neoformans. J Biol Chem. 2018 Jun 29;293(26):9995–10008.

50. Wambaugh MA, Denham ST, Ayala M, Brammer B, Stonhill MA, Brown JC. Synergistic and antagonistic drug interactions in the treatment of systemic fungal infections. eLife. 2020 May 5;9:e54160.

51. Giaever G, Chu AM, Ni L, Connelly C, Riles L, Véronneau S, et al. Functional profiling of the Saccharomyces cerevisiae genome. Nature. 2002 Jul;418(6896):387–91.

52. Kim DU, Hayles J, Kim D, Wood V, Park HO, Won M, et al. Analysis of a genome-wide set of gene deletions in the fission yeast Schizosaccharomyces pombe. Nat Biotechnol. 2010 Jun;28(6):617–23.

53. Rutherford KM, Lera-Ramírez M, Wood V. PomBase: a Global Core Biodata Resource—growth, collaboration, and sustainability. Genetics. 2024 Feb 20;iyae007.

54. Wong ED, Miyasato SR, Aleksander S, Karra K, Nash RS, Skrzypek MS, et al. Saccharomyces genome database update: server architecture, pan-genome nomenclature, and external resources. Genetics. 2023 May 4;224(1):iyac191.

55. Dietrich FS, Voegeli S, Brachat S, Lerch A, Gates K, Steiner S, et al. The Ashbya gossypii Genome as a Tool for Mapping the Ancient Saccharomyces cerevisiae Genome. Science. 2004 Apr 9;304(5668):304–7.

56. Wolfe KH, Shields DC. Molecular evidence for an ancient duplication of the entire yeast genome. Nature. 1997 Jun 12;387(6634):708–13.

57. Garnaud C, García-Oliver E, Wang Y, Maubon D, Bailly S, Despinasse Q, et al. The Rim Pathway Mediates Antifungal Tolerance in Candida albicans through Newly Identified Rim101 Transcriptional Targets, Including Hsp90 and Ipt1. Antimicrob Agents Chemother. 2018 Feb 23;62(3):e01785-17.

58. Ost KS, O'Meara TR, Huda N, Esher SK, Alspaugh JA. The Cryptococcus neoformans Alkaline Response Pathway: Identification of a Novel Rim Pathway Activator. PLoS Genet. 2015 Apr 10;11(4):e1005159.

59. Nishino K, Obara K, Kihara A. The C-terminal Cytosolic Region of Rim21 Senses Alterations in Plasma Membrane Lipid Composition: INSIGHTS INTO SENSING MECHANISMS FOR PLASMA MEMBRANE LIPID ASYMMETRY. J Biol Chem. 2015 Dec 25;290(52):30797–805.

60. Brown HE, Ost KS, Esher SK, Pianalto KM, Saelens JW, Guan Z, et al. Identifying a novel connection between the fungal plasma membrane and pH-sensing. Mol Microbiol. 2018;109(4):474–93.

61. Huang W, Liao G, Baker GM, Wang Y, Lau R, Paderu P, et al. Lipid Flippase Subunit Cdc50 Mediates Drug Resistance and Virulence in Cryptococcus neoformans. mBio. 2016 May 10;7(3):10.1128/mbio.00478-16.

62. Altamirano S, Simmons C, Kozubowski L. Colony and Single Cell Level Analysis of the Heterogeneous Response of Cryptococcus neoformans to Fluconazole. Front Cell Infect Microbiol [Internet]. 2018 Jun 19 [cited 2024 Apr 24];8. Available from: https://www.frontiersin.org/articles/10.3389/fcimb.2018.00203

63. Rosconi F, Rudmann E, Li J, Surujon D, Anthony J, Frank M, et al. A bacterial pan-genome makes gene essentiality strain-dependent and evolvable. Nat Microbiol. 2022;7(10):1580–92.

64. Opijnen T van, Dedrick S, Bento J. Strain Dependent Genetic Networks for Antibiotic-Sensitivity in a Bacterial Pathogen with a Large Pan-Genome. PLOS Pathog. 2016 Sep 8;12(9):e1005869.

65. Lofgren LA, Ross BS, Cramer RA, Stajich JE. The pan-genome of Aspergillus fumigatus provides a high-resolution view of its population structure revealing high levels of lineage-specific diversity driven by recombination. PLoS Biol. 2022 Nov;20(11):e3001890.

66. Barber AE, Sae-Ong T, Kang K, Seelbinder B, Li J, Walther G, et al. Aspergillus fumigatus pan-genome analysis identifies genetic variants associated with human infection. Nat Microbiol. 2021 Dec;6(12):1526–36.

67. Roemer T, Jiang B, Davison J, Ketela T, Veillette K, Breton A, et al. Large-scale essential gene identification in Candida albicans and applications to antifungal drug discovery. Mol Microbiol. 2003;50(1):167–81.

68. Breslow DK, Cameron DM, Collins SR, Schuldiner M, Stewart-Ornstein J, Newman HW, et al. A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. Nat Methods. 2008 Aug;5(8):711–8.

69. Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, Gotta M, et al. Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. Nature. 2003 Jan;421(6920):231–7.

70. Dietzl G, Chen D, Schnorrer F, Su KC, Barinova Y, Fellner M, et al. A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. Nature. 2007 Jul;448(7150):151–6.

71. Zhang YM, Zheng L, Xie K. CRISPR/dCas9-Mediated Gene Silencing in Two Plant Fungal Pathogens. mSphere. 2023 Feb 21;8(1):e0059422.

72. Billerbeck S, Prins RC, Marquardt M. A Modular Cloning Toolkit Including CRISPRi for the Engineering of the Human Fungal Pathogen and Biotechnology Host Candida glabrata. ACS Synth Biol. 2023 Apr 21;12(4):1358–63.

73. Anzai IA, Shaket L, Adesina O, Baym M, Barstow B. Rapid curation of gene disruption collections using Knockout Sudoku. Nat Protoc. 2017 Oct;12(10):2110–37.

74. Fan Y, Lin X. Multiple Applications of a Transient CRISPR-Cas9 Coupled with Electroporation (TRACE) System in the Cryptococcus neoformans Species Complex. Genetics. 2018 Apr;208(4):1357–72.

75. Weil CF, Kunze R. Transposition of maize Ac/Ds transposable elements in the yeast Saccharomyces cerevisiae. Nat Genet. 2000 Oct;26(2):187–90.

76. Lazarow K, Du ML, Weimer R, Kunze R. A Hyperactive Transposase of the Maize Transposable Element Activator (Ac). Genetics. 2012 Jul;191(3):747–56.

77. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009 Jul 15;25(14):1754–60.

78. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinforma Oxf Engl. 2009 Aug 15;25(16):2078–9.

79. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010 Mar 15;26(6):841–2.

