Diversity in the Regulatory Landscape Governing *Candida albicans* Biofilm Formation

By

Manning Y. Huang

A dissertation submitted to the

Faculty of the Mellon College of Science

Carnegie Mellon University

In partial fulfilment of the requirements for the degree of

Doctor of Philosophy in Biological Sciences

Under the supervision of

Professor Aaron. P. Mitchell

December, 2019

Abstract

The diploid dimorphic fungus *Candida albicans* displays a remarkable degree of phenotypic diversity among members of the species. In particular, strains display significant variability in capacity for filamentation and biofilm formation, virulence determinants of this clinically relevant human pathogen. Phenotypic diversity has not thus far been investigated in our studies of *C. albicans* virulence determinants, primarily because genetic engineering in this species relies on a single laboratory strain lineage: SC5314 and its engineered derivatives. In this dissertation, we discuss several methods to rapidly investigate gene function in multiple clinical isolates, and use both Nanostring and RNA-seq to compare regulatory circuitry governing filamentation and biofilm formation in *C. albicans*. We identify widespread circuit diversification, a phenomenon where a target gene is under the control of regulator A in background A, yet under the control of regulator B in background B. Finally, we explore possible mechanisms that may underlie circuit diversification, and investigate multi-strain core targets for gene function in filamentation.

Acknowledgements

I would like to thank all the members of the Mitchell Lab, past or present: Dr. Carol Woolford, Dr. Yuichi Ichikawa, for their many years of teaching and guidance; Tatyana Aleynikov, for her infinite assistance; Dr. Jigar Desai, Dr. Kayden Min, and Dr. Jill Blankenship for all their help and encouragement; excellent undergrads Julia Carter, Mia Burke, Sophie Halpern; Maximus Cravener (cray-veh-ner), for his inquisitiveness and friendship; and my benchmate Dr. Katie Lagree, for her friendship, instruction, and our many scientific discussions. Most of all, my supervisor Dr. Aaron P. Mitchell, for his endless forbearance, invaluable guidance, infinite support, many bon mots, and for – in his words – ensuring that we were well entertained!

I would like to thank Dr. Frederick Lanni, whose teaching, depth of knowledge, and curiosity has served as great inspiration, and for our many enjoyable discussions. I thank my committee members, Dr. John Woolford, Dr. Luisa Hiller, Dr. Damian J. Krysan, and Dr. Xiaorong Lin, for their instruction, encouragement, and support over many years. I thank our collaborators, Dr. C. Joel McManus, Gemma May, Dr. Scott Filler and Norma V. Solis, for their support and willingness to answer my many questions. I thank my many colleagues in the department: Surya D. Aggarwal, Karina M. Brown; my fellow cohort members, Scott Keith, Dan Wilson, Amber LaPeruta, Teresa Spix, Iris Zhipeng Yang, for their great company.

Lastly, I dedicate this to my family and friends, for their love and support. I thank my mother, Hui-chen Chuke, my father, Wen-jia Huang, Yvonne and Ben, Ryan and all the rest of my family. I thank Jessie Park, Peggy Shih, Yolanta Siu, Anni Greene for many late night discussions. I thank Yuzo Ishikawa and Steven Teng-Wei Tseng for certainly

not sabotaging me in the last hours of my writing process (in addition to many years of close brotherhood). I thank Edmund Tang, Katie Jiang and Tina Xu for keeping me sane and humble, and Zanetta Chang, for many comments and suggestions, endless encouragement, and most excellent camaraderie.

Thank you, everyone!

Table of Contents

Chapter 1: Introduction

Candida albicans is a diploid dimorphic fungus consistently identified as a significant component of the human myco/microbiome (1). The species may grow alternatively in yeast form - round, budding cells similar in appearance to *Saccharomyces cerevisiae* - or long filamentous hyphae that may extend beyond several hundreds of microns in length. *C. albicans* asymptomatically colonizes a number of niches in healthy individuals, and is the most abundant fungal species within the healthy human gut (2). However, in immunocompromised individuals, dissemination of *Candida* into the bloodstream (e.g. by peritonitis or through implanted medical devices) is frequently fatal (3). *Candida albicans* accounts for more than half of all cases of candidemia, and conservative estimates identify approximately 400,000 yearly cases of candidemia worldwide from all *Candida* species with an attributable mortality of 27% (4,5). Diagonosis of candidemia can be difficult, and treatment may also be complicated by a limited selection of antifungal drugs. Even the use of broad spectrum antibiotics may be counterproductive, as historical observations suggest that elimination of the normal microbial flora allows overgrowth of *C. albicans* that may then cause disease (6,7).

The primary strain utilized in the study of *C. albicans* has been the clinical isolate SC5314, isolated from a patient with a “generalized Candida infection” prior to at least 1968 (8). However, *C. albicans* strains are both genotypically and phenotypically diverse, with greatly varying capacity to cause disease (9). Compared to *S. cerevisiae, C. albicans* is much more genetically intractable. For this reason, investigators have relied heavily on the SC5314 type strain and its engineered auxotrophic derivatives.

Few dominant selectable markers are available for use. Furthermore, *C. albicans* lacks a complete sexual cycle, engaging in a parasexual cycle instead of a traditional meiotic cycle, and lacks a native plasmid system. As a result, engineering this diploid organism has generally been a time-consuming challenge, precluding the examination of gene function in multiple isolates. Herein, we discuss our work in tool development to accelerate the dissection of *C. albicans* genetics and our subsequent analysis of intraspecific variation in this clinically relevant pathogen.

Candida albicans Virulence Determinants

C. albicans expresses multiple virulence determinants, including toxins, secreted proteases, phospholipases, adhesins, and invasins (reviewed in 10,11). One unifying aspect seems to govern expression of these virulence factors: hypha formation. Hypha formation (interchangeable with the term filamentation) occurs under a distinct transcriptional program in response to a number of stress signals (12). Expression of these virulence determinants, such as the toxin Candidalysin encoded by the *ECE1* gene, occurs within this hyphal transcriptional program. Subsequently, *C. albicans* hyphae are both morphologically and compositionally distinct from yeast form cells. These long tubular hyphal cells adhere more efficiently to both human tissue and inorganic surfaces due to increased expression of cell wall linked (GPI-anchored) adhesins such as Als1, Als3, and Hwp1 (13). Yeast form cells, while still able to adhere to surfaces, lack robust expression of these adhesins and also express the anti-adhesin Ywp1 (14). Filamentation additionally plays roles in immune evasion, oxidative stress

responses, and trace metal acquisition (15). *C. albicans* is able to escape phagocytosis by macrophages through physical disruption of the phagosomal membrane through hyphal elongation (16,17). Oxidative stress tolerance occurs through the expression of hyphal associated superoxide dismutases such as *SOD5* (18). Zinc acquisition is mediated by expression of the *PRA1* gene in hyphal cells, and Als3 can also function in iron uptake (19,20). Unsurprisingly, this morphogenetic switch from yeast form to hyphal form is often considered the primary virulence determinant of *C. albicans*, and is an important focus for research.

One significant consequence of hyphal growth and the expression of cell-cell and cell-surface adhesins is biofilm formation. A *C. albicans* biofilm is a complex multicellular community comprised of both yeast and hyphae covered in extracellular matrix material. Biofilm formation can occur on abiotic surfaces, namely implanted medical devices, and biofilms are thought to act as reservoirs of cells that may be shed and disseminate into the bloodstream. The process *in vivo* is thought to begin with the adherence of yeast cells to some substrate (21). It is unclear if hyphal cells may also seed biofilm formation in this context, but this is thought to be unlikely as hyphal cells are less likely to disseminate to new sites, and the vast majority of cells shed from *in vitro* biofilms are in the yeast form. Following attachment, yeast cells multiply to form microcolonies and proceed into an initiation step, during which hyphae and pseudohyphae begin to form. The microcolonies are thought to give rise to a concentrated layer of yeast cells that has often been found at the basal layer of the mature biofilm, but has in our hands been observed to sometimes reside in middle biofilm layers. At this step, extracellular matrix

is also produced, and has been thought to be distributed throughout the intracellular spaces of the biofilm, but has mostly been directly observed in distinct blanketing layers on the apical surface of the biofilm (22). As the biofilm begins to mature, extracellular matrix will continue to accumulate and hyphae continue to elongate in a semi-ordered fashion, generating tall, cohesive biofilm architecture. The final stage of biofilm maturation involves dispersal of primarily yeast cells that may then cause infection and disease. However, dispersal may not be limited solely to the end stages of biofilm formation, and may occur throughout the entire process of biofilm formation as a response to nutritional depletion (23). Using their flow cell model, Uppuluri and colleagues demonstrated that biofilm-dispersed cells are phenotypically distinct and express increased adhesion and pathogenicity.

Compared to bacterial biofilms, *C. albicans* biofilms are thicker and contain greater biomass. Several *in vitro* and *in vivo* models exist for the study of biofilm formation and outcomes may vary depending on the model employed (24). The model discussed in this work relies on growth on silicone substrata to assay biofilm formation, and in some media conditions may produce biofilms that are more than half a millimetre in thickness. Similar to bacterial biofilms, fungal biofilms are highly resistant to antifungal treatment (25). Biofilm antifungal resistance occurs mainly through sequestration of drug molecules by the extracellular matrix component beta-1,3 glucan (26). Secondly, during early biofilm growth, biofilm induced increased expression of drug efflux pump genes *MDR1, CDR1,* and *CDR2* contribute to antifungal resistance, though these efflux pumps play a reduced role in mature biofilms (25). Additionally, both extracellular DNA in the

extracellular matrix, and other biofilm-induced stress responses contribute to the overall tolerance of biofilm cells to antifungal treatment.

Currently, three classes of antifungal drugs exist for use against *Candida* species: polyenes, azoles, and echinocandins. Polyenes incur nephrotoxicity, and thus cannot be employed at the necessary concentrations to treat candidemia (25,27). Resistance to each of these drug classes has been observed in some isolates. Common mechanisms for two of these resistance cases are known: point mutations in the ergosterol synthesis pathway gene *ERG11* confer resistance to azoles; point mutations in the 1,3-beta glucan synthase gene *FKS1* confer resistsance to echinocandins (25,27). Compared to cells in planktonic growth conditions, biofilm cells are 8-fold more resistant to polyenes, 1000-fold more resistant to azoles, and 2-20 fold more resistant to echinocandins (25). These observations highlight the need for development of novel antifungal drugs and the identification of additional drug targets.

Regulation of Filamentation and Biofilm Formation

Given that filamentation stands at the nexus of both virulence and antifungal resistance, the mechanisms that govern this process have been an attractive topic for the identification of novel drug targets (28). A number of environmental cues may induce *C. albicans* to form hyphae, including growth at human body temperature, physiological levels of carbon dioxide, nutritional stressors, presence of serum, neutral pH, presence of N-acetylglucosamine, and growth under embedded conditions (29,30). Opposingly,

filamentation is inhibited by the quorum sensing molecule farnesol (31). Broad integration of filamentation-inducing signals occurs through the evolutionarily conserved cAMP-PKA and MAP kinase pathways, which also govern morphogenesis in *S. cerevisiae*. These pathways signal into a highly interconnected hub of transcription factor genes. These are *BRG1, BCR1, ROB1, TEC1, EFG1, FLO8,* and *NDT80*, sometimes referred to as “master” regulators of biofilm formation (32–34). Each of these regulators directly binds the promoters and induces expression of some of the other members of this hub (32–34). Their primacy derives from the magnitude of their deletion mutant phenotypic severity, but many other transcription factors are connected to this master regulatory hub and appear to play some role in regulating filamentation and biofilm formation. For example, expression of hypha-specific genes is globally repressed by the transcription factors Nrg1, Tup1, and Rbf1. Some transcription factors play temporal roles in regulation of filamentation of biofilm formation, such as Ume6, Rfx2, and Gal4 (33,35). At the time of writing, between 29 and 66 transcription factor genes have been verified to regulate filamentous growth or biofilm formation (36,37). A number of these transcription factors play roles in pathways that bypass the major cAMP-PKA and MAP kinase pathways. The target of rapamycin (TOR) pathway interacts directly with Brg1, Bcr1, and Efg1, as well as other regulators in the master transcription factor hub (38,39). Growth under embedded conditions is signaled through the transcription factor Czf1, which bypasses Efg1 to activate hyphal associated gene expression (40). Taken together with the kinases, transcription cofactors, and chromatin remodeling factors shown to regulate filamentation and biofilm formation, we see that filamentation is a process under incredibly complex regulation with densely entangled

circuitry (12). In this work, we will primarily be concerned with the interactions of the hyphal associated transcription factors Efg1, Brg1, Bcr1, and Ume6. These transcription factors are among the most well characterized, with strong impact on filamentation and biofilm formation under several conditions.

Foremost of these is the transcription factor Efg1, which lies at the terminus of the cAMP-PKA pathway and upstream of several of the other master hub transcription factors (12,32). Efg1 was first identified by Stoldt and colleagues in 1997 to be involved in filamentation as disruption of a single allele led to a decrease in hyphal growth and overexpression induced hyperfilamentation in both solid and liquid conditions (41). Shortly after, Lo and colleagues demonstrated that a homozygous *efg1* deletion mutant both failed to form hyphae and showed reduced virulence in a murine disseminated infection model (42). Subsequent work has shown that Efg1 is required for filamentation in response to all signals except growth under embedded conditions (43). Unsurprisingly, *EFG1* is required for normal biofilm formation (44). Interestingly, an *efg1* mutant is hyperfilamentous under embedded growth conditions, suggesting that *EFG1* serves as a transcriptional activator in response to some signals, but a transcriptional repressor in response to others (43). Additionally, Efg1 plays a role in two other pathways tangentially connected to filamentation and biofilm formation. Metabolic gene expression and regulatory adaptation are involved in biofilm formation, although the causative relationships remain uncharacterized (45,46). Efg1 has been shown to regulate glycolytic gene expression indirectly through the transcription factor gene *TYE7*

(47). Efg1 also plays a prominent role in white-opaque switching, the morphogenetic switch involved in the incomplete sexual cycle of *C. albicans* (48).

Brg1 was first identified by Uhl and colleagues in a haploinsufficiency transposon mutagenesis mutant screen as an ortholog of *S. cerevisiae* Gat2 (49). Uhl and colleagues as well as Homann and colleagues demonstrated that *C. albicans* Gat2 (interchangeable with the later popularized name Brg1) was a positive regulator of filamentous colony morphology (49,50). Du and colleagues later demonstrated that a homozygous *brg1* mutant failed to form biofilms on silicone and plastic, and failed to filament under liquid hypha-inducing growth conditions (51). In addition to the positive regulation of *BRG1* by Efg1 and other master hub transcription factors, *BRG1* integrates signals from the TOR pathway and interacts antagonistically with Nrg1, one of the global inhibitors of filamentation. Lu and colleagues demonstrated that activation of the cAMP-PKA pathway combined with reduced TOR signalling induced *BRG1* expression and Brg1-mediated recruitment of a histone deacetylase, Hda1, to the promoters of hypha associated genes (38). Hda1-mediated nucleosome repositioning then led to the obstruction of Nrg1 binding sites, allowing maintenance of hypha associated gene expression and hyphal growth (38).