80. Liaw A, Wiener M. Classification and Regression by randomForest. R News. 2(3):18–22.

Figure 1

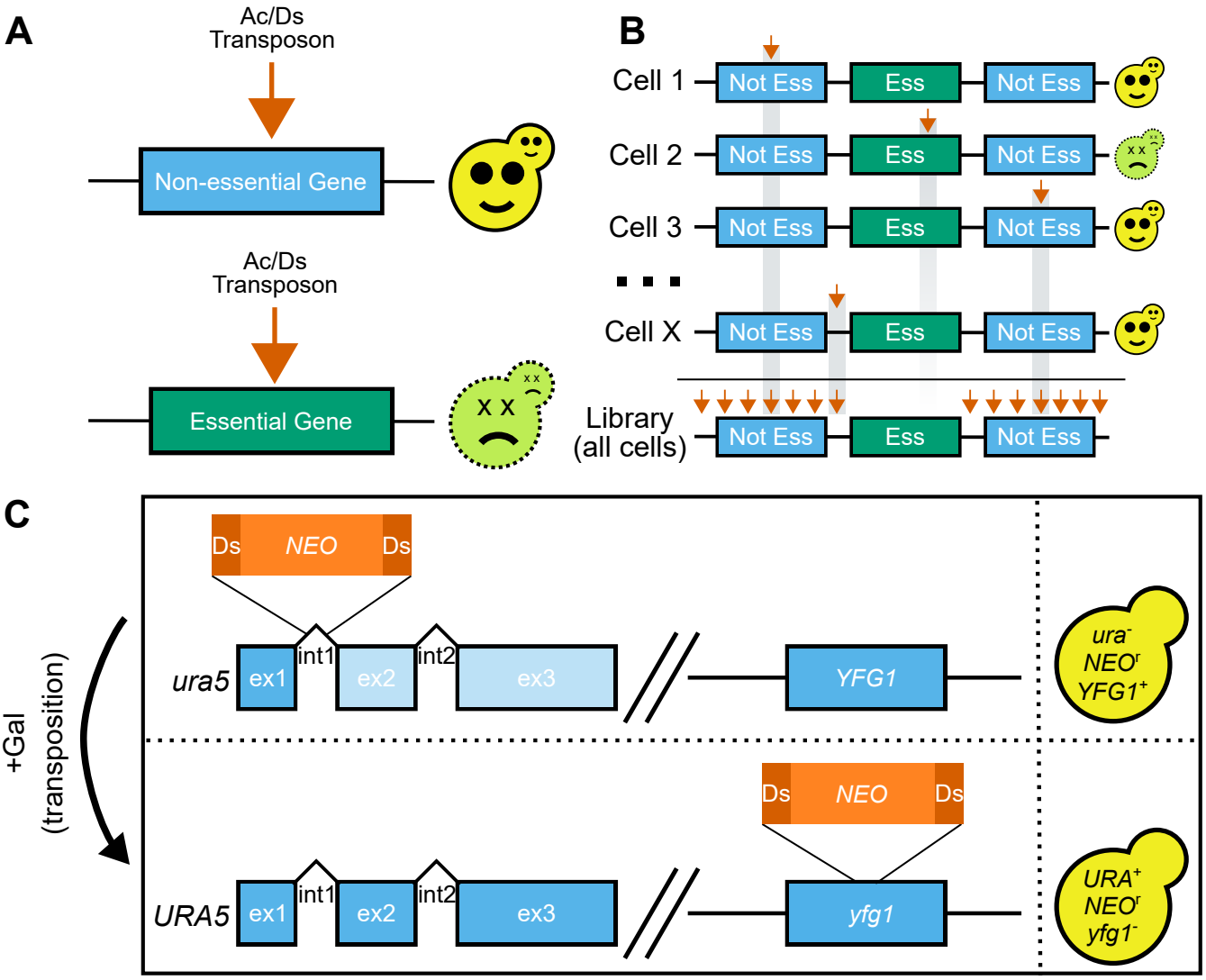

A
Ac/Ds
Transposon
Non-essential Gene
Ac/Ds
Transposon
Essential Gene
B
Cell 1
Cell 2
Cell 3
Cell X
Library
(all cells)
Not Ess
Ess
C
+Gal
(transposition)
Ds
NEO
Ds
ura5
ex1
int1
ex2
int2
ex3
YFG1
ura⁻
NEO^r
YFG1⁺
URA5
yfg1
URA⁺
NEO^r
yfg1⁻

Figure 2

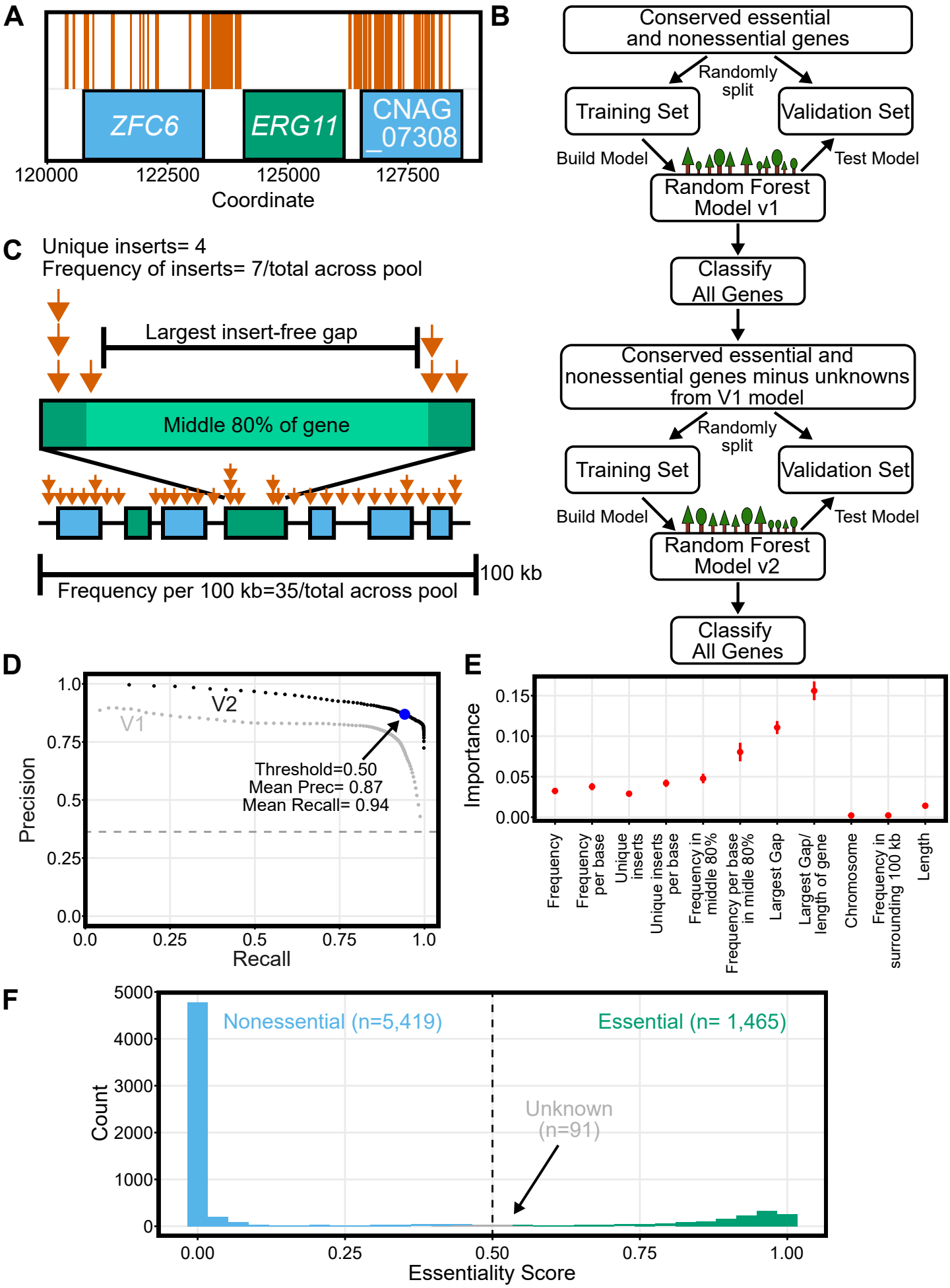
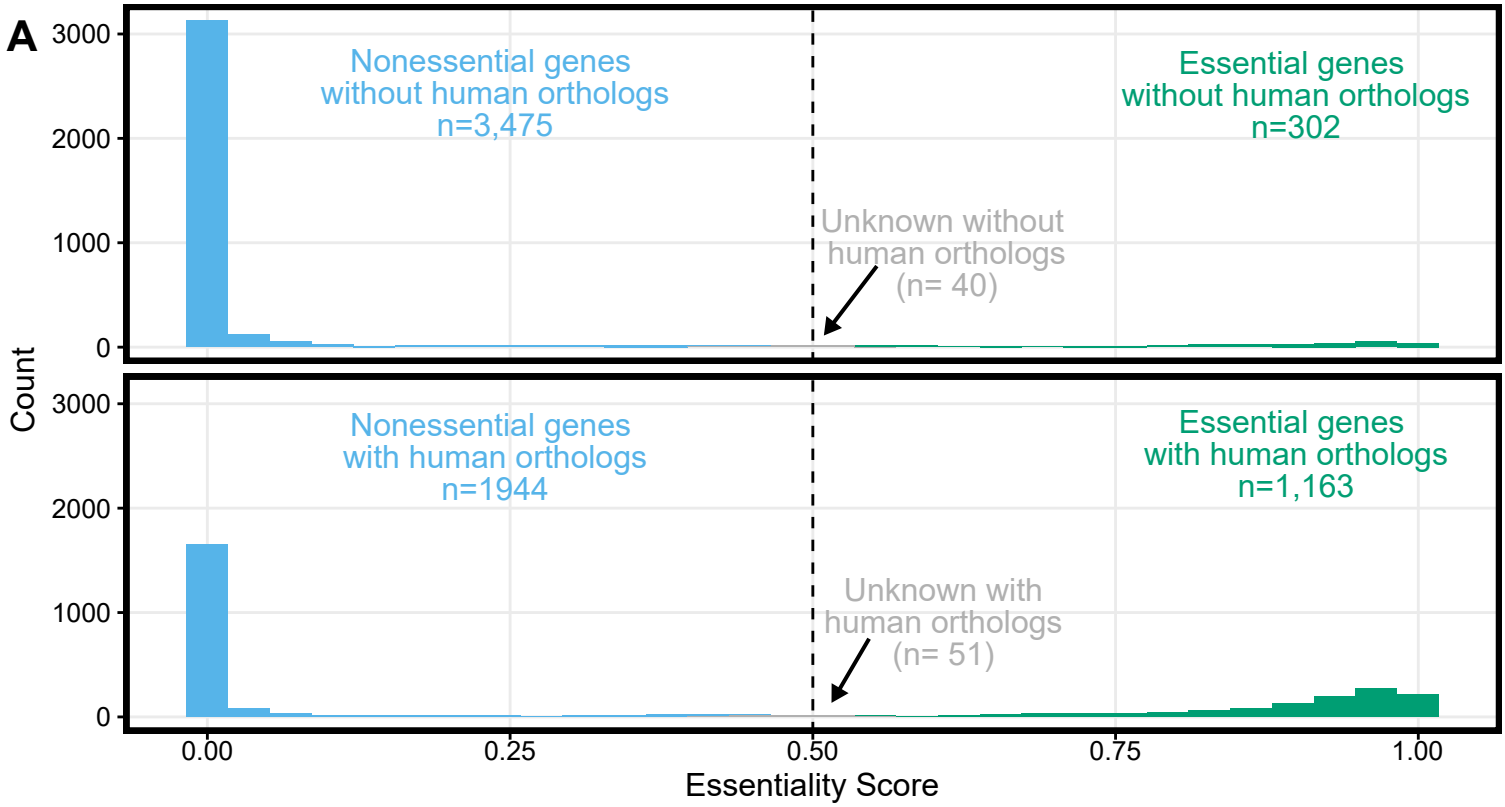

A
ZFC6
ERG11
CNAG_07308
120000
122500
125000
127500
Coordinate
B
Conserved essential and nonessential genes
Randomly split
Training Set
Validation Set
Build Model
Test Model
Random Forest Model v1
Classify All Genes
Conserved essential and nonessential genes minus unknowns from V1 model
Randomly split
Training Set
Validation Set
Build Model
Test Model
Random Forest Model v2
Classify All Genes
C
Unique inserts= 4
Frequency of inserts= 7/total across pool
Largest insert-free gap
Middle 80% of gene
100 kb
Frequency per 100 kb=35/total across pool
D
Precision
Recall
V2
V1
Threshold=0.50
Mean Prec= 0.87
Mean Recall= 0.94
E
Importance
Frequency
Frequency per base
Unique inserts
Unique inserts per base
Frequency in middle 80%
Frequency per base in midle 80%
Largest Gap
Largest Gap/ length of gene
Chromosome
Frequency in surrounding 100 kb
Length
F
Count
Nonessential (n=5,419)
Essential (n= 1,465)
Unknown (n=91)
Essentiality Score

Figure 3


A
Count
3000
2000
1000
0
Nonessential genes
without human orthologs
n=3,475
Essential genes
without human orthologs
n=302
Unknown without
human orthologs
(n= 40)
Nonessential genes
with human orthologs
n=1944
Essential genes
with human orthologs
n=1,163
Unknown with
human orthologs
(n= 51)
0.00
0.25
0.50
0.75
1.00
Essentiality Score