Bcr1 canonically regulates biofilm formation, but not filamentation. In a transcription factor transposon mutant library screen, Nobile and Mitchell demonstrated that a *bcr1* mutant failed to form a biofilm on silicone substrata, but was able to form normal hyphae that were apparent in the surrounding growth media (52). Microarray analysis revealed

that Bcr1 positively regulated expression of cell wall genes, including the adhesin genes *HWP1* and *ALS3*. These data initially suggested that Bcr1 may act as a downstream regulator, responsible for the expression of a functional subset of the master regulatory hub targets (53). Bcr1 also plays a role in biofilm-mediated antifungal resistance, as well as general resistance to antimicrobial peptide (54,55). Srikantha and colleagues showed that Bcr1 played different roles in the regulation of extracellular matrix composition between a/alpha and a/a *C. albicans* biofilms (54). By investigating the genes activated by Bcr1 overexpression, Srikantha and colleagues identified 8 target genes that explained the higher antifungal resistance of a/alpha biofilms through increased impermeability and impenetrability. Bcr1 is necessary for virulence in a mouse model of oropharyngeal candidiasis, and may be bypassed by other regulators under in vivo conditions (56,57).

Ume6 demonstrates complex interactions with inhibitors of filamentation and plays a role in biofilm maturation. Banerjee and colleagues demonstrated that a *ume6* mutant was defective in filamentation in a number of solid and liquid hypha-inducing growth conditions, but only modestly defective in biofilm formation (35,58). Under liquid growth conditions, a *ume6* mutant retained the ability to form germ tubes, but failed to form complete, elongated hyphae (35). While expression of hypha associated genes was similar between the *UME6* mutant and wildtype at early time points, expression was significantly decreased at later time points, indicating a role for Ume6 in regulating hyphal extension (35). *UME6* expression is subject to complex temporal mechanics - *UME6* transcript levels spike shortly after induction of hyphal growth, corresponding with

a decrease in Nrg1 transcript levels, before diminishing over time (35). Like Bcr1, Ume6 has been considered a “downstream” regulator - *UME6* expression is repressed by both Nrg1-Rfg1 as well as Nrg1-Tup1 (35). Additionally, *UME6* expression is regulated by Brg1 and Hda1, and Ume6 protein levels are maintained by Ofd1 in response to hypoxia and high carbon dioxide (59,60).

It is important to note that the interconnectivity of these transcription factors does not provide robustness to the overall network. Instead, modeling of the network properties of the master regulatory hub suggests a highly efficient, but fragile network (34). Glazier and colleagues support this network model hypothesis through haploinsufficiency experiments demonstrating that modest changes to the regulatory hub lead to large changes in target gene expression and biofilm phenotype (34,61). Although many functional target genes have been identified downstream of these transcription factors, the total set of effectors required for filamentation and biofilm formation is not yet known.

It may seem surprising that the biofilm regulatory network involves so many transcription factors intertwined with complex regulatory connections. However, Sorrells and Johnson argue that this level of complexity is not without precedent, and that complex regulatory networks are common for many biological processes (62). With respect to biofilm formation, network complexity may be required to integrate the diverse signals encountered by *C. albicans* in changing environments, or fine tune outputs in response to complex signals (32,62,63). However, they propose an

alternative hypothesis, where complexity in biofilm regulatory networks may not represent an optimized solution, but rather may have developed coincidentally during neutral evolution (62). They theorize that when a regulator gains connections to a set of target genes, this is unlikely to occur on a gene by gene basis. Instead, the new connection might be established indirectly through a second regulator that has existing regulatory connections to the set of target genes. Sorrells and Johnson support their argument by highlighting ChIP-Seq experiments in the biofilm master regulatory hub where a transcription factor binding event is detected, but target gene expression is the same between transcription factor mutant and wild type. Although this observation may be explained by complex genetic interactions, Sorrells and Johnson suggest that some proportion of a transcription factor's regulatory targets will be non-functional.

Natural Variation Among *C. albicans* Clinical Isolates

C. albicans natural isolates display remarkable phenotypic diversity. Strains differ significantly in growth rate, with a 1.5 fold range separating slowest and fastest doubling strains (64). Further phenotypic variation has been observed in responses to cell wall stress, adhesiveness, and filamentation on both solid and liquid media (64,65). Variation in antifungal resistance associated with increased efflux pump expression has been observed (66). Such inter-strain variation was used to identify the zinc-finger transcription factor gene *MRR1*, whose increased expression in resistant isolates induced increased expression of the efflux pump gene *MDR1* (67). Individual strains differ greatly in pathogenic potential between total mortality by five days, out through no

mortality by 24 days (9). Wu and colleagues demonstrated that this variability in pathogenic potential was strongly associated with homozygosity or heterozygosity at the mating type locus (9).

Variability in biofilm formation has also been demonstrated. Biofilm formation capacity may differ by at least 10-fold between the weakest and strongest biofilm formers (68,69). Furthermore, virulence outcomes have been shown to vary correspondingly between different clinical isolates. Isolates with higher biofilm formation showed increased virulence in a murine model of disseminated candidiasis, and high biofilm forming isolates were more virulent than low biofilm forming isolates in a *Galleria mellonella* infection model (68,70). Outcomes in association with biofilm formation ranged between complete mortality by 5 days for high biofilm formers, 80% mortality by 10 days for low biofilm formers, and no mortality through 30 days for non-biofilm forming strains (68). Clinical studies by Rajendran and colleagues; and Tumbarella and colleagues reported that high biofilm formation was associated with significantly increased patient mortality (68,69). As expected, high biofilm formers were also more resistant to antifungal treatment (69,70).

High and low biofilm forming strains differ in biofilm composition, and a significant correlation was observed with increased cell surface hydrophobicity and high biofilm formation (70). Furthermore, transcriptional differences were observed between clinical isolates with varying biofilm formation capacity. Surprisingly, while a few hypha associated genes were significantly differentially expressed, overall expression of hypha

associated genes was highly similar between high and low biofilm forming strains (70). A separate study showed using RNA-seq of high and low biofilm forming isolates that only 4% of differentially regulated genes between high and low biofilm formers mapped to biofilm formation or cell adhesion (46). Instead, approximately half the differentially expressed genes were involved in transport, response to chemicals, and response to stress (46). Network analysis also showed that expression of fatty acid metabolism, pyruvate metabolism, as well as arginine and proline metabolism pathways was upregulated in high biofilm formers (46).

What is the extent of genetic variation between clinical strains? The *C. albicans* population genetics field informally employs the term “clade” to refer to a group of closely related strains. Multi-locus sequence typing (MLST) efforts have identified 17 such clades, and subsequent analyses reveal that 97% of isolates may be assigned to one of these 17 clades (64,71). *C. albicans* lacks a complete meiotic cycle, and recombination occurs through a parasexual cycle. Two MTL homozygous cells of opposite mating types may fuse, becoming tetraploid, and subsequently undergo concerted chromosome loss to return to normal diploidy (72). This parasexual cycle is rarely observed and correspondingly, population studies have identified that *C. albicans* propagates primarily by clonal descent (64,71,73).

Overall, strains are genetically highly similar. Several collections of clinical isolates have been fully sequenced. The average nucleotide diversity between strains is 0.37%, comparable to the diversity between *S. cerevisiae* strains and to the diversity between

two unrelated humans (64,74,75). Genetic variation manifests mostly as a collection of single nucleotide polymorphisms (SNPs), copy number variation events (CNVs), aneuploidy, and loss of heterozygosity (LOH) (64). Strains within a clade differ by approximately 1 base every 1400 bps, and across clades by approximately 1 base every 250 bps (64). In an examination of 21 strains, truncations were distributed across 175 genes and frameshift mutations were distributed across 391 genes (64).

Connecting Genotypic Variation and Phenotypic Variation

Genotypic variation is commonly connected to phenotypic variation through differences in gene content. This is most famously observed in bacterial pathogens, whose distributed genome may include genetic elements required for virulence. For example, pathogenic potential in *Escherichia coli* is linked to acquisition of the shigella toxin genes *stx1* and *stx2*, and pathogenic potential in *Salmonella* species depends on the presence of a number of large gene cassettes (pathogenicity islands) (76,77). In *C. albicans*, gene content across strains does not vary drastically; on average, 98.1% of 6189 open reading frames were present in all 21 other examined strains (64). Subsequently, differences in gene content explain phenotypic outcomes observed in only 1 of 21 examined strains. The strain P94015 is defective in biofilm formation and filamentation, and this strain was shown to harbor a homozygous nonsense mutation in the *EFG1* open reading frame (ORF) (64). Of the remaining strains, two are heterozygous for an *EFG1* loss of function allele, but may still form robust biofilms (Max Cravener, personal correspondence). Overall, it does not appear that loss or gain of

function mutations account for variability in filamentation and biofilm phenotypes in clinical isolates.

In cases where inter-strain gene content is similar, variation in allelic strength may explain phenotypic differences. Genetic background interactions were suggested to affect differences in gene essentiality in *S. cerevisiae* isolates. By examining deletion libraries in two different *S. cerevisiae* strains, Dowell and colleagues identified that 6% of the essential genes were only essential in one background (78). Hybrid crosses demonstrated that most of the observed conditional essentiality resulted from interactions with at least four modifiers (78). Vu and colleagues examined the mutant phenotypic severity of RNAi knockdowns of 1400 genes in two *Caenorhabditis elegans* isolates. Vu and colleagues only examined genes wherein RNAi efficiency did not drive observed phenotypic differences (79). Approximately 20% of examined genes had different mutant phenotypic severity between isolates, despite containing no differences within their coding regions (79). Instead, genetic background interactions appeared to induce the majority of observed differences. Lower wild-type expression of the analyzed gene or other genes in the same functional grouping led to increased mutant phenotypic severity (79). Chin and colleagues formally described a mechanism connecting genetic interactions with regulatory variation in *S. cerevisiae*. While the filamentation MAP kinase pathway is required for *FLO11* mediated adhesion in the *S. cerevisiae* Sigma background, this pathway is dispensable for expression of *FLO11* and adhesion in the S288c background (75). With allele swap experiments, Chin and colleagues demonstrated that the S288c *RPI1* allele, but not the Sigma *RPI1* allele, was sufficient

for expression of *FLO11* in either strain background (75). Chin and colleagues coined the term “circuit diversification” to describe this mechanism, where a target gene is under the control of regulator A in strain A, but under the control of regulator B in strain B.

Quantitative trait loci mapping is a popular method used to assign heritability to genetic elements, and therefore indirectly identify differences in allelic activity. Lin and colleagues employed this approach in *Cryptococcus neoformans* to connect allelic differences in the transcription factor gene *MAC1* to phenotypic variation in melanization, filamentation, and copper homeostasis (80). Gallagher and colleagues employed QTL mapping to investigate the genetic basis of resistance to 4-nitroquinoline 1-oxide in *S. cerevisiae*, and identified a specific allele of the transcription factor *YRR1* that conferred resistance (81). Using ChIP-Seq, Gallagher and colleagues demonstrated that these two Yrr1 alleles bound different targets in different backgrounds, with a greater number of targets bound by the resistance-conferring allele (81). While differences in gene content may satisfactorily explain simple traits, variation in complex traits appears to be primarily driven by non-coding variants that affect gene regulation (82,83). Non-coding variants that affect gene expression are termed expression quantitative trait loci (eQTL), and may be divided between *cis* and *trans* eQTL based upon their mechanism of action (84). Recent evidence has suggested that *trans* acting eQTL are responsible for more than half of the observed variance in expression levels (84,85).

What is the extent of regulatory variation in *C. albicans* biofilm regulatory networks? Although QTL and eQTL experiments have been employed to great effect in many species, the lack of facile mating in *C. albicans* complicates and practically prevents the use of QTL and eQTL mapping approaches. Additionally, even when causal variants of phenotypic differences are identified by eQTL mapping, the specific mechanisms by which gene expression differences are generated may be unclear. For example, if a set of SNPs is identified as a *cis* acting eQTL, each SNP may affect binding of a number of different transcription factors. While an eQTL may affect epigenetic modifications, pre-mRNA splicing, post-transcriptional regulation, as well as post-translational regulation, altered transcription factor binding is thought to be the primary mechanism driving regulatory variation (85).

How broadly will any given transcription factor's target genes vary? Are altered transcription factor binding, such as in the case of *YRR1*, and circuit diversification, such as in the case of *RPI1*, common phenomena, or limited to specific hyper-variable transcription factors? We reasoned that if regulatory differences contribute to phenotypic divergence in *C. albicans* biofilm formation, then some of the transcription factors may have different target genes in different backgrounds. In particular, we should expect that transcription factors with apparently different allelic strengths may have hyper-variable target genes. Herein, we employ a simple approach to investigate this question. We examine the mutant phenotypic impact of transcription factor deletions in parallel in a panel of *C. albicans* clinical isolates. As mutant strain construction in *C.*

albicans can be a difficult task, this analysis was made possible by the development of several CRISPR-Cas9 engineering tools.

Advances in *C. albicans* Strain Engineering

C. albicans studies have primarily been conducted in the SC5314 type strain background and its auxotrophic derivatives BWP17 and SN152. BWP17 is a Ura- His- Arg- auxotroph, while SN152 is a His- Leu- Arg- auxotroph (86,87). Three auxotrophic markers were commonly required before the deployment of CRISPR-Cas9 methods - two markers to manipulate each allelic copy of a gene of interest (GOI), and a final marker to complement the GOI mutation. Three dominant selectable markers have been examined in *C. albicans* isolates: *NAT1*, conferring resistance to nourseothricin; *CaHygB*, conferring resistance to hygromycin; and *IMH3*, conferring resistance to mycophenolic acid (MPA) (88,89). Only the *NAT1* marker has been deployed with reliable success, as MPA and hygromycin resistance can be highly variable across strains (90). Furthermore, selection for MPA resistant colonies is slow, requiring up to 7 days of growth (91).

The first recyclable marker deployed in *C. albicans* studies was the URA-blaster, adapted from *S. cerevisiae* studies, and later optimized into the URA-blister marker (92–94). These markers allowed excision of the *URA3* gene by counterselection on 5-fluoroorotic acid (5FOA). Unfortunately, counterselection on 5FOA is evidently mutagenic, and was shown to induce various chromosomal changes (95). Furthermore,

specific care is required for the use of the *URA3* marker, as Ura auxotrophy affected filamentation and virulence outcomes (96). To address the need for a better recyclable marker, the recyclable SAT1 flipper was developed, containing FLP target sequences flanking a *FLP* recombinase gene and the nourseothricin resistance marker *SAT1* (sourced from *E. coli* whereas *NAT1* was sourced from the *Streptomyces noursei* genome) (91). While the *NAT1* marker and SAT1 flipper methods are highly efficient, the potentially prohibitive cost of nourseothricin renders these methods unattractive to investigators. A Cre-*loxP* was adapted for use in *C. albicans*, with *loxP* flanked His, Arg, and Ura markers (97). This method requires three total transformation steps – two to delete a given GOI, and one two introduce the Cre recombinase, but will yield a fully auxotrophic homozygous deletion mutant at the end of the process (97).

One objective in refining *C. albicans* strain engineering has been to reduce the total number of required manipulations. While reducing the time required for strain engineering is a common motivator in many species, in *C. albicans*, reducing the number of manipulation steps is thought to reduce the risk for undesired genome changes (98). The *UAU1* method is one such method developed to allow the manipulation of a homozygous deletion mutant in a single transformation. One allele is disrupted by a split *URA3* marker interrupted by the *ARG4* marker. Subsequent gene conversion disrupts the second allele of the target gene through a recombinational junction that reveals the *URA3* marker (99). In addition to the disadvantages connected to the *URA3* marker, use of the *UAU1* method is complicated by the potential for significant loss of heterozygosity during gene conversion, and aneuploidy at the target

locus in the resultant strains (99). Nonetheless, the *UAU1* method has proven to be an effective tool for basic research in *C. albicans*.