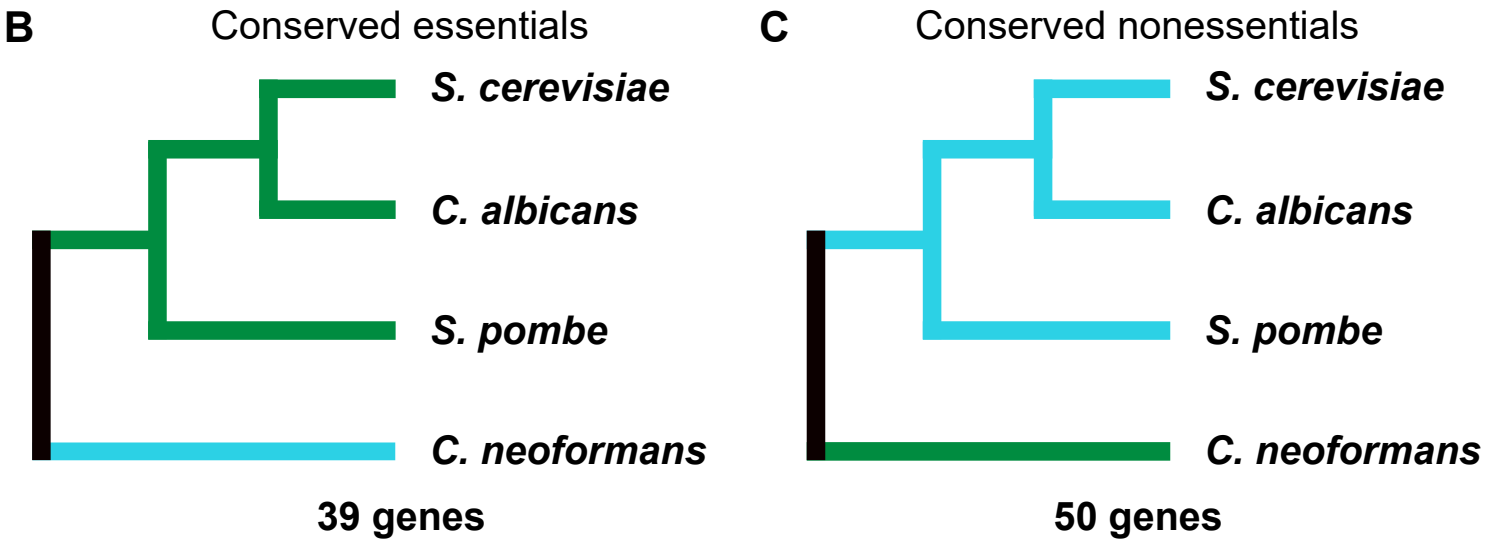

B
Conserved essentials
S. cerevisiae
C. albicans
S. pombe
C. neoformans
39 genes
C
Conserved nonessentials
S. cerevisiae
C. albicans
S. pombe
C. neoformans
50 genes

Figure 4

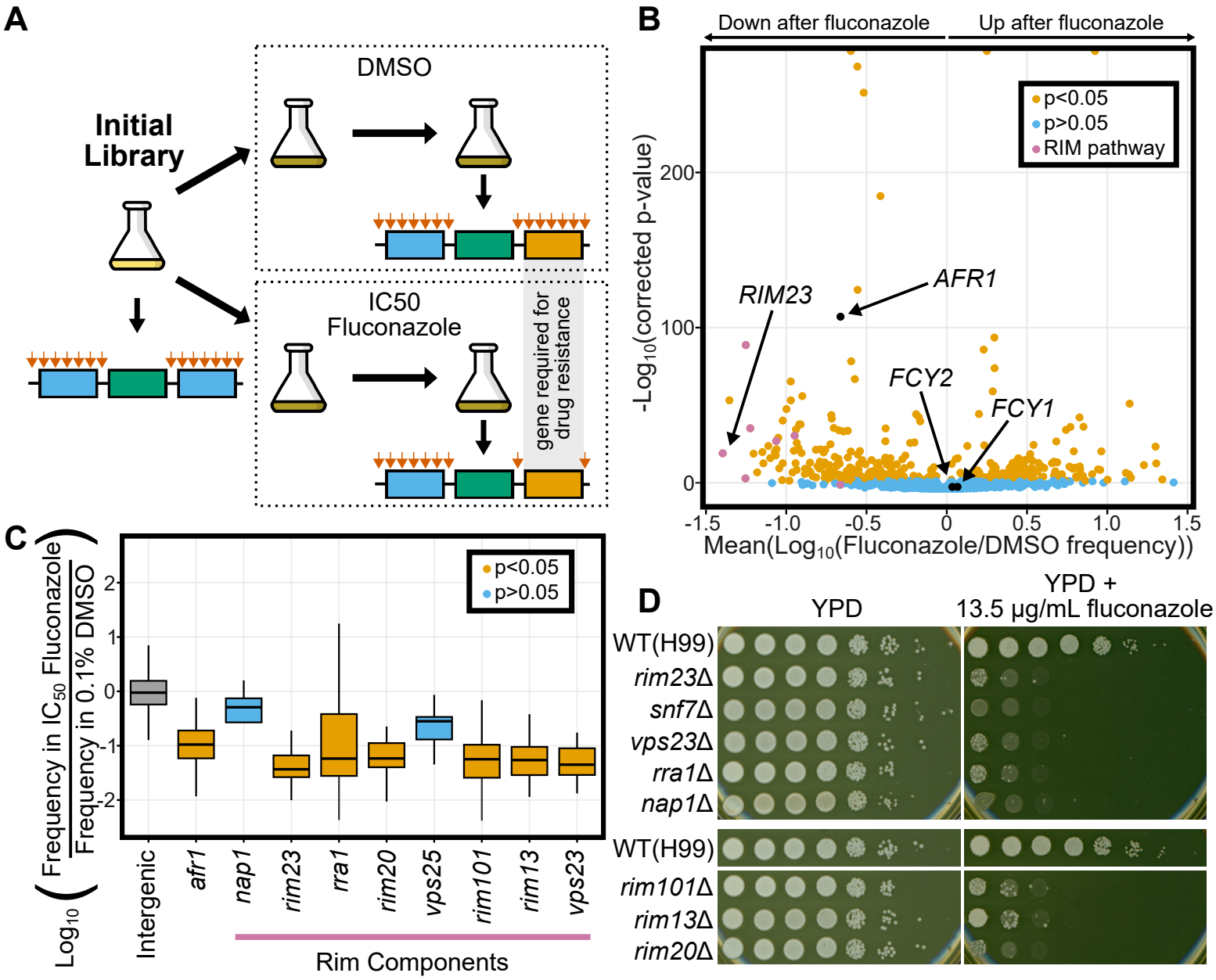
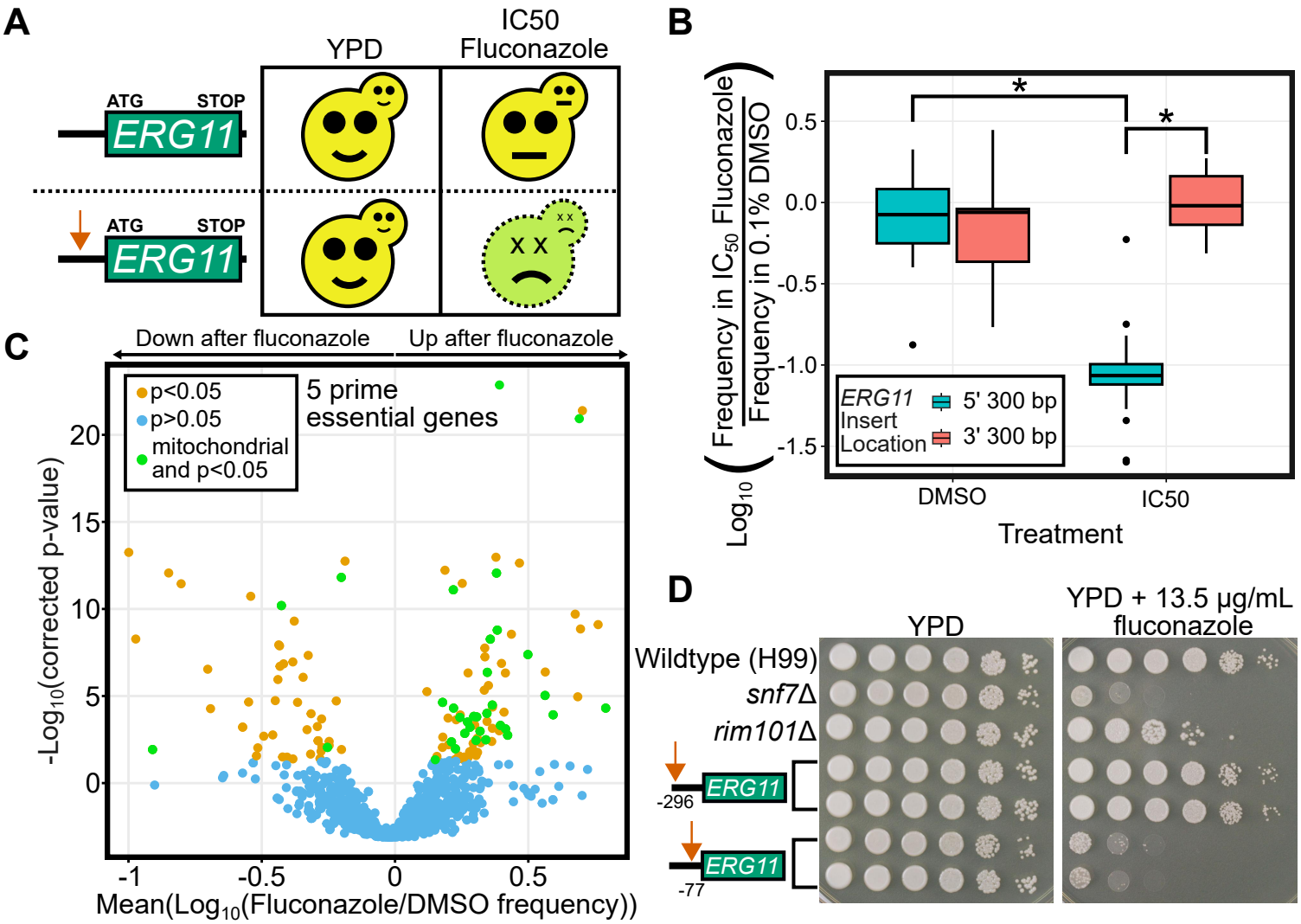
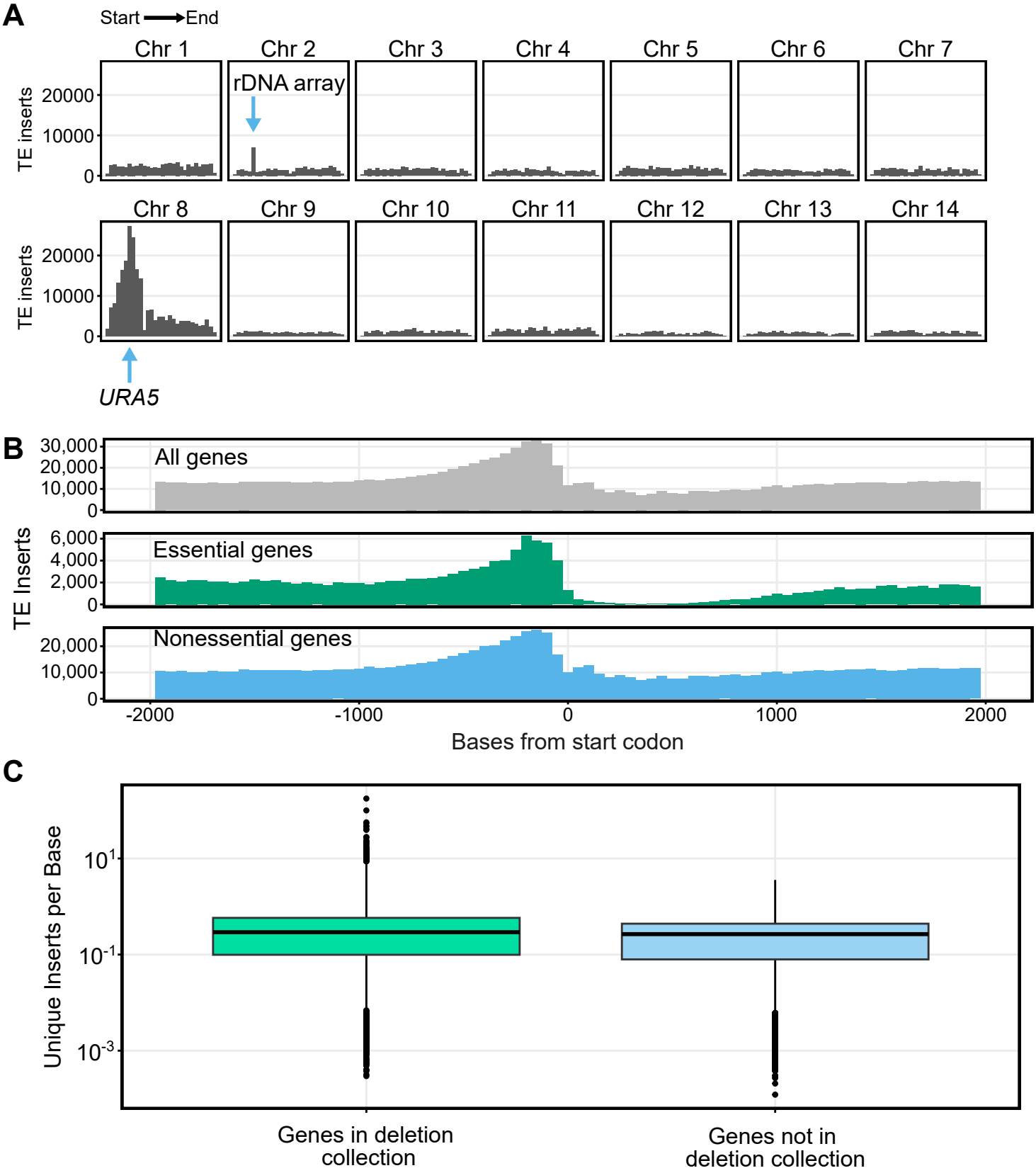