The use of the “programmable nuclease” CRISPR-Cas9 has greatly accelerated genome manipulation in many species (100,101). The first CRISPR-Cas9 system developed for use in *C. albicans* relied on the integration at the *ENO1* locus of a large cassette containing the *CAS9* gene and single-guide RNA (sgRNA) expression components (available in two similar versions as the “Solo” and “Duet” systems) (102). The cassette is co-transformed into the desired *C. albicans* strain along with a markerless repair template spanning the desired editing site (102). This method was demonstrated with good editing efficiency in the simultaneous mutation of up to four loci (102). One of the leading concerns with CRISPR-Cas9 editing is the possibility of off-target cleavage. Genome-integrated systems may carry a higher risk of inducing unwanted genome changes. This risk may be addressed by the use of a transient CRISPR-Cas9 system. Min and colleagues performed a modification of the “solo” CRISPR-Cas9 transformation protocol and selected for a marker-bearing repair template instead of integration of the *CAS9* and sgRNA expressing vector (103). They observed that many transformants containing a homozygous deletion of their GOI did not contain the unselected *CAS9* and sgRNA expressing vector (103). Min and colleagues demonstrated that high CRISPR-Cas9 editing efficiency could be achieved by the simple inclusion of *CAS9* and sgRNA expression cassettes as individual linearized DNA fragments generated by PCR amplification (103).

Both the genome-integrated system from Vyas and colleagues and the transient system developed by Min and colleagues allow for the construction of homozygous deletion mutants using only a single selectable marker and in less than half the time required by pre-CRISPR-Cas9 methods. While this would therefore allow an investigator to examine a homozygous deletion of a GOI in a clinical isolate, additional markers or recyclable markers are required for analysis of genetic interactions as well as mutant validation. While the “solo” and “duet” system were designed to employ a recyclable NAT1 marker, the transient system lacks a compatible marker recycling strategy. In order to employ the transient CRISPR-Cas9 system in the manipulation of *C. albicans* clinical isolates, we developed a CRISPR-Cas9 based marker recycling strategy for both dominant selectable markers as well as auxotrophic markers (described in Chapter 2). Using insights gained during this development process, we also present a CRISPR-Cas9 toolkit for gene concatenation that has been useful in both mutant validation and in other cloning-free manipulations (described in Chapter 3).

Literature Cited, Chapter 1

1. Hallen-Adams HE, Suhr MJ. Fungi in the healthy human gastrointestinal tract. Virulence. 2017 Apr 3;8(3):352–8.
2. Cohen R, Roth FJ, Delgado E, Ahearn DG, Kalser MH. Fungal Flora of the Normal Human Small and Large Intestine. N Engl J Med. 1969 Mar 20;280(12):638–41.
3. Kullberg BJ, Arendrup MC. Invasive Candidiasis. Campion EW, editor. N Engl J Med. 2015 Oct 8;373(15):1445–56.
4. Pfaller MA, Diekema DJ. Epidemiology of invasive candidiasis: a persistent public health problem. Clin Microbiol Rev. 2007 Jan;20(1):133–63.
5. Brown GD, Denning DW, Gow NAR, Levitz SM, Netea MG, White TC. Hidden killers: human fungal infections. Sci Transl Med. 2012 Dec 19;4(165):165rv13.
6. Casadevall A. The Pathogenic Potential of a Microbe. mSphere. 2(1).
7. Krause W, Matheis H, Wulf K. FUNGÆMIA AND FUNGURIA AFTER ORAL ADMINISTRATION OF CANDIDA ALBICANS. Lancet. 1969 Mar 22;293(7595):598–9.
8. Odds FC, Brown AJP, Gow NAR. Candida albicans genome sequence: a platform for genomics in the absence of genetics. Genome Biol. 2004;5(7):230.
9. Wu W, Lockhart SR, Pujol C, Srikantha T, Soll DR. Heterozygosity of genes on the sex chromosome regulates Candida albicans virulence. Mol Microbiol. 2007 Jun;64(6):1587–604.
10. Naglik JR, Gaffen SL, Hube B. Candidalysin: discovery and function in Candida albicans infections. Curr Opin Microbiol. 2019 Jul 6;52:100–9.
11. Calderone RA, Fonzi WA. Virulence factors of Candida albicans. Trends Microbiol. 2001 Jul;9(7):327–35.
12. Basso V, D'Enfert C, Znaidi S, Bachellier-Bassi S. From Genes to Networks: The Regulatory Circuitry Controlling Candida albicans Morphogenesis. Curr Top Microbiol Immunol. 2019;422:61–99.
13. de Groot PWJ, Bader O, de Boer AD, Weig M, Chauhan N. Adhesins in human fungal pathogens: glue with plenty of stick. Eukaryot Cell. 2013 Apr;12(4):470–81.
14. Granger BL. Insight into the antiadhesive effect of yeast wall protein 1 of Candida albicans. Eukaryot Cell. 2012 Jun;11(6):795–805.
15. Wilson D, Naglik JR, Hube B. The Missing Link between Candida albicans Hyphal Morphogenesis and Host Cell Damage. PLoS Pathog. 2016;12(10):e1005867.
16. Westman J, Moran G, Mogavero S, Hube B, Grinstein S. Candida albicans Hyphal Expansion Causes Phagosomal Membrane Damage and Luminal Alkalinization. MBio. 2018;9(5).
17. May RC, Casadevall A. In Fungal Intracellular Pathogenesis, Form Determines Fate. MBio. 2018;9(5).
18. Martchenko M, Alarco A-M, Harcus D, Whiteway M. Superoxide dismutases in Candida albicans: transcriptional regulation and functional characterization of the

hyphal-induced SOD5 gene. Mol Biol Cell. 2004 Feb;15(2):456–67.
19. Citiulo F, Jacobsen ID, Miramón P, Schild L, Brunke S, Zipfel P, et al. Candida albicans scavenges host zinc via Pra1 during endothelial invasion. PLoS Pathog. 2012;8(6):e1002777.
20. Almeida RS, Brunke S, Albrecht A, Thewes S, Laue M, Edwards JE, et al. The Hyphal-Associated Adhesin and Invasin Als3 of Candida albicans Mediates Iron Acquisition from Host Ferritin. Mitchell AP, editor. PLoS Pathog. 2008 Nov 21;4(11):e1000217.
21. Finkel JS, Mitchell AP. Genetic control of Candida albicans biofilm development. Nat Rev Microbiol. 2011;9(2):109–18.
22. Vila T, Fonseca BB, DA Cunha MML, Dos Santos GRC, Ishida K, Barreto-Bergter E, et al. Candida albicans biofilms: comparative analysis of room-temperature and cryofixation for scanning electron microscopy. J Microsc. 2017;267(3):409–19.
23. Uppuluri P, Chaturvedi AK, Srinivasan A, Banerjee M, Ramasubramaniam AK, Köhler JR, et al. Dispersion as an important step in the Candida albicans biofilm developmental cycle. PLoS Pathog. 2010 Mar 26;6(3):e1000828.
24. Soll DR, Daniels KJ. Plasticity of Candida albicans Biofilms. Microbiol Mol Biol Rev. 2016;80(3):565–95.
25. Taff HT, Mitchell KF, Edward JA, Andes DR. Mechanisms of Candida biofilm drug resistance. Future Microbiol. 2013 Oct;8(10):1325–37.
26. Nett J, Lincoln L, Marchillo K, Massey R, Holoyda K, Hoff B, et al. Putative role of beta-1,3 glucans in Candida albicans biofilm resistance. Antimicrob Agents Chemother. 2007 Feb;51(2):510–20.
27. Marie C, White TC. Genetic Basis of Antifungal Drug Resistance. Curr Fungal Infect Rep. 2009 Sep 1;3(3):163–9.
28. Vila T, Romo JA, Pierce CG, McHardy SF, Saville SP, Lopez-Ribot JL. Targeting Candida albicans filamentation for antifungal drug development. Virulence. 2017;8(2):150–8.
29. Berman J, Sudbery PE. Candida Albicans: a molecular revolution built on lessons from budding yeast. Nat Rev Genet. 2002 Dec;3(12):918–30.
30. Sudbery PE. Growth of Candida albicans hyphae. Nat Rev Microbiol. 2011 Aug 16;9(10):737–48.
31. Kruppa M. Quorum sensing and Candida albicans. Mycoses. 2009 Jan;52(1):1–10.
32. Nobile CJ, Fox EP, Nett JE, Sorrells TR, Mitrovich QM, Hernday AD, et al. A recently evolved transcriptional network controls biofilm development in Candida albicans. Cell. 2012;148(1–2):126–38.
33. Fox EP, Bui CK, Nett JE, Hartooni N, Mui MC, Andes DR, et al. An expanded regulatory network temporally controls Candida albicans biofilm formation. Mol Microbiol. 2015 Jun;96(6):1226–39.
34. Glazier VE, Krysan DJ. Transcription factor network efficiency in the regulation of Candida albicans biofilms: it is a small world. Curr Genet. 2018;0(0):0.
35. Banerjee M, Thompson DS, Lazzell A, Carlisle PL, Pierce C, Monteagudo C, et al. UME6, a novel filament-specific regulator of Candida albicans hyphal extension and virulence. Mol Biol Cell. 2008 Apr;19(4):1354–65.
36. Bonhomme J, D'Enfert C. Candida albicans biofilms: building a heterogeneous,

drug-tolerant environment. Curr Opin Microbiol. 2013 Aug;16(4):398–403.
37. Skrzypek MS, Binkley J, Binkley G, Miyasato SR, Simison M, Sherlock G. The Candida Genome Database (CGD): incorporation of Assembly 22, systematic identifiers and visualization of high throughput sequencing data. Nucleic Acids Res. 2017;45(D1):D592–6.
38. Lu Y, Su C, Liu H. A GATA transcription factor recruits Hda1 in response to reduced Tor1 signaling to establish a hyphal chromatin state in Candida albicans. PLoS Pathog. 2012;8(4):e1002663.
39. Bastidas RJ, Heitman J, Cardenas ME. The protein kinase Tor1 regulates adhesin gene expression in Candida albicans. PLoS Pathog. 2009 Feb;5(2):e1000294.
40. Brown DH, Giusani AD, Chen X, Kumamoto CA. Filamentous growth of Candida albicans in response to physical environmental cues and its regulation by the unique CZF1 gene. Mol Microbiol. 1999 Nov;34(4):651–62.
41. Stoldt VR, Sonneborn A, Leuker CE, Ernst JF. Efg1p, an essential regulator of morphogenesis of the human pathogen Candida albicans, is a member of a conserved class of bHLH proteins regulating morphogenetic processes in fungi. EMBO J. 1997 Apr 15;16(8):1982–91.
42. Lo HJ, Köhler JR, Didomenico B, Loebenberg D, Cacciapuoti A, Fink GR. Nonfilamentous *C. albicans* mutants are avirulent. Cell. 1997;90(5):939–49.
43. Giusani AD, Vinces M, Kumamoto CA. Invasive filamentous growth of Candida albicans is promoted by Czf1p-dependent relief of Efg1p-mediated repression. Genetics. 2002 Apr;160(4):1749–53.
44. Ramage G, VandeWalle K, López-Ribot JL, Wickes BL. The filamentation pathway controlled by the Efg1 regulator protein is required for normal biofilm formation and development in Candida albicans. FEMS Microbiol Lett. 2002 Aug 27;214(1):95–100.
45. Nett JE, Lepak AJ, Marchillo K, Andes DR. Time course global gene expression analysis of an in vivo Candida biofilm. J Infect Dis. 2009 Jul 15;200(2):307–13.
46. Rajendran R, May A, Sherry L, Kean R, Williams C, Jones BL, et al. Integrating Candida albicans metabolism with biofilm heterogeneity by transcriptome mapping. Sci Rep. 2016;6(October):1–11.
47. Stichternoth C, Ernst JF. Hypoxic adaptation by Efg1 regulates biofilm formation by Candida albicans. Appl Environ Microbiol. 2009 Jun;75(11):3663–72.
48. Sonneborn A, Tebarth B, Ernst JF. Control of white-opaque phenotypic switching in Candida albicans by the Efg1p morphogenetic regulator. Infect Immun. 1999 Sep;67(9):4655–60.
49. Uhl MA, Biery M, Craig N, Johnson AD. Haploinsufficiency-based large-scale forward genetic analysis of filamentous growth in the diploid human fungal pathogen C.albicans. EMBO J. 2003 Jun 2;22(11):2668–78.
50. Homann OR, Dea J, Noble SM, Johnson AD. A phenotypic profile of the Candida albicans regulatory network. PLoS Genet. 2009 Dec;5(12):e1000783.
51. Du H, Guan G, Xie J, Sun Y, Tong Y, Zhang L, et al. Roles of Candida albicans Gat2, a GATA-type zinc finger transcription factor, in biofilm formation, filamentous growth and virulence. PLoS One. 2012;7(1):e29707.
52. Nobile CJ, Mitchell AP. Regulation of cell-surface genes and biofilm formation by

the C. albicans transcription factor Bcr1p. Curr Biol. 2005 Jun 21;15(12):1150–5.
53. López-Ribot JL. Candida albicans biofilms: more than filamentation. Curr Biol. 2005 Jun 21;15(12):R453-5.
54. Srikantha T, Daniels KJ, Pujol C, Kim E, Soll DR. Identification of genes upregulated by the transcription factor Bcr1 that are involved in impermeability, impenetrability, and drug resistance of Candida albicans a/α biofilms. Eukaryot Cell. 2013 Jun;12(6):875–88.
55. Jung S-I, Finkel JS, Solis N V, Chaili S, Mitchell AP, Yeaman MR, et al. Bcr1 functions downstream of Ssd1 to mediate antimicrobial peptide resistance in Candida albicans. Eukaryot Cell. 2013 Mar;12(3):411–9.
56. Dwivedi P, Thompson A, Xie Z, Kashleva H, Ganguly S, Mitchell AP, et al. Role of Bcr1-activated genes Hwp1 and Hyr1 in Candida albicans oral mucosal biofilms and neutrophil evasion. PLoS One. 2011 Jan 25;6(1):e16218.
57. Fanning S, Xu W, Solis N, Woolford CA, Filler SG, Mitchell AP. Divergent targets of Candida albicans biofilm regulator Bcr1 in vitro and in vivo. Eukaryot Cell. 2012 Jul;11(7):896–904.
58. Holland LM, Schröder MS, Turner SA, Taff H, Andes D, Grózer Z, et al. Comparative phenotypic analysis of the major fungal pathogens Candida parapsilosis and Candida albicans. PLoS Pathog. 2014 Sep;10(9):e1004365.
59. Lu Y, Su C, Solis N V, Filler SG, Liu H. Synergistic regulation of hyphal elongation by hypoxia, CO(2), and nutrient conditions controls the virulence of Candida albicans. Cell Host Microbe. 2013 Nov 13;14(5):499–509.
60. Lu Y, Su C, Liu H. Candida albicans hyphal initiation and elongation. Trends Microbiol. 2014;22(12):707–14.
61. Glazier VE, Murante T, Murante D, Koselny K, Liu Y, Kim D, et al. Genetic analysis of the Candida albicans biofilm transcription factor network using simple and complex haploinsufficiency. PLoS Genet. 2017 Aug;13(8):e1006948.
62. Sorrells TR, Johnson AD. Making sense of transcription networks. Cell. 2015 May 7;161(4):714–23.
63. Müller D, Stelling J. Precise regulation of gene expression dynamics favors complex promoter architectures. PLoS Comput Biol. 2009 Jan;5(1):e1000279.
64. Hirakawa MP, Martinez DA, Sakthikumar S, Anderson MZ, Berlin A, Gujja S, et al. Genetic and phenotypic intra-species variation in Candida albicans. Genome Res. 2015 Mar;25(3):413–25.
65. MacCallum DM, Castillo L, Nather K, Munro CA, Brown AJP, Gow NAR, et al. Property differences among the four major Candida albicans strain clades. Eukaryot Cell. 2009 Mar;8(3):373–87.
66. Karababa M, Coste AT, Rognon B, Bille J, Sanglard D. Comparison of gene expression profiles of Candida albicans azole-resistant clinical isolates and laboratory strains exposed to drugs inducing multidrug transporters. Antimicrob Agents Chemother. 2004 Aug;48(8):3064–79.
67. Morschhäuser J, Barker KS, Liu TT, Blaß-Warmuth J, Homayouni R, Rogers PD. The transcription factor Mrr1p controls expression of the MDR1 efflux pump and mediates multidrug resistance in Candida albicans. PLoS Pathog. 2007;3(11):1603–16.
68. Hasan F, Xess I, Wang X, Jain N, Fries BC. Biofilm formation in clinical Candida

isolates and its association with virulence. Microbes Infect. 2009 Jul;11(8–9):753–61.
69. Rajendran R, Sherry L, Nile CJ, Sherriff A, Johnson EM, Hanson MF, et al. Biofilm formation is a risk factor for mortality in patients with Candida albicans bloodstream infection-Scotland, 2012-2013. Clin Microbiol Infect. 2016 Jan;22(1):87–93.
70. Sherry L, Rajendran R, Lappin DF, Borghi E, Perdoni F, Falleni M, et al. Biofilms formed by Candida albicans bloodstream isolates display phenotypic and transcriptional heterogeneity that are associated with resistance and pathogenicity. BMC Microbiol. 2014 Jul 5;14(1):182.
71. Odds FC, Jacobsen MD. Multilocus sequence typing of pathogenic Candida species. Eukaryot Cell. 2008 Jul;7(7):1075–84.
72. Bennett RJ, Johnson AD. Completion of a parasexual cycle in Candida albicans by induced chromosome loss in tetraploid strains. EMBO J. 2003 May 15;22(10):2505–15.
73. Gräser Y, Volovsek M, Arrington J, Schönian G, Presber W, Mitchell TG, et al. Molecular markers reveal that population structure of the human pathogen Candida albicans exhibits both clonality and recombination. Proc Natl Acad Sci U S A. 1996 Oct 29;93(22):12473–7.
74. Gasch AP, Payseur BA, Pool JE. The Power of Natural Variation for Model Organism Biology. Trends Genet. 2016 Mar;32(3):147–54.
75. Chin BL, Ryan O, Lewitter F, Boone C, Fink GR. Genetic variation in Saccharomyces cerevisiae: circuit diversification in a signal transduction network. Genetics. 2012 Dec;192(4):1523–32.
76. Dallman TJ, Ashton PM, Byrne L, Perry NT, Petrovska L, Ellis R, et al. Applying phylogenomics to understand the emergence of Shiga-toxin-producing Escherichia coli O157:H7 strains causing severe human disease in the UK. Microb genomics. 2015 Sep;1(3):e000029.
77. Marcus SL, Brumell JH, Pfeifer CG, Finlay BB. Salmonella pathogenicity islands: big virulence in small packages. Microbes Infect. 2000 Feb;2(2):145–56.
78. Dowell RD, Ryan O, Jansen A, Cheung D, Agarwala S, Danford T, et al. Genotype to phenotype: a complex problem. Science. 2010 Apr 23;328(5977):469.
79. Vu V, Verster AJ, Schertzberg M, Chuluunbaatar T, Spensley M, Pajkic D, et al. Natural Variation in Gene Expression Modulates the Severity of Mutant Phenotypes. Cell. 2015;162(2):391–402.
80. Lin X, Huang JC, Mitchell TG, Heitman J. Virulence attributes and hyphal growth of C. neoformans are quantitative traits and the MATalpha allele enhances filamentation. PLoS Genet. 2006 Nov 17;2(11):e187.
81. Gallagher JEG, Zheng W, Rong X, Miranda N, Lin Z, Dunn B, et al. Divergence in a master variator generates distinct phenotypes and transcriptional responses. Genes Dev. 2014;28(4):409–21.
82. Zheng W, Gianoulis TA, Karczewski KJ, Zhao H, Snyder M. Regulatory Variation Within and Between Species. Annu Rev Genomics Hum Genet. 2011;12(1):327–46.
83. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From

Polygenic to Omnigenic. Cell. 2017 Jun 15;169(7):1177–86.
84. Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. Nat Rev Genet. 2015 Apr 6;16(4):197–212.
85. Pai AA, Pritchard JK, Gilad Y. The Genetic and Mechanistic Basis for Variation in Gene Regulation. PLoS Genet. 2015;11(1).
86. Wilson RB, Davis D, Mitchell AP. Rapid hypothesis testing with Candida albicans through gene disruption with short homology regions. J Bacteriol. 1999 Mar;181(6):1868–74.
87. Noble SM, Johnson AD. Strains and strategies for large-scale gene deletion studies of the diploid human fungal pathogen Candida albicans. Eukaryot Cell. 2005 Feb;4(2):298–309.
88. Papon N, Courdavault V, Clastre M, Simkin AJ, Crèche J, Giglioli-Guivarc'h N. Deus ex Candida genetics: Overcoming the hurdles for the development of a molecular toolbox in the CTG clade. Vol. 158, Microbiology. 2012. p. 585–600.
89. Basso LR, Bartiss A, Mao Y, Gast CE, Coelho PSR, Snyder M, et al. Transformation of Candida albicans with a synthetic hygromycin B resistance gene. Yeast. 2010 Dec;27(12):1039–48.
90. Samaranayake DP, Hanes SD. Milestones in Candida albicans gene manipulation. Fungal Genet Biol. 2011 Sep;48(9):858–65.
91. Reuss O, Vik A, Kolter R, Morschhäuser J. The SAT1 flipper, an optimized tool for gene disruption in Candida albicans. Gene. 2004 Oct 27;341(1–2):119–27.
92. Alani E, Cao L, Kleckner N. A method for gene disruption that allows repeated use of URA3 selection in the construction of multiply disrupted yeast strains. Genetics. 1987 Aug;116(4):541–5.
93. Fonzi WA, Irwin MY. Isogenic strain construction and gene mapping in Candida albicans. Genetics. 1993 Jul;134(3):717–28.
94. Wilson RB, Davis D, Enloe BM, Mitchell AP. A recyclable Candida albicans URA3 cassette for PCR product-directed gene disruptions. Yeast. 2000 Jan 15;16(1):65–70.
95. Wellington M, Anaul M, Rustchenko KE. 5-fluoro-orotic acid induces chromosome alterations in genetically manipulated strains of Candida albicans. Mycologia. 2006 May 1;98(3):393–8.
96. Staab JF, Sundstrom P. URA3 as a selectable marker for disruption and virulence assessment of Candida albicans genes. Trends Microbiol. 2003;11(2):69–73.
97. Dennison PMJ, Ramsdale M, Manson CL, Brown AJP. Gene disruption in Candida albicans using a synthetic, codon-optimised Cre-loxP system. Fungal Genet Biol. 2005 Sep;42(9):737–48.
98. Abbey D, Hickman M, Gresham D, Berman J. High-Resolution SNP/CGH Microarrays Reveal the Accumulation of Loss of Heterozygosity in Commonly Used Candida albicans Strains. G3 (Bethesda). 2011 Dec;1(7):523–30.
99. Enloe B, Diamond A, Mitchell AP. A single-transformation gene function test in diploid Candida albicans. J Bacteriol. 2000 Oct;182(20):5730–6.
100. Sternberg SH, Doudna JA. Expanding the Biologist's Toolkit with CRISPR-Cas9. Mol Cell. 2015 May 21;58(4):568–74.
101. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial

immunity. Science. 2012 Aug 17;337(6096):816–21.
102. Vyas VK, Barrasa MI, Fink GR. A Candida albicans CRISPR system permits genetic engineering of essential genes and gene families. Sci Adv. 2015;1(3):e1500248.
103. Min K, Ichikawa Y, Woolford CA, Mitchell AP. Candida albicans Gene Deletion with a Transient CRISPR-Cas9 System. mSphere. 2016;1(3):1–9.

Chapter 2: Marker recycling in *Candida albicans* through CRISPR-Cas9-induced marker excision

Manning Y. Huang and Aaron P. Mitchell

FOREWORD

This chapter is lightly modified from the publication version as appeared in mSphere, March 2017. All referenced figures are available at the end of the chapter.

Supplemental files referenced in this chapter are available online at:

https://doi.org/10.1128/mSphere.00050-17

INTRODUCTION

Engineered genetic manipulations almost always require selection markers, and for many organisms only a few markers are useful. The spectrum of selection markers may be limited by an organism's intrinsic resistance to drugs, the complexity of medium formulations, phenotypic impact of a growth requirement, or other factors. Therefore, it is helpful to be able to use a single selection marker repeatedly. The repeated use of the same marker for genetic constructs that are integrated stably in the genome is achieved through an approach called marker recycling, in which a strategy to promote or detect loss of a marker can be applied after the initial selection for the marker.

Marker recycling has been achieved through two general approaches: positive/negative selection or recombinase-promoted excision. In the positive/negative selection approach, a marker cassette is used that permits growth under one condition and prevents growth under another condition. The cassette includes flanking directly repeated sequences that allow low-frequency homologous recombination events to excise the marker, leaving behind one copy of the repeated sequence. This approach was popularized with the development of the "Ura-blaster" for *Saccharomyces cerevisiae*, which was rapidly adapted for use in other fungi (1, 2). In the recombinase-promoted excision approach, the marker cassette includes both a selection marker and an inducible site-specific recombinase gene. Target sites for the recombinase lie at the ends of the cassette, so that induction of the recombinase causes high-frequency

excision of the marker cassette. In essence, the activity of the recombinase increases the cassette excision rate sufficiently so that no selection against the marker cassette is necessary to detect loss events. This approach was popularized with the development of the "*SAT1* flipper" in *Candida albicans* (3), and is related to the Cre-lox system, used to create conditional knockouts in mouse (4).

Here we present a marker recycling approach that builds upon CRISPR-Cas9 systems. These systems use a programmable nuclease to make a targeted double strand break in DNA (5). Targeting is accomplished by base-pairing between one genomic DNA strand and the single-guide RNA (sgRNA) that is complexed with the Cas9 nuclease (5, 6). An investigator can choose the site at which a double strand break will be induced by designing an appropriate sgRNA. We apply our approach to the fungal pathogen *C. albicans* for proof of principle. *C. albicans* is extremely important clinically (7), and presents challenging genetics because it is naturally diploid and lacks a complete sexual cycle (8). In most cases, a recessive loss-of-function mutation must be homozygous in order to manifest a prominent phenotype, so gene function analysis in this organism has typically required at least two successive transformations. The creation of homozygous mutants was accelerated dramatically through the work of Vyas et al., who developed a CRISPR-Cas system for *C. albicans* (9). They showed that homozygous mutations in one or even several genes could be created in a single transformation. We previously modified their system to create complete gene deletion mutations, and found that the genes specifying Cas9 and the sgRNA could be introduced into cells transiently and without direct selection (10). The marker recycling

approach we describe here is tested specifically in *C. albicans*, and is based upon the general properties of CRISPR-Cas9 systems and the native recombination and repair machinery of the cell. Because CRISPR-Cas9 systems have been deployed in a broad spectrum of organisms, we believe that our marker recycling strategy may be generally useful.

RESULTS

Rationale for CRISPR-Cas9-induced marker excision.

A double strand break in a genomic region flanked by directly repeated sequences should yield a deletion that fuses the flanking repeats (Figure 1). This expectation is founded on the pioneering study by Sugawara and Haber of break-induced recombination (11). We reasoned that this recombination process should allow loss, through excision, of any selection marker. Moreover, if a cell was homozygous for the entire region depicted, then both alleles could undergo the same recombination process, provided that both alleles were subjected to a double strand break.

Such marker excision events could be implemented with the use of CRISPR-Cas9 to create a marker recycling system. Consider that an investigator seeks to make a *C. albicans* strain with homozygous deletion mutations in three genes - *YFG1, YFG2,* and *YFG3* - and can use only two selection markers, *M1* and *M2*. The specific marker cassettes would include flanking direct repeats, and we call the cassettes *rM1r* and *rM2r*. The construction of the homozygous triple mutant could be accomplished in three successive transformations (Figure 2). In the first transformation, the *YFG1* gene is replaced with *yfg1Δ::rM1r* at both alleles. Biallelic replacement is accomplished by including in the transformation mix genes that specify Cas9 and a *YFG1*-targeting sgRNA along with the *yfg1Δ::rM1r* repair template. The second transformation is carried out with the strain that resulted from the first transformation. In the second transformation, the *YFG2* gene is replaced with *yfg2Δ::rM2r* at both alleles, and in addition the *rM1r* marker is excised to leave behind *yfg1Δ::r* at both *yfg1Δ* alleles.

These two biallelic events are accomplished by including in the transformation mix genes that specify Cas9, a *YFG2*-targeting sgRNA, and an *M1*-targeting sgRNA, along with the *yfg2Δ::rM2r* repair template. Hence the resulting strain lacks the *M1* marker, so that marker can be used for selection again. This strain is used for the third transformation. In the third transformation, the *YFG3* gene is replaced with *yfg3Δ::rM1r* at both alleles, and in addition the *rM2r* marker is excised to leave behind *yfg2Δ::r* at both *yfg2Δ* alleles. These two biallelic events are accomplished analogously to those of the second transformation. Specifically, the transformation mix includes genes that specify Cas9, a *YFG3*-targeting sgRNA, and an *M2*-targeting sgRNA, along with the *yfg3Δ::rM1r* repair template. These three transformations yield a *yfg1 yfg2 yfg3* homozygous triple deletion mutant that carries the *M1* marker but lacks the *M2* marker, so that the *M2* marker can be used for selection again. We refer to each marker excision step (the conversion of *rM1r* to *r*, or the conversion of *rM2r* to *r*) as CRISPR-Cas9-Induced Marker Excision, which we abbreviate CRIME.

Application of CRIME.

To see whether CRIME may work in practice, we set out to create a homozygous *ume6Δ brg1Δ bcr1Δ* triple mutant. Each of the genes chosen for deletion is a positive regulator of filamentation and biofilm formation (12–14). We used the popular strain SN152, which is homozygous for the mutations *his1*Δ and *leu2*Δ (15). (It is also homozygous for *arg4*Δ, but we did not use the *ARG4* gene in our studies.) The *yfgΔ::rMr* repair templates comprised two overlapping PCR products, each with a single "r" repeat sequence, to create a split-marker template (16) as detailed in Methods.

(Note that our split-marker transformations include two overlapping fragments of the repair template, and their final assembly requires cellular recombination machinery.) There were different flanking repeat sequences for *r1HIS1r1* and *r2LEU2r2* in order to minimize the possibility of recombinational interaction between the cassettes. The *r1HIS1r1* marker included flanking repeats of 360 bp derived from the vector pRS424 (17). The *r2LEU2r2* marker included flanking repeats of 252 bp derived from the vector YEp24 (18). These materials allowed us to carry out the triple mutant strain construction outlined above.

In Construction 1, we created a homozygous *ume6Δ::r1HIS1r1* mutant (Table 1). All transformations included the gene specifying Cas9. Inclusion of the split-marker template yielded His+ transformants (compare transformations 1 and 2 in Table 1). Inclusion of a *UME6*-targeting sgRNA increased the recovery of selected His+ transformants considerably (compare transformations 2 and 3 in Table 1), as expected if the Cas9-sgRNA complex is functional. PCR genotyping (Figure 3A) showed that 4 out of 10 transformants tested were homozygous for the *ume6Δ::r1HIS1r1* mutation (Figure 3B, isolates 1, 2, 9, 10). This frequency of homozygous marked deletion mutants was similar to what we found previously (10).