A
Initial Library
DMSO
IC50 Fluconazole
gene required for drug resistance
B
Down after fluconazole
Up after fluconazole
p<0.05
p>0.05
RIM pathway
RIM23
AFR1
FCY2
FCY1
-Log10(corrected p-value)
Mean(Log10(Fluconazole/DMSO frequency))
C
Log10 (Frequency in IC50 Fluconazole / Frequency in 0.1% DMSO)
p<0.05
p>0.05
Intergenic
afr1
nap1
rim23
rra1
rim20
vps25
rim101
rim13
vps23
Rim Components
D
YPD
YPD + 13.5 µg/mL fluconazole
WT(H99)
rim23Δ
snf7Δ
vps23Δ
rra1Δ
nap1Δ
WT(H99)
rim101Δ
rim13Δ
rim20Δ

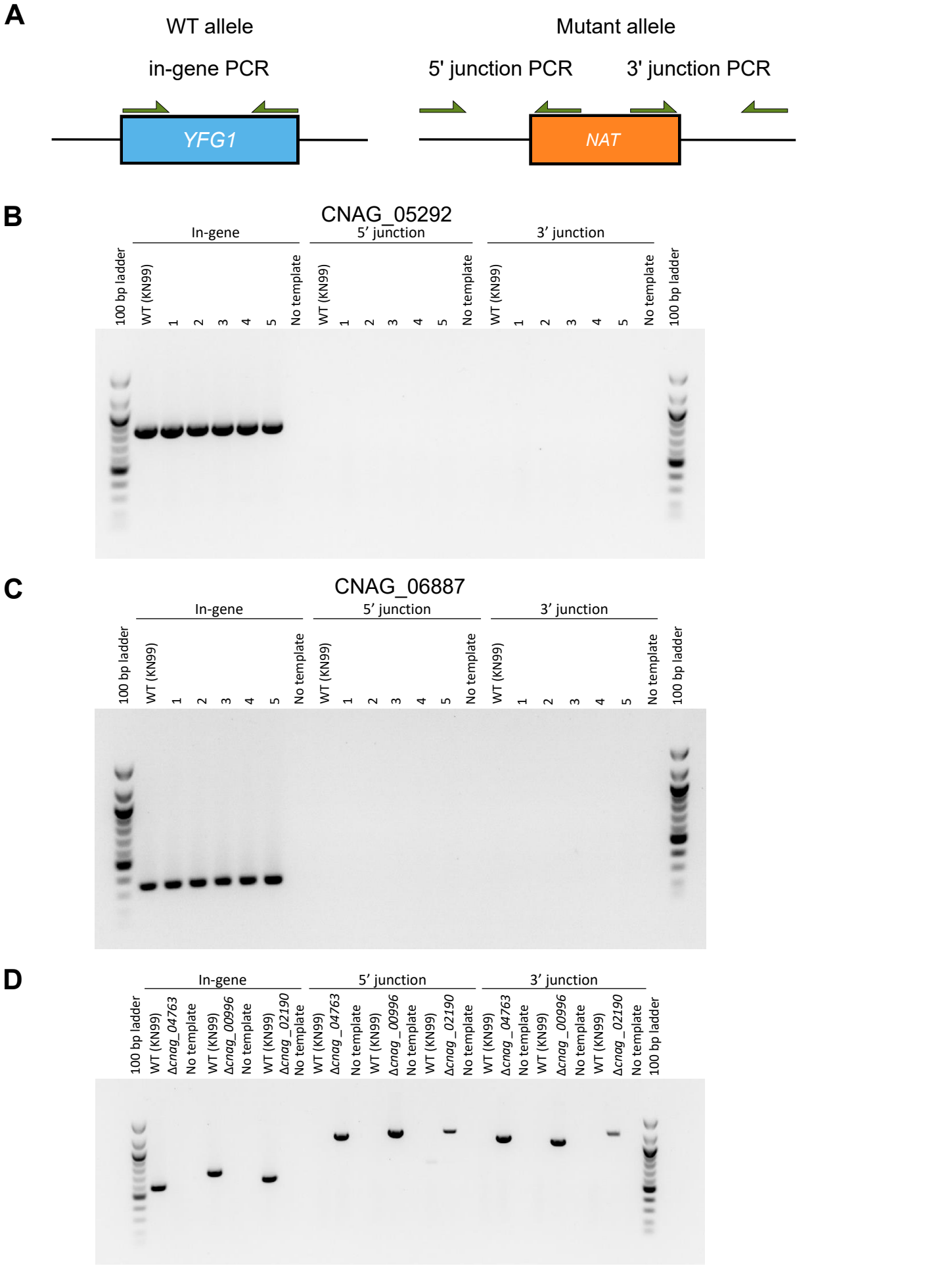
Figure 5


A
YPD
IC50
Fluconazole
ATG
STOP
ERG11
ATG
STOP
ERG11
B
Frequency in IC50 Fluconazole
Frequency in 0.1% DMSO
Log10
0.5
0.0
-0.5
-1.0
-1.5
*
*
ERG11
Insert
Location
5' 300 bp
3' 300 bp
DMSO
IC50
Treatment
C
Down after fluconazole
Up after fluconazole
p<0.05
p>0.05
mitochondrial
and p<0.05
5 prime
essential genes
-Log10(corrected p-value)
20
15
10
5
0
-1
-0.5
0
0.5
Mean(Log10(Fluconazole/DMSO frequency))
D
YPD
YPD + 13.5 µg/mL
fluconazole
Wildtype (H99)
snf7Δ
rim101Δ
ERG11
-296
ERG11
-77