In Construction 2, we used a *ume6Δ::r1HIS1r1* homozygous strain as a recipient to introduce a homozygous *brg1Δ::r2LEU2r2* mutation, and simultaneously used CRIME to convert the *ume6Δ::r1HIS1r1* alleles to *ume6Δ::r1* alleles. With a split-marker repair template, inclusion of a *BRG1*-targeting sgRNA increased the recovery of selected Leu+

transformants (compare transformations 5 and 6 in Table 1). Inclusion of an additional *HIS1*-targeting sgRNA resulted in 4-19% of the Leu+ transformants being His-, depending upon the amount of the HIS1 sgRNA gene (compare transformations 7-9 in Table 1). PCR genotyping (Figure 3E) indicated that 13 of 14 His- transformants tested were homozygous for a single repeat sequence marking the *ume6Δ::r1* alleles (Figure 3F, isolates 1-13). In addition, 12 of 14 His- transformants tested were homozygous for the *brg1Δ::r2LEU2r2* mutation (Figure 3C, isolates 1-9, 11, 12, 14). The results of this construction show that CRIME allows recycling of the *r1HIS1r1* marker cassette.

In Construction 3, we used a *ume6Δ::r1 brg1Δ::r2LEU2r2* homozygous double mutant strain as a recipient to introduce a homozygous *bcr1Δ::r1HIS1r1* mutation, and simultaneously used CRIME to convert the *brg1Δ::r2LEU2r2* alleles to *brg1Δ::r2* alleles. The *r1HIS1r1* cassette we used to select for the *bcr1Δ* allele was the same cassette we used in Construction 1 to select for the *ume6Δ* allele. Once again a split-marker repair template was employed, and inclusion of a *BCR1*-targeting sgRNA increased the recovery of selected His+ transformants (compare transformations 11 and 12 in Table 1). Inclusion of an additional *LEU2*-targeting sgRNA resulted in 47% of the His+ transformants being Leu- (transformation 13 in Table 1). PCR genotyping indicated that 11 out of 12 Leu- transformants tested were homozygous for a single repeat sequence marking the *brg1Δ::r2* alleles (Figure 3G, isolates 1, 3-12). In addition, 7 out of 12 Leu- transformants tested were homozygous for the *bcr1Δ::r1HIS1r1* mutation (Figure 3D, isolates 3-6, 8-10). The results of this construction show that CRIME allows recycling of the *r2LEU2r2* marker cassette.

DISCUSSION

We have presented a new approach to marker recycling. Marker recycling has played an important role in genetic manipulation, as illustrated by the hundreds of citations to previous descriptions of marker recycling approaches (1–3). Use of these strategies is especially prominent in fungal studies, where the number of selection markers may be limited (19), and where nutritional requirements can impact diverse phenotypes. Our CRIME approach is conceptually a hybrid between the positive/negative selection strategy and the recombinase-promoted excision strategy. Like the positive/negative selection strategy, CRIME makes use of the cell's native recombination and repair machinery to excise the DNA between directly repeated sequences. Like the recombinase-promoted excision strategy, CRIME makes use of controlled DNA cleavage events to increase the frequency of recombination in a specific genomic region. CRIME has one major advantage over the prior strategies: speed. This point is illustrated by the fact that each construction in Table 1 required just over 1 week from start to finish, including time for genotyping. In effect, an investigator can use CRIME to steal a little extra time.

The ability of CRISPR-Cas9 systems to be multiplexed is of critical importance for the CRIME strategy. Specifically, Cas9 nuclease subunits can interact with multiple different sgRNAs to target multiple genomic sites for cleavage (9). This multiplexing capability is exploited by CRIME in use of a single transformation for both the deletion of one gene and the recycling of the previously used selection marker. One feature we saw consistently was that inclusion of a second sgRNA reduced the frequency of the

transformant class promoted by the first sgRNA (transformations 6-9 and 12-13 in Table 1). These results are expected if there is competition between two sgRNAs or their respective genes. Recognition of the competition phenomenon should prompt investigators to try a range of sgRNA gene concentrations in multiplexed transformations.

Our detailed method employs transformation mixes that contain only PCR products and not cloned DNA segments. The approach builds upon the rapid transient CRISPR-Cas9 approach (10). The use of PCR products saves time compared to cloning-dependent genetic approaches.

The recombinational marker excision event in CRIME seems to be efficient. In our two examples (Figure 3), 13/14 and 11/12 marker loss events occurred through excision between repeated sequences at both alleles. It is well known that excision between directly repeated regions of homology can be used for double strand break repair in human cells (20) and yeast cells (21). In the yeast *S. cerevisiae*, the single-strand annealing pathway is operative when repeats are 200 bp or more in length; the microhomology-mediated end joining pathway is used when repeats are 5-25 bp in length (21). Thus the simplest hypothesis is that our CRIME system uses the single-strand annealing pathway. The most important observation in this regard is that small mutations that inactivate the marker gene were rare in our studies. The efficiency of marker loss events from CRIME is important in order to reuse a marker cassette to

target a new locus, because extensive homology in the genome might promote integration of the marker at the mutant alleles created previously.

One feature of CRIME that may be viewed as a weakness compared to other marker recycling methods is that two markers are required for CRIME, and only one is required for the Ura-blaster and *SAT1* flipper. This consideration will have to be weighed against the time-saving advantage of CRIME in choosing a method to use. A second consideration, more relevant for other fungi than for *C. albicans*, is the relative frequency of marker loss through excision between repeats as opposed to indels or more complex rearrangements. The first report of CRISPR-Cas9 usage in *Aspergillus fumigatus* presented the startling result that inactivation of a targeted gene was often accompanied by nonhomologous integration of input DNA, in particular the sgRNA gene, at the break site (22). Whether this event would still occur predominantly if flanking repeats were present is unknown. These points illustrate that it is useful to have a few different approaches for any genetic manipulation, because biology and technology often have to reach a compromise when mutations are engineered.

One striking observation is that the efficiency of CRISPR-Cas9-promoted integration seemed to vary widely. Integration of the *brg1Δ* construct was more efficient than the *ume6Δ* construct, and both were more efficient than *bcr1Δ* construct. These transformations are not precisely comparable, though, because they employed different markers and selections, different strains, and different sgRNAs. It may be useful to compare sgRNA efficiencies under parallel conditions, as has been done in human cells

(see (23) for example), to see if sgRNA design principles pertinent to *C. albicans* can be deduced.

When we look toward future genetic studies of *C. albicans*, we have a recommendation. Our recommendation is that newly created deletion alleles should be made with repeat-flanked marker cassettes such as *r1HIS1r1* or *r2LEU2r2.* Many investigators create double or triple mutant strains in which deletion mutations are combined to provide an appraisal of pathway relationships or functional redundancy. In the past in our lab, multi-mutant strain constructions often begin with remaking a single mutant using a recyclable marker cassette. If most mutant strains in most labs were initially made with recyclable cassettes, then it would be unnecessary to remake mutant strains for genetic interaction studies.

CRISPR-Cas9 systems have been implemented in numerous organisms (9, 22, 24). We suggest that the CRIME approach to marker recycling may be useful in many organisms as well. It relies upon general features of CRISPR-Cas9 systems as well native double strand break repair machinery, which is highly conserved. Therefore, CRIME seems poised to be applied to diverse genetic systems.

METHODS

Strains and culture conditions.

All yeast strains are listed in the Supplementary Data file. Strains were grown at 30°C in YPD+uri (2% Bacto peptone, 2% dextrose, 1% yeast extract, and 80ug/ml uridine) with shaking. *C. albicans* transformants were selected on CSM plates lacking either histidine or leucine. All strains were saved as frozen stocks at -80°C in 15% glycerol. All transformations were performed with the lithium acetate transformation method (25) with DNA quantities as previously described (10).

Plasmids and DNA

A) Overview of Partner Plasmids for CRIME Markers:

All primers are listed in the Supplementary Data file, along with DNA sequences for plasmids pMH01-04 We utilized a strategy built around split-marker recombination (16) to generate direct repeat flanked marker cassettes. Briefly, two plasmids, derived from the same parent, each contain a selectable marker introduced at different restriction sites (Figure 4A). PCR products that each contain only segments of the whole marker are amplified from these partner plasmids (Figure 4A). One product contains at its 5' end an 80 bp region of homology to the upstream region of the gene of interest introduced by a primer. This is followed by the repeat sequence and an incomplete segment of the selectable marker (Figure 4B). The other product contains at its 5' end an incomplete segment of the selectable marker and the repeat sequence. This is followed by an 80 bp region of homology to the downstream region of the gene of interest introduced by another primer (Figure 4B). The two amplicons reconstitute the

complete direct-repeat flanked marker *in situ* via split-marker recombination following transformation (Figure 4B).

B) Partner Plasmids:

To construct pMH01 and pMH02, each containing the *Candida dubliniensis HIS1* gene, we used the following methods. An aliquot of 1 ug of pRS424 (17) plasmid DNA was digested with the restriction enzyme KpnI, which was then heat inactivated. A second aliquot of pRS424 plasmid DNA was digested with the restriction enzyme SapI, which was then heat inactivated.

To make partner plasmid pMH01 (sequence in Supplementary Data file), a 2.3 kb fragment containing the *Candida dubliniensis HIS1* gene was amplified by PCR from pSN52 (15) using primers KpnI_pRS424_H+AdapN/F and KpnI_pRS424_H+AdapN/R. 4 ul of the PCR product were co-transformed with 1 ul of pRS424 cut with KpnI into *S. cerevisiae* strain BJ8918 (26) to insert the *HIS1* gene into pRS424 at the KpnI restriction site via gap repair (Figure 4A).

To make partner plasmid pMH02 (sequence in Supplementary Data file), a 2.3 kb fragment containing the *Candida dubliniensis HIS1* gene was amplified by PCR from pSN52 using primers SapI_pRS424_H+AdapN/F and SapI_pRS424_H+AdapN/R. 4 ul of the PCR product were co-transformed with 1 ul of pRS424 cut with SapI into strain BJ8918 to insert the *HIS1* gene into pRS424 at the SapI restriction site via gap repair (Figure 4A).

Plasmids were recovered from BJ8918 using the Zymoprep™ Yeast Plasmid Miniprep II kit.

To construct pMH03 and pMH04 (sequences in Supplementary Data file), each containing the *Candida maltosa LEU2* gene, we used the following methods. An aliquot of YEp24 (18) was digested with the restriction enzyme BamHI, followed by heat inactivation. A second aliquot of pRS424 plasmid DNA was digested with the restriction enzyme SalI, followed by heat inactivation.

To make pMH03, a 2.2 kb fragment containing the *Candida maltosa LEU2* gene was amplified by PCR from pSN40 using primers BamHI_YEp24_H+AdapN/F and BamHI_YEp24_H+AdapN/R. 4 ul of the PCR product were co-transformed with 1 ul of pRS424 cut with BamHI into strain BJ8918 (26) to insert the *LEU2* gene into YEp24 at the BamHI restriction site via gap repair.

To make pMH04, a 2.2 kb fragment containing the *Candida maltosa LEU2* gene was amplified by PCR from pSN40 using primers SalI_YEp24_H+AdapN/F and SalI_YEp24_H+AdapN/R. 4 ul of the PCR product were co-transformed with 1 ul of pRS424 cut with SalI into strain BJ8918 to insert the *LEU2* gene into YEp24 at the SapI restriction site via gap repair.

Plasmids were again recovered from BJ8918 using the Zymoprep™ Yeast Plasmid Miniprep II kit.

C) CRIME Markers

The *ume6Δ::r1HIS1r1* cassette was amplified from pMH01 and pMH02. The aft product was generated by amplification from pMH01 using primers UME6-SapI/R, which contains an 80bp segment of homology downstream of *UME6*, and HIS1 CRIME/F. The fore product was generated by amplification from pMH02 using primers UME6-KpnI/F, containing 80bp of homology upstream of *UME6*, and HIS1 CRIME/R.

The *bcr1Δ::r1HIS1r1* cassette was amplified from pMH01 and pMH02. The aft product was generated by amplification from pMH01 using primers BCR1-SapI/R, which contains an 80bp segment of homology downstream of *BCR1*, and UME6-SapI/R. The fore product was generated by amplification from pMH02 using primers BCR1-KpnI/F, containing 80bp of homology upstream of *BCR1*, and HIS1 CRIME/R.

The *brg1Δ::r2LEU2r2* cassette was amplified from pMH03 and pMH04. The aft product was generated by PCR amplification from pMH03 using primers BRG1-SalI/R, containing an 80 bp segment of homology downstream of *BRG1*, and LEU2 CRIME/F. The fore product was generated by PCR amplification from pMH04 using primers BRG1-BamHI/F, containing an 80 bp segment of homology upstream of *BRG1*, and LEU2 CRIME/R.

D) Other DNA cassettes:

The approximately 5kb CaCas9 cassette containing an *ENO1* promoter, the CaCas9 ORF, and a *CYC1* terminator was amplified from pV1093 (9) using primers CaCas9/For and CaCas9/Rev. The sgRNA cassettes for *UME6*, *BRG1*, *BCR1*, *C. maltosa HIS1*, and *C. dubliniensis LEU2* were amplified via split joint PCR as previously described (10) using primers UME6-sgRNA/F and UME6-sgRNA/R; BRG1-sgRNA/F and BRG1-sgRNA/R; BCR1-sgRNA/F and BCR1-sgRNA/R; Cd.HIS1-sgRNA/F and Cd.HIS1-sgRNA/R; and Cm.LEU2-sgRNA/F and Cm.LEU2-sgRNA/R respectively. The methods described in Min et al (10) may be summarized as follows: YFG single guide RNA sequences were first selected, either from the *Candida albicans* CRISPR target sequence database kindly supplied by Vyas et al. (*UME6*, *BRG1*, *BCR1*) (9), or otherwise designed by hand (*C. maltosa HIS1* and *C. dubliniensis LEU2*) (9). The guide sequence was designed into the YFG-sgRNA/F primer sequence by removing the NGG PAM sequence and adding our sgRNA scaffold adapter sequence in its place.
e.g.) 5'-[YFG target without PAM]GTTTTAGAGCTAGAAATAGCAAGTTAAA-3'
The YFG-SNR52/R primer sequence was designed with the reverse complement.
e.g.) 5'-[reverse complement]CAAATTAAAAATAGTTTACGCAAGTC-3'

The promoter region was then amplified via PCR with primers SNR52/F and YFG-SNR52/R, while the scaffold and terminator regions was amplified via PCR with primers YFG-sgRNA/F and sgRNA/R. Standard TAKARA Ex-Taq protocols were applied for this reaction. Products were then purified using the protocols and materials provided in the Thermo Fisher GeneJet PCR Purification kit.

The second round of PCR was roughly modified from the standard TAKARA Ex-Taq protocol. To join the *SNR52* promoter amplicon to the sgRNA scaffold and terminator amplicon, equimolar quantities of each amplicon (up to 1000 ng) were combined roughly as follows: 2.5 µl of purified *SNR52* promoter amplicon, 2.5 µl of purified sgRNA amplicon, 2.5 µl of 10X TAKARA Taq Buffer, 2.0 µl of dNTPs, and 0.25 µl of TAKARA ExTaq polymerase, with double distilled water to a total volume of 25 µl.

The second round of PCR ran for 10 cycles with 30 seconds at melting temperature, 10 minutes at 58°C to anneal the two amplicons, and a 5 minute elongation phase. 1 µl of the second round PCR product was then amplified in a third round of PCR with primers SNR52/N and sgRNA/N using standard protocols.

ACKNOWLEDGMENTS

We are grateful to Tatyana Aleynikov for expert technical assistance and lab management, to Fred Lanni for encouragement, and to Katie Lagree and Carol Woolford for many insightful discussions and very helpful comments on this manuscript.

LITERATURE CITED, CHAPTER 2

1. Alani E, Cao L, Kleckner N. 1987. A method for gene disruption that allows repeated use of URA3 selection in the construction of multiply disrupted yeast strains. Genetics 116:541–545.
2. Fonzi WA, Irwin MY. 1993. Isogenic strain construction and gene mapping in Candida albicans. Genetics 134:717–728.
3. Reuß O, Vik Å, Kolter R, Morschhäuser J. 2004. The SAT1 flipper, an optimized tool for gene disruption in Candida albicans. Gene 341:119–127.
4. Sauer B, Henderson N. 1988. Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. Proc Natl Acad Sci U S A 85:5166–70.
5. Sternberg SH, Doudna JA. 2015. Expanding the Biologist's Toolkit with CRISPR-Cas9. Mol Cell 58:568–574.
6. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A Programmable Dual-RNA – Guided 337:816–822.
7. Brown GD, Denning DW, Gow NAR, Levitz SM, Netea MG, White TC. 2012. Hidden Killers: Human Fungal Infections. Sci Transl Med 4:1–9.
8. Bennett RJ, Johnson AD. 2005. Mating in Candida albicans and the search for a sexual cycle. Annu Rev Microbiol 59:233–55.
9. Vyas V, Barrasa M, Fink G. 2015. A Candida albicans CRISPR system permits genetic engineering of essential genes and gene families. Sci Adv.
10. Min K, Ichikawa Y, Woolford CA, Mitchell AP. 2016. Candida albicans Gene Deletion with a Transient CRISPR-Cas9 System. mSphere 1:1–9.
11. Sugawara N, Haber JE. 1992. Characterization of Double-Strand Break-Induced Recombination : Homology Requirements and Single-Stranded DNA Formation 12:563–575.
12. Banerjee M, Thompson DS, Lazzell A, Carlisle PL, Pierce C, Monteagudo C, Lo L, Kadosh D. 2008. UME6 , a Novel Filament-specific Regulator of Candida albicans Hyphal Extension and Virulence 19:1354–1365.
13. Nobile CJ, Fox EP, Nett JE, Sorrells TR, Mitrovich QM, Hernday AD, Tuch BB, Andes DR, Johnson AD. 2012. A Recently Evolved Transcriptional Network Controls Biofilm Development in Candida albicans. Cell 148:126–138.
14. Nobile CJ, Andes DR, Nett JE, Smith FJ, Yue F, Phan QT, Edwards JE, Filler SG, Mitchell AP. 2006. Critical role of Bcr1-dependent adhesins in C. albicans biofilm formation in vitro and in vivo. PLoS Pathog 2:636–649.
15. Noble SM, Johnson AD. 2005. Strains and Strategies for Large-Scale Gene Deletion Studies of the Diploid Human Fungal Pathogen Candida albicans Strains and Strategies for Large-Scale Gene Deletion Studies of the Diploid Human Fungal Pathogen Candida albicans. Eukaryot Cell 4:298–309.
16. de Hoogt R, Luyten WH, Contreras R, De Backer MD. 2000. PCR- and ligation-mediated synthesis of split-marker cassettes with long flanking homology regions for gene disruption in Candida albicans. Biotechniques 28:1112–6.
17. Sikorski RS, Hieter P. 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. Genetics

122:19–27.
18. Botstein D, Falco SC, Stewart SE, Brennan M, Scherer S, Stinchcomb DT, Struhl K, Davis RW. 1979. Sterile host yeasts (SHY): A eukaryotic system of biological containment for recombinant DNA experiments. Gene 8:17–24.
19. Papon N, Courdavault V, Clastre M, Simkin AJ, Crèche J, Giglioli-Guivarc'h N. 2012. Deus ex Candida genetics: Overcoming the hurdles for the development of a molecular toolbox in the CTG clade. Microbiology 158:585–600.
20. Bae S, Kweon J, Kim HS, Kim J. 2014. correspondEnce Microhomology-based choice of Cas9 nuclease target sites. Nat Publ Gr 11:705–706.
21. Symington LS, Rothstein R, Lisby M. 2014. Mechanisms and regulation of mitotic recombination in saccharomyces cerevisiae. Genetics 198:795–835.
22. Fuller KK, Chen S, Loros JJ, Dunlap C. 2015. Development of the CRISPR / Cas9 System for Targeted Gene Disruption in Aspergillus fumigatus. Eukaryot Cell 14:1073–1080.
23. Wang T, Wei JJ, Sabatini DM, Lander ES. 2014. Genetic screens in human cells using the CRISPR-Cas9 system. Science 343:80–4.
24. Sander JD, Joung JK. 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol 32:347–55.
25. Walther A, Wendland J. 2003. An improved transformation protocol for the human fungal pathogen Candida albicans. Curr Genet 42:339–343.
26. Jones EW, Berget PB, Burnette III JM, Anderson C, Asafu-Adjei D, Avetisian S, Barrie F, Chen R, Chu B, Conroy S, Conroy S, Dill A, Eimer W, Garrity D, Greenwood A, Hamilton T, Hucko S, Jackson C, Livesey K, Monaco T, Onorato C, Otsuka M, Pai S, Schaeffer G, Shung S, Spath S, Stahlman J, Sweeney B, Wiltrout E, Yurovsky D, Zonneveld A. 2008. The spectrum of Trp− mutants isolated as 5-fluoroanthranilate-resistant clones in Saccharomyces bayanus, S. mikatae and S. paradoxus. Yeast 25:41–46.

FIGURE LEGENDS, CHAPTER 2

Figure 1. Break-induced marker excision concept. A selection marker is flanked by directly repeated sequences (blue arrows) in the genomic region designated "A B C D E F" (top line). A double strand break within the marker (second line) results in a recombination event between the repeats (third line) that, when resolved, deletes the marker (fourth line). This type of excision reaction was shown to occur in vivo in *S. cerevisiae* by Sugawara and Haber (11). Mechanistically, the recombination event may occur through the single strand annealing pathway, the microhomology-mediated end joining pathway, or other homology-promoted repair events (21). A textbook-style crossover is depicted for simplicity of visualization.

Figure 2. Strategy for marker recycling through CRISPR-Cas9-induced marker excision. Consider a situation in which an investigator seeks to make a *C. albicans* strain with homozygous deletion mutations in three genes - *YFG1, YFG2,* and *YFG3* - using only two selection markers, *M1* and *M2.* The marker cassettes, modified to include flanking direct repeats, are designated *rM1r* and *rM2r*. The three *YFG* genes are shown on separate blue, red, and violet chromosomes. The construction is carried out with only three transformations. Transformation 1: First, a homozygous *yfg1*Δ*::rM1r* mutant is created through transformation of the strain with a *yfg1*Δ*::rM1r* PCR product, an sgRNA gene that targets *YFG1,* and a *CAS9* gene. The M1+ phenotype is selected.

Transformation 2: After genotyping, a homozygous *yfg1Δ::rM1r* mutant is chosen and transformed to create a homozygous *yfg2Δ::rM2r* mutation. The transformation mix includes a *yfg2Δ::rM2r* PCR product, an sgRNA gene that targets *YFG2,* and a *CAS9* gene. In addition, in order to eliminate the *M1* marker by recombination between flanking repeats, an sgRNA gene that targets *M1* itself is also included. For this transformation, the M2+ phenotype is selected. Among M2+ transformants, some are M1-. The M1- transformants are genotyped to identify homozygous *yfg2Δ::rM2r* mutants. In addition, PCR genotyping is used to verify that *yfg1Δ::r* is homozygous, marked only with a repeat sequence and not with the entire *M1* marker. Transformation 3: A strain homozygous for *yfg1Δ::r yfg2Δ::rM2r* is chosen, and the strain is transformed to create a homozygous *yfg3Δ::rM1r* mutation. The transformation mix includes a *yfg3Δ::rM1r* PCR product, an sgRNA gene that targets *YFG3,* and a *CAS9* gene. In addition, in order to eliminate the *M2* marker by recombination between flanking repeats, an sgRNA gene that targets *M2* itself is also included. For this transformation, the M1+ phenotype is selected once again, just as it was in the initial transformation. Among M1+ transformants, some are M2-. The M2- transformants are genotyped to identify homozygous *yfg3Δ::rM1r* mutants. In addition, PCR genotyping is used to verify that *yfg2Δ::r* is homozygous, marked only with a repeat sequence and not with the entire *M2* marker.

Figure 3. PCR genotype analysis. A. Primer pairs for detection of deletion alleles. The designation "*YFG*" refers to any of the genes *UME6, BRG1,* or

BCR1. The designation "*yfgΔ::rMr*" refers to any of the deletion alleles *ume6Δ::r1HIS1r1*, *brg1Δ::r2LEU2r2,* or *bcr1Δ::r1HIS1r1.* Primer 1 anneals to a region flanking the *YFG* gene; primer 2 anneals to a region internal to *yfgΔ::rMr* and absent from *YFG*; primer 3 anneals to a region internal to *YFG* and absent from *yfgΔ::rMr*.

Figure 3B. Primer 1: UME6 Check/F. Primer 2: HIS1 Check Int/R. Primer 3: UME6 Check Int/R. Genotype assays for 10 His+ transformants from transformation 3 with primers for *UME6* alleles. Transformants 1, 2, 9, and 10 yield PCR products with primers 1 and 2, but not with 1 and 3, as expected for homozygous *ume6Δ::r1HIS1r1* mutants. Transformants 4-8 yield PCR products with primers 1 and 2 and with 1 and 3, as expected for heterozygous *UME6/ume6Δ::r1HIS1r1* mutants.

Figure 3C. Primer 1: BRG1 Check/F. Primer 2: LEU2 Check Int/R. Primer 3: BRG1 Check Int/R. Genotype assays for 14 Leu+ His- transformants from transformation 9 with primers for *BRG1* alleles. Transformants 1-9, 11,12, and 14 yield PCR products with primers 1 and 2, but not with 1 and 3, as expected for homozygous *brg1Δ::r2LEU2r2* mutants. Transformant 10 yields PCR products expected for a heterozygous mutant. Transformant 13 yields a PCR product indicative of a genetic rearrangement. The parent strain (lane P) is included as a control.

Figure 3D. Primer 1: BCR1 Check/F. Primer 2: HIS1 Check Int/R. Primer 3: BCR1 Check Int/R. Genotype assays for 12 His+ Leu- transformants from transformation 13 with primers for *BCR1* alleles. Transformants 3-6 and 8-10

yield PCR products as expected for homozygous *bcr1Δ::r1HIS1r1* mutants. Transformants 1, 2, 7, 11, and 12 yield PCR products expected for heterozygous mutants. The parent strain (lane P) is included as a control.

Figure 3E. Primer pairs for detection of marker loss. The designation "*yfgΔ::r*" refers to any of the deletion alleles that have lost the selection marker by recombination between repeats, including *ume6Δ::r1* or *brg1Δ::r2.* The designation "*yfgΔ::rmr*" refers to any of the deletion alleles that have lost a functional marker by a mutation at or near the Cas9-sgRNA cleavage site (represented by the red line segment), which would be designated *ume6Δ::r1HIS1r1*, or *brg1Δ::r2leu2r2.* Primer 1 anneals to a region flanking the *YFG* gene; primer 2 anneals to a region internal to *yfgΔ::rMr*; primer 4 anneals to a region flanking the *YFG* gene on the opposite side from primer 1.

Figure 3F. Primer 1: UME6 Check/F. Primer 2: HIS1 Check Int/R. Primer 4: UME6 Check Down/R. Genotype assays for 14 Leu+ His- transformants from transformation 9 with primers for *UME6* alleles. These transformants were the same ones analyzed in panel **C**. Transformants 1-13 yield PCR products with primers 1 and 4, but not with 1 and 2, as expected for homozygous *ume6Δ::r1* mutants. Transformant 14 yields PCR products expected for a homozygous *ume6Δ::r1HIS1r1* mutant. The parent strain (lane P) is included as a control. The single asterisks mark a minor PCR product that is expected from repeat annealing in the *r1HIS1r1* and *r1HIS1r1* cassettes (16).

Figure 3G. Primer 1: BRG1 Check/F. Primer 2: LEU2 Check Int/R. Primer 4: BRG1 Check Down/R. Genotype assays for 12 His+ Leu- transformants from

transformation 13 with primers for *BRG1* alleles. These transformants were the same ones analyzed in panel **D**. Transformants 1 and 3-12 yield PCR products expected for homozygous *brg1Δ::r2* mutants. Transformant 2 yields PCR products expected for a heterozygous *brg1Δ::r2/brg1Δ::r2LEU2r2* mutant. The parent strain (lane P) is included as a control. The single asterisk marks a minor PCR product that is expected from repeat annealing in the *r1HIS1r1* and *r1HIS1r1* cassettes (16). The double asterisk marks the PCR product expected for the *brg1Δ::r2LEU2r2* allele, which is diminished in yield due to presence of the smaller PCR product from the *brg1Δ::r2* allele.

Figure 4. Cloning of Partner Plasmids and Amplification of the *ume6Δ::r1HIS1r1* cassette. A. pMH01 and pMH02 are derived from pRS424, which contains KpnI and SapI restriction sites. The vector sequence between these restriction sites becomes the repeat sequence for the *ume6Δ::r1HIS1r1* cassette. This repeat sequence can be lengthened or shortened through use of different restriction enzymes. **B.** Amplification from pMH01 using primers His1 CRIME/F and UME6-SapI/R, the latter primer containing an 80 bp region of homology to the downstream region of UME6, generates one of the two halves of the *ume6Δ::r1HIS1r1* cassette. Amplification from pMH02 using primers His1 CRIME/R and UME6-KpnI/F, the latter primer containing an 80 bp region of homology to the upstream region of UME6, generates the other half of the *ume6Δ::r1HIS1r1* cassette. Following transformation, split-marker recombination

(16) reconstitutes the whole *ume6Δ::r1HIS1r1* cassette revealing the direct repeat.

Supplementary data. A list of strain genotypes, primer sequences, and assembled sequence files for partner plasmids pMH01, pMH02, pMH03, and pMH04.