Supplemental Figure 1

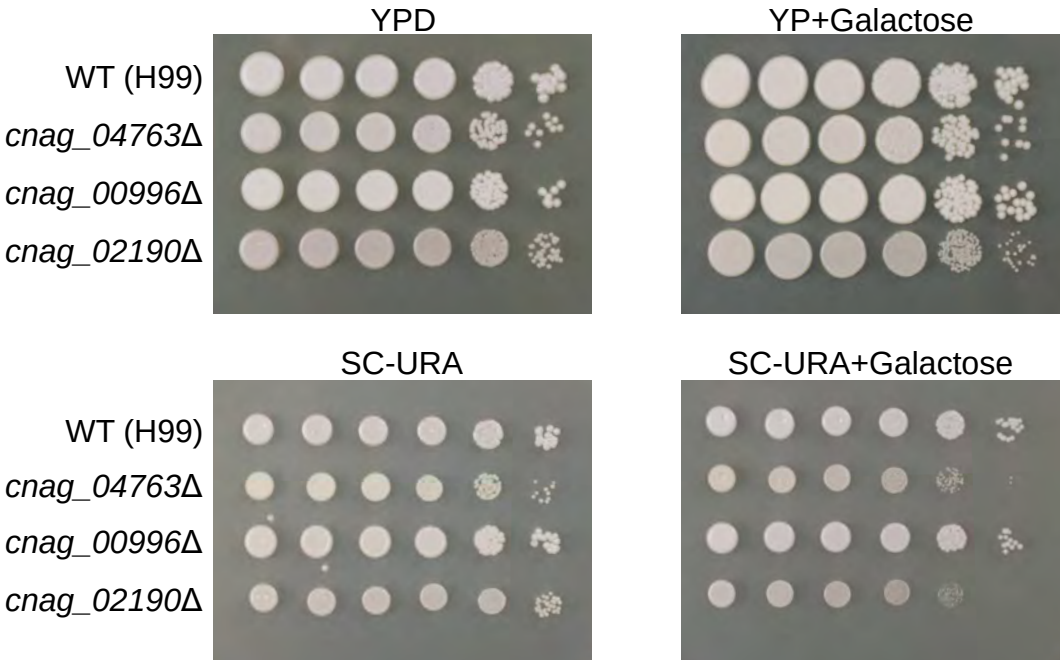


Supplemental Figure 2

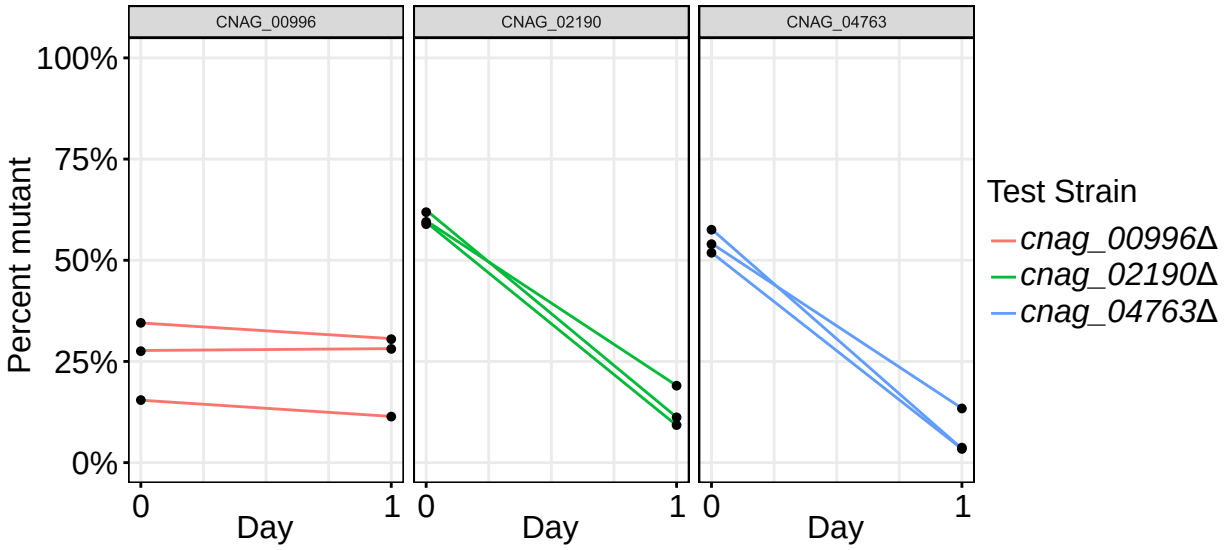


Supplemental Figure 3

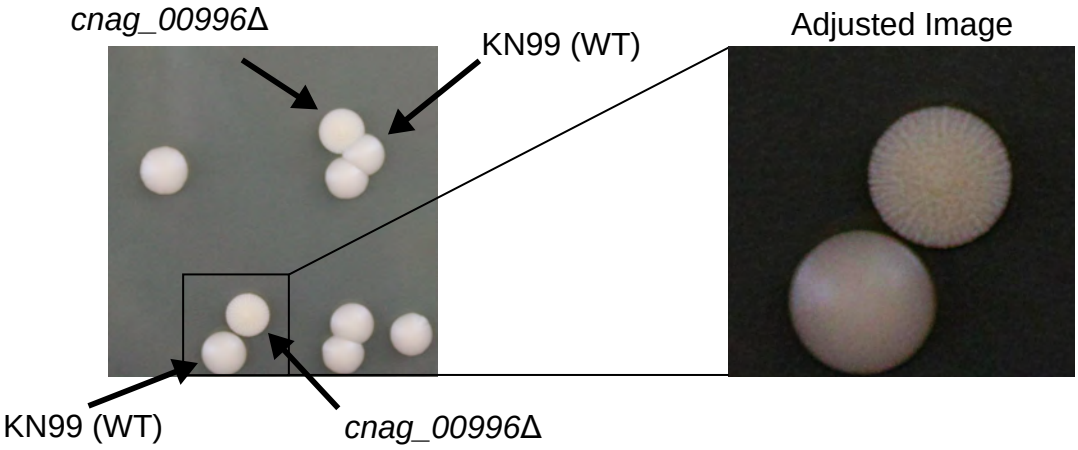
A



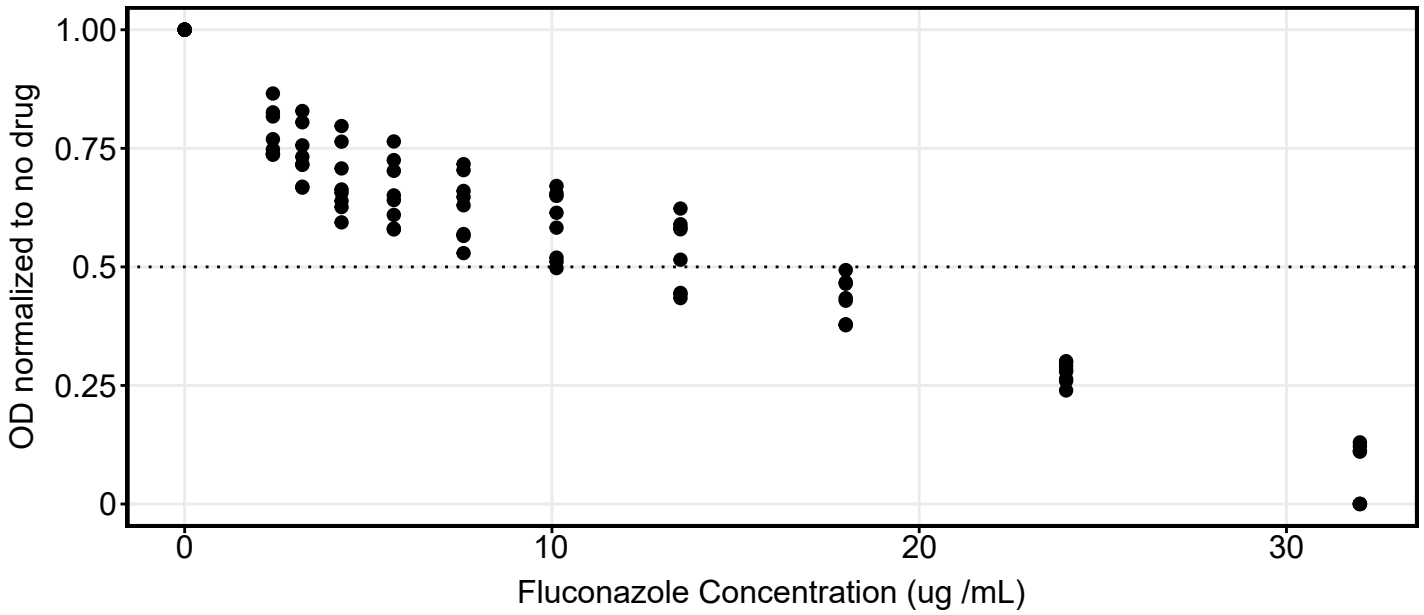
B



C