FIGURES, CHAPTER 2

Figure 1

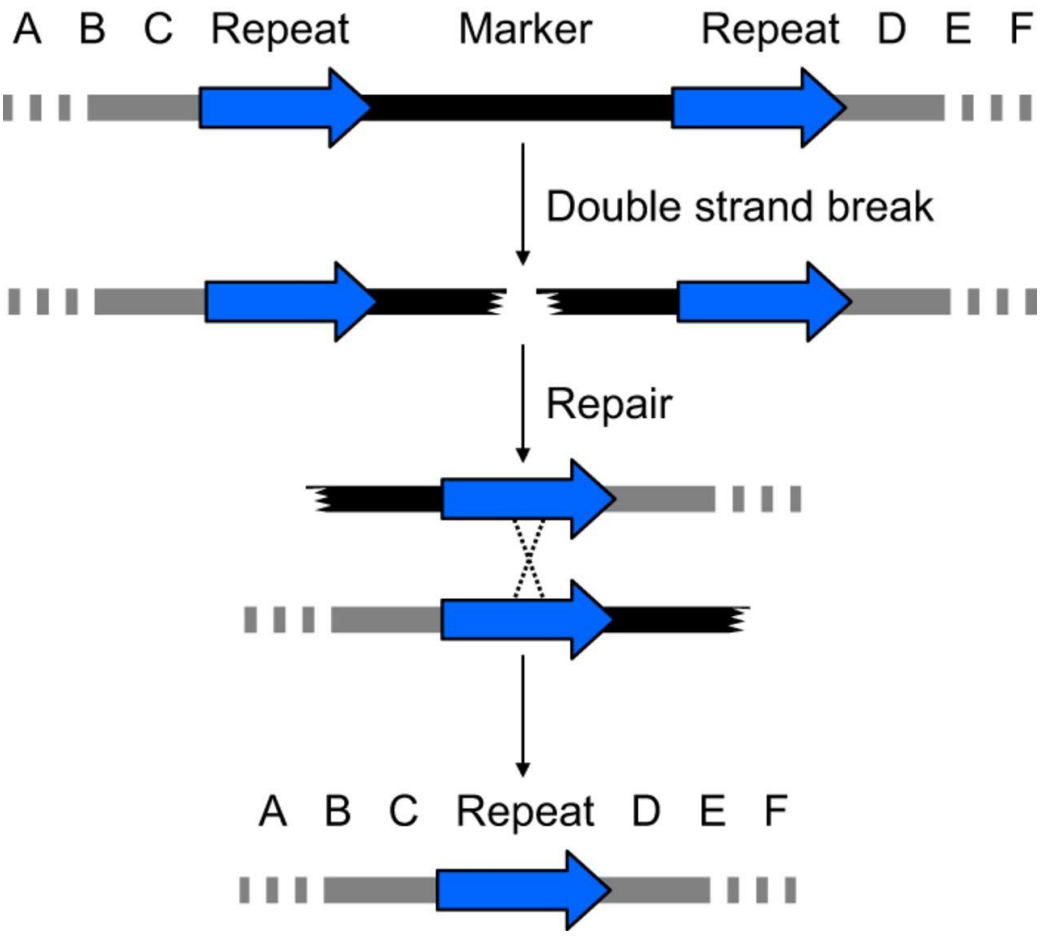


Figure 2

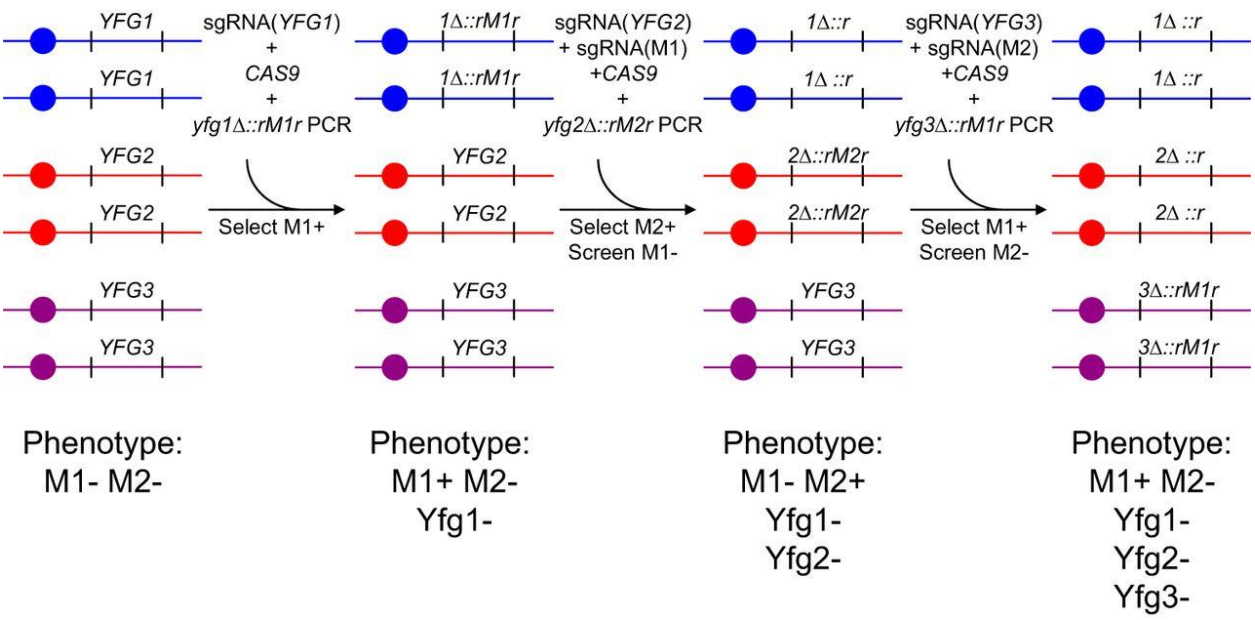


Figure 3

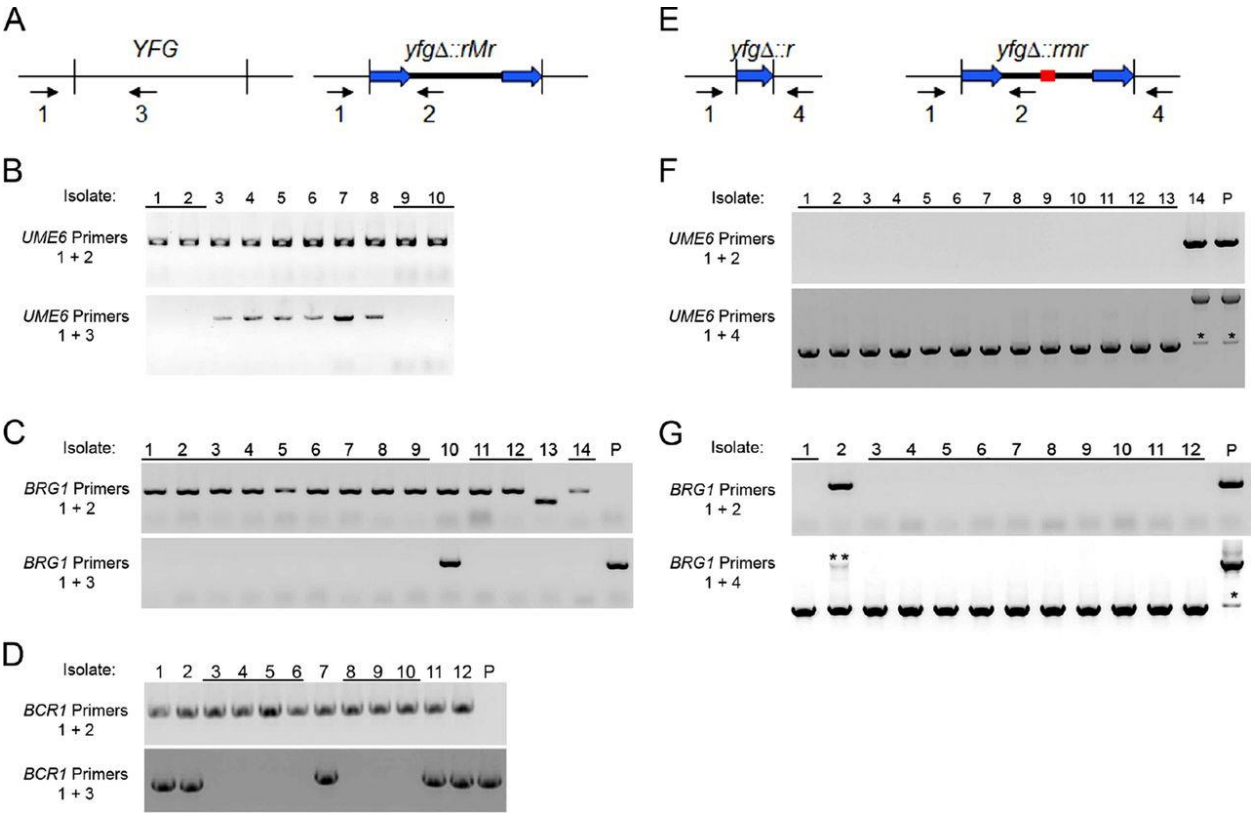
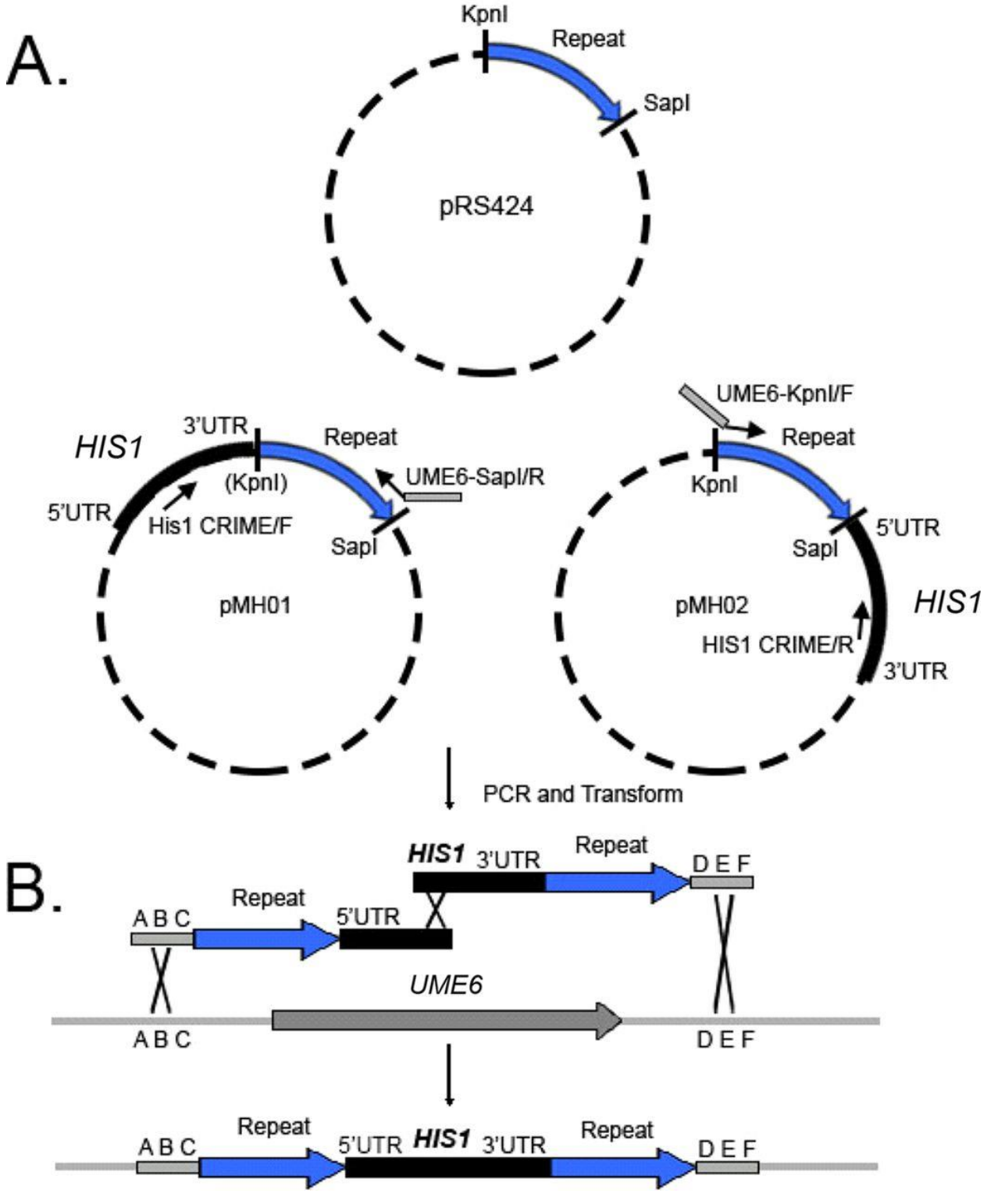


Figure 4


A.
KpnI
Repeat
SapI
pRS424
HIS1
3'UTR
Repeat
(KpnI)
UME6-SapI/R
5'UTR
His1 CRIME/F
SapI
pMH01
UME6-KpnI/F
Repeat
KpnI
5'UTR
SapI
pMH02
HIS1
HIS1 CRIME/R
3'UTR
PCR and Transform
B.
HIS1 3'UTR
Repeat
D E F
A B C
Repeat
5'UTR
UME6
A B C
D E F
Repeat
A B C
5'UTR HIS1 3'UTR
Repeat
D E F

TABLES, CHAPTER 2

Table 1. Transformation outcomes

Table 1. Transformation outcomes

Construction	Transformation[1]	Recipient strain[2]	sgRNA1	sgRNA2	Split-marker repair template	Total	His-	Leu-
1. *UME6* deletion	1	SN152	0	0	0	0		
	2	SN152	0	0	*ume6Δ::rHIS1r*	39		
	3	SN152	UME6 (1 ug)	0	*ume6Δ::rHIS1r*	333		
2. *BRG1* deletion	4	MH101	0	0	0	0		
and *HIS1* excision	5	MH101	0	0	*brg1Δ::rLEU1r*	12		
in *ume6Δ::rHIS1r*	6	MH101	BRG1 (1 ug)	0	*brg1Δ::rLEU1r*	1120	0	
	7	MH101	BRG1 (1 ug)	HIS1 (1 ug)	*brg1Δ::rLEU1r*	564	22	
	8	MH101	BRG1 (1 ug)	HIS1 (3 ug)	*brg1Δ::rLEU1r*	276	42	
	9	MH101	BRG1 (1 ug)	HIS1 (9 ug)	*brg1Δ::rLEU1r*	156	30	
3. *BCR1* deletion	10	MH110	0	0	0	0		
and *LEU2* excision	11	MH110	0	0	*bcr1Δ::rHIS1r*	26		
in *ume6Δ::r*	12	MH110	BCR1 (1 ug)	0	*bcr1Δ::rHIS1r*	134		0
brg1Δ::rLEU2r	13	MH110	BCR1 (1 ug)	LEU2 (1 ug)	*bcr1Δ::rHIS1r*	47		22

Footnotes:

1. All transformations included a *CAS9* gene, following the method of Min et al. (ref).

2. All strains are of genotype *his1Δ/his1 Δ leu2Δ/leu2Δ arg4Δ/arg4Δ.* MH101 has the additional genotype *ume6Δ::rHIS1r/ume6Δ::rHIS1r*. MH110 has the additional genotype *ume6Δ::r /ume6Δ::r brg1Δ::rLEU2r/brg1Δ::rLEU2r,* in which the *ume6Δ::r* allele is marked only with one copy of the repeat that had flanked our *rHIS1r* marker cassette.

Chapter 3: Rapid gene concatenation for genetic rescue of multi-gene mutants in *Candida albicans*

Manning Y. Huang, Carol A. Woolford and Aaron P. Mitchell

FOREWORD

This chapter is lightly modified from the publication version as appeared in mSphere, April 2018. All referenced figures are available at the end of the chapter. Dr. Carol Woolford carried out the Nanostring experiments described in this chapter, and I analyzed the data. I also carried out the rest of the work described in this chapter. Supplemental files referenced in this chapter are available online at:

https://doi.org/10.1128/mSphere.00169-18

INTRODUCTION

For many organisms, our understanding of gene function is based upon analysis of engineered mutations at predetermined genomic sites. The manipulations that yield engineered mutations can also yield adventitious mutations elsewhere in the genome. For this reason, when characterizing a defined mutant strain there is a concern that its phenotype may be due to adventitious mutations rather than the engineered mutation. For microbial pathogens, mutant strain characterization can be extensive, involving for example animal infection models. Moreover, strain phenotypes may appear to be complex because many mechanistic underpinnings of virulence have yet to be discovered. For these organisms, the possible contributions of adventitious mutations to phenotype are often assessed through analysis of a complemented or reconstituted strain, in which a wild-type copy of the gene that had been altered by engineering is introduced at a distant genomic site (complementation) or at the native locus (reconstitution). The importance of inclusion of complemented or reconstituted strains in pathogen phenotypic characterization was emphasized as the third postulate among the cornerstone "molecular Koch's postulates" coined by Falkow (13).