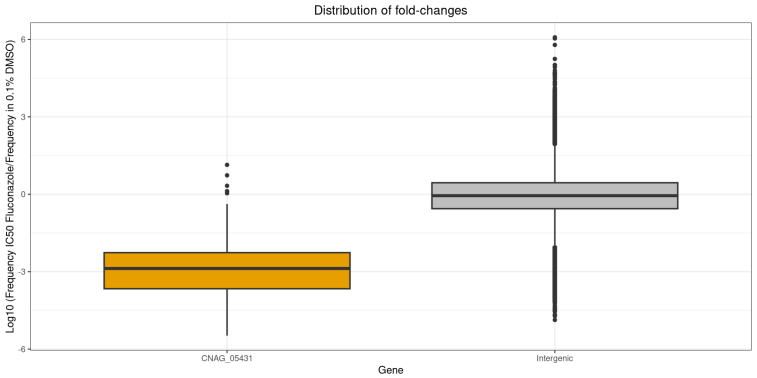
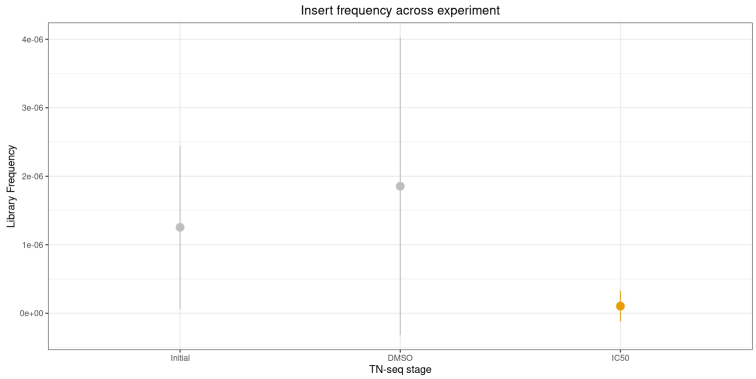
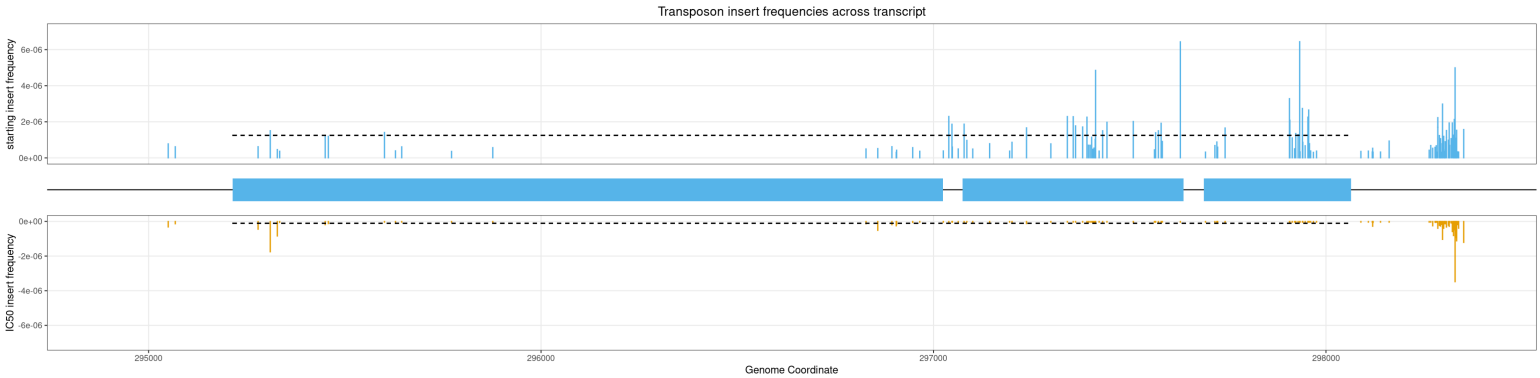
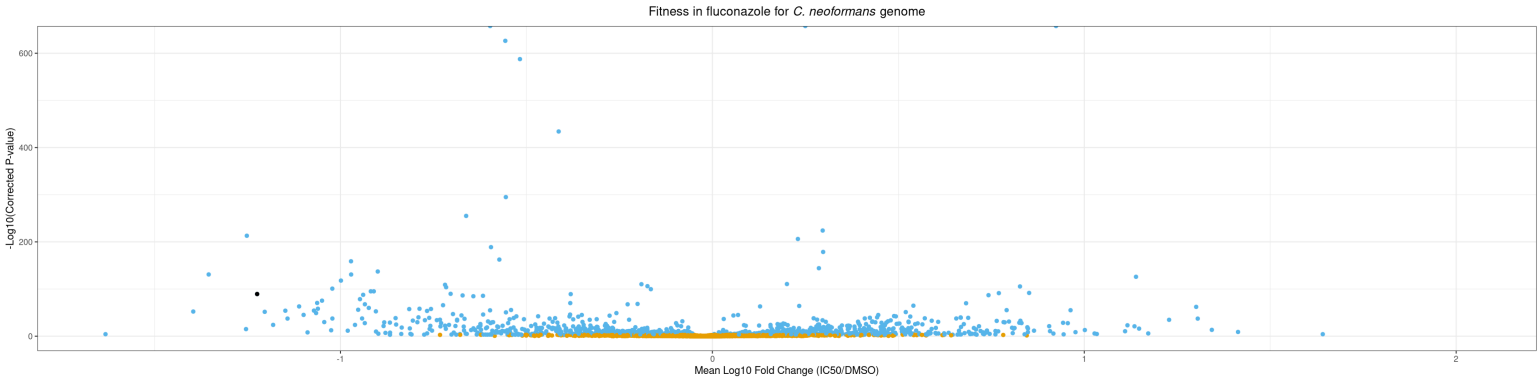
Supplemental Figure 4



Supplemental Figure 5

Crypto TN-seq

Gene Name:

CNAG_05431

Link to FungiDB