Once there is a panel of well characterized genes in an organism, it is useful to build upon that knowledge in the discovery or characterization of new genes. The approach of genetic interaction analysis, also called epistasis analysis, is a widely used strategy that can connect the functions of two genes. In premier genetic organisms like

Saccharomyces cerevisiae, systematic gene interaction analysis has been applied on a large scale to reveal new gene functions as well as novel connections among seemingly distinct biological processes (3). Our organism of interest, *Candida albicans*, is not as amenable to genetic manipulation as *S. cerevisiae*, but nonetheless has a growing panel of characterized genes that provide the foundation for genetic interaction analysis. In fact, the use of double and multi-gene mutants in *C. albicans* to draw functional inferences has a long history [e.g., (4-6, 11, 20-22, 30, 33)]. Studies of *C. albicans* double and multi-gene mutants have provided key insights into functional redundancy, pathway relationships, and major effectors that mediate pathway outputs. The analysis of double and multi-gene mutants has been of exceptional value for the understanding of *C. albicans*, as it has for diverse other organisms.

It is likely that analysis of *C. albicans* multi-gene mutants will have an increasingly prominent role in elucidation of the basis for key traits such as pathogenicity and drug resistance. This expectation is based upon the value of the information from these strains as well as the increasing ease with which multi-gene mutants can be created. Strain construction in *C. albicans* has been accelerated through implementation of CRISPR-Cas9-based methods (31) and, recently, a gene-drive system designed specifically for creation of double mutant strains (30). Unfortunately, the development of multi-gene complementation or reconstitution strategies, which we refer to together as genetic rescue strategies, has not kept pace with these developments. Strain validation through genetic rescue generally requires vector-based cloning, an often tedious procedure when multiple genes must be introduced together.

Here, we present a genetic rescue strategy that relies on recombinational assembly in *C. albicans* of PCR products. Integration of the concatenated assembly into a mutant strain is augmented by CRISPR-Cas9 cleavage at the targeted locus. Then the mutant strain that carries the concatenated assembly is tested to see if a wild-type phenotype is restored. We report proof-of-principle studies for phenotypic rescue of auxotrophic markers and for mutants defective for biofilm formation and filamentation. This approach, which we call Concatemer Assembly for Rescue of Mutant Abilities (CARMA), has the potential to accelerate multi-gene mutant validation in *C. albicans* and other organisms.

RESULTS

Rationale

We developed a multi-gene rescue strategy that employs three DNA segments (Figure 1). In this illustration, we seek to introduce functional *YFG1* and *YFG2* genes to assay phenotypic rescue of a *yfg1*Δ/Δ *yfg2*Δ/Δ double mutant. We include the marker *M1* in order to select for transformation and integration of a concatemer of the three segments. Each segment is synthesized through a PCR templated by the genome or an existing plasmid, thus eliminating any need for custom cloning. The segments have homology at their ends to one another or to the genome (regions A-F in Figure 1), so that cellular homologous recombination machinery can assemble the segments in a predictable order. We refer to an organization of homology that directs concatenation of the DNA segments as "concatenating homology."

The first (leftmost) segment includes one of the coding regions, *YFG1*, that corresponds to one of the deletions in the *yfg1*Δ/Δ *yfg2*Δ/Δ double mutant. It also includes *YFG1* 5' and 3' flanking regions. The overall assembly will be directed to integrate at the *YFG1* genomic locus to replace a *yfg1*Δ coding region deletion allele. Therefore, the *YFG1* 5' flanking region does not need to include all sequences necessary for proper expression; integration by recombination at the "A" region will join the *YFG1* DNA to the complete genomic 5' flanking region that extends to the left in the diagram. The segment needs to include sufficient 3' flanking sequences for *YFG1* expression, extending to the "B" region. This requirement arises because the 3' flanking end of the *YFG1* segment will

be fused to novel sequences from the *YFG2* locus, the "C" region, in the integrated assembly.

The second (middle) segment includes the second coding region, *YFG2*, that corresponds to a deletion in the *yfg1*Δ/Δ *yfg2*Δ/Δ double mutant. This segment includes *YFG2* 5' and 3' flanking regions that extend to regions labeled "C" and "D." These flanking regions must include all sequences necessary for proper expression because both regions are joined to novel neighboring sequences, the "B" and "E" regions, specified by the PCR primers.

The third (rightmost) segment includes a marker for selection, *M1,* along with all necessary 5' and 3' flanking sequences for expression. It has homology to the *YFG2* segment ("D-E" region) at one end and to the genome neighboring the *yfg1*Δ allele ("F" region) at the other end. The 5' to 3' orientation of each gene (*M1*, *YFG1*, and *YFG2*) is arbitrary; the diagram illustrates the actual relative orientations of the genes in the experiments we present below.

Homologous recombination should not allow integration of *M1* on its own. However, the *YFG1* and *YFG2* segments may serve as adaptors to enable homologous recombination events to insert a *YFG1:YFG2:M1* concatenated assembly in place of the *yfg1*Δ allele (Figure 1). Genomic integration of the concatemer may be stimulated by double strand cleavage of the *yfg1*Δ allele directed by the CRISPR-Cas9 system (Figure 1).

Genetic rescue of *arg4, leu2,* and *his1* mutations in strain SN152

We chose the popular laboratory strain SN152, which has *arg4*Δ/Δ*, leu2*Δ/Δ*,* and *his1*Δ/Δ mutations, for an initial test of the multi-gene rescue strategy. The strategy is expected to yield Arg+ Leu+ His+ transformants through recombinational concatemerization and integration of *ARG4*, *LEU2*, and *HIS1* cassettes. We designed three cassettes with concatenating homology as depicted in Figure 1, using the *C. albicans ARG4* gene in the position of *YFG1*, the *Candida albicans LEU2* gene in the position of *YFG2*, and the *Candida dubliniensis HIS1* gene in the position of *M1*. Integration of the cassettes was targeted by homology to the *arg4*Δ::*dpl200* mutant locus in the SN152 genome, and transformants were selected through acquisition of only a His+ phenotype. If the other two genes were assembled into an integrating concatemer, then His+ transformants should be Arg+ and Leu+ as well. Transformation of approximately 3 µg of each cassette yielded seven His+ transformants from three separate transformations (Table 1). One transformant could not be propagated on media selective for His+ strains and was not studied further. Five of the remaining six transformants grew when replica-plated to minimal SD medium and thus were Arg+ Leu+ His+. This observation indicates that the *ARG4, LEU2,* and *HIS1* genes were all maintained. Thus genetic rescue with three cassettes is possible, though the transformant recovery frequency is low (see (24, 32)).

The CRISPR-Cas9 system has been used to increase the rate of DNA integration through generation of a genomic double strand break that stimulates homology directed

repair. To test the effect of CRISPR-Cas9 on concatemer integration, we included in our transformation mixture DNA cassettes specifying Cas9 and a single guide RNA that targets the *arg4*Δ::*dpl200* locus. These components increased the recovery of His+ transformants by roughly 300 fold (Table 1). Approximately 94% of His+ transformants grew when replica-plated to SD media, thus indicating once again that the *ARG4, LEU2,* and *HIS1* genes were all maintained. This result indicates that multiple cassettes with concatenating homology may successfully integrate with selection for only a single cassette, and that integration is greatly enhanced by the introduction of a genomic double strand break at the integration site.

Genetic rescue of a *ume6*Δ/Δ *brg1*Δ/Δ mutant strain

We also tested our genetic rescue strategy with a *ume6*Δ/Δ *brg1*Δ/Δ double mutant strain. *UME6* and *BRG1* encode transcription factors that promote *C. albicans* filamentation and biofilm formation (2, 9, 12, 26). These two genes have large promoter and regulatory regions, and cloning of the two large regions for a conventional complementing construct could be laborious. We designed three PCR products with concatenating homology (Figure 1): a *UME6* segment (*YFG1*), a *BRG1* segment (*YFG2*), and a *C. dubliniensis HIS1* segment (*M1*). Integration of the concatenated cassettes was targeted to the *ume6*Δ::*r1* locus. The *ume6*Δ::*r1* locus has a deletion of the *UME6* coding region; it contains the *UME6* upstream and downstream regions flanking a 360 bp segment (designated *r1*) of vector pRS424 (17). The strain was transformed with approximately 3 µg of each PCR product along with cassettes

specifying Cas9 and a single guide RNA that targets the *ume6*Δ::*r1* allele. We recovered more than 200 His+ transformants from a single transformation. Ten transformants were selected for PCR genotyping. All ten transformants yielded the PCR product (primers 1+2, Figure 2A, B) expected from joining of the *UME6* coding region with the *UME6* 5' flanking region. All ten transformants also yielded the PCR product (primers 3+4, Figure 2A, C) expected from presence of the *UME6* coding region adjacent to the *BRG1* coding region. In addition, all ten transformants yielded the PCR product (primers 5+6, Figure 2A, D) expected from presence of the *BRG1* coding region adjacent to the *HIS1* coding region. Finally, all ten transformants yielded the PCR product (primers 7+8, Figure 2A, E) expected from presence of the *HIS1* coding region adjacent to the *UME6* 3' region. These results are consistent with the structure expected for integration of a concatenated *UME6:BRG1:HIS1* DNA segment at the *ume6*Δ::*r1* locus in the genome. Therefore, these genotyping results indicate that concatemer integration occurred in the majority of transformants.

We anticipated that we may recover homozygous integrants as a result of Cas9-sgRNA cleavage of both *ume6*Δ::*r1* alleles (31). PCR analysis indicated that nine of the ten genotyped transformants lacked any detectable *ume6*Δ::*r1* allele and thus were likely homozygous integrants (primers 9+10, Figure 2A, F). One genotyped transformant was apparently heterozygous (primers 9+10, Figure 2A, F, G). These results argue that the majority of our transformants were homozygous for the recombinational concatemer, integrated to replace the *ume6*Δ::*r1* locus in the genome.

For phenotypic rescue assays we focused on one transformant, strain MH281, that carried the integrated *UME6:BRG1:HIS1* concatenation product and was presumably homozygous due to lack of a *ume6*Δ::*r1* allele. We refer to this strain as the validation strain. Three assays were used to assess phenotypic rescue: filamentation, biofilm formation, and expression of *UME6*- and *BRG1*-responsive genes (2, 9, 12, 26, 33). We compared these phenotypes in the wild-type strain, a *ume6*Δ/Δ single mutant, a *brg1*Δ/Δ single mutant, a *ume6*Δ/Δ *brg1*Δ/Δ double mutant, and the *UME6:BRG1* validation strain. We sought to use strains with matched nutritional requirements; therefore, for example, the *ume6*Δ/Δ *brg1*Δ/Δ strain was a His+ derivative of the His- parental *ume6*Δ/Δ *brg1*Δ/Δ double mutant that had been transformed to create the validation strain. Similarly, the other strains had auxotrophies complemented.

Filamentation was assayed by microscopic observation after growth in YPD + serum at 37°C (Figure 3). The wild-type strain formed long hyphae with few detached yeast-form cells. The *ume6*Δ/Δ *brg1*Δ/Δ double mutant failed to form filaments under these conditions. The *ume6*Δ/Δ and *brg1*Δ/Δ mutants each formed short hyphae intermixed with a large proportion of yeast-form cells. The *UME6:BRG1* validation strain, like the wild type, formed long hyphae and few yeast-form cells. These observations indicate that the inability of the double mutant strain to undergo filamentation is due to the *ume6*Δ/Δ *brg1*Δ/Δ genotype, because (a) each single mutant has a filamentation defect, and (b) the *UME6:BRG1* concatemer restores filamentation when introduced into the *ume6*Δ/Δ *brg1*Δ/Δ background.

Biofilm formation was assayed by confocal microscopy after growth on silicone squares in YPD alone at 37°C for 24 hrs (Figure 4). The wild-type strain formed a robust biofilm of 300 μm in depth with abundant hyphae evident in both side-view and apical-view projections. The *ume6*Δ/Δ *brg1*Δ/Δ mutant was severely defective in biofilm formation: few cells adhered to the silicone surface, and no hyphae were apparent (Figure 4A, B). The *ume6*Δ/Δ and *brg1*Δ/Δ single mutants were also defective, yielding biofilms composed exclusively of yeast-form cells and of depth ~50-70 μm, as reported previously for 48 hr biofilms (33). The *UME6:BRG1* validation strain produced a biofilm of depth intermediate between the wild-type and the single mutant strains. Abundant hyphae were evident in side-view and apical-view projections (Figure 4A, B). These observations indicate that the inability of the double mutant strain to produce a biofilm is due to the *ume6*Δ/Δ *brg1*Δ/Δ genotype, because (a) each single mutant has a biofilm defect, and (b) the *UME6:BRG1* concatemer restores biofilm when introduced into the *ume6*Δ/Δ *brg1*Δ/Δ background. However, the *UME6:BRG1* concatemer enables only partial rescue of the biofilm defect, because the validation strain does not produce as extensive a biofilm as the wild-type strain.

Gene expression was measured by Nanostring on RNA samples from cells grown for 4 hours in YPD at 37°C. The probe set assayed RNA levels for a panel of 182 genes, including 166 environmentally responsive genes (33). The wild-type strain had significantly different RNA levels (>2-fold change, $p<0.05$) than the *ume6*Δ/Δ *brg1*Δ/Δ double mutant for 18 genes in addition to the deleted genes *UME6* and *BRG1* (Figure 5A). Almost all of these RNA levels were also significantly different in the wild-type

strain compared to the *ume6*Δ/Δ and *brg1*Δ/Δ single mutant strains. The affected genes included the core hyphal-associated genes *ALS3, ECE1, HWP1, IHD1,* and *RBT1* (Figure 5B), as expected (2, 9, 12, 26). The validation strain expressed the affected RNAs at levels similar to the wild-type strain (Figure 5A). Most of the core hyphal-associated genes had RNA levels of roughly 50% of the level in the wild type. These observations confirm that the concatenated *UME6:BRG1* construct partially rescues the biofilm defect of the *ume6*Δ/Δ *brg1*Δ/Δ double mutant.

Partial rescue by *UME6:BRG1* may result from incompletely restored expression of the *UME6* or *BRG1* genes. We were concerned in particular about *BRG1* expression for two reasons. First, its genomic 5' flanking region extends for over 10 Kbp to the next open reading frame, but only 1.6 Kbp had been included in our concatenation cassette. Second, Brg1 is a positive regulator of *UME6* expression, so a reduction in *BRG1* expression may cause a reduction in *UME6* expression as well (26). Our Nanostring measurements indicated that the validation strain expressed *BRG1* at ~30% of the wild-type level, and expressed *UME6* at ~60% of the wild-type level (Figure 5C). These results are consistent with the model that the *UME6:BRG1* concatenated construct allows partial phenotypic rescue because expression of the *BRG1* and *UME6* genes is only partially restored.

DISCUSSION

Genetic interaction analysis is a vital tool to help decipher functional relationships in pathways, networks, and protein complexes (3). Central to this analysis is the ability to construct and validate multi-gene mutants. In this report we have described a simple approach to introduce multiple DNA segments as an integrated genomic concatemer to generate a validation strain. The validation strain is then used in assays for rescue of multi-gene mutant phenotypes. We refer to the approach as Concatemer Assembly for Rescue of Mutant Abilities (CARMA) in order to convey features of both the technical procedure and the experimental objective.

The CARMA approach builds upon well-founded principles. Homologous recombination can seal a break in a target DNA, as demonstrated by Ma et al. with plasmid DNA in *S. cerevisiae* (23). Multiple DNA fragments with overlapping homology can be assembled through in vivo recombination in *S. cerevisiae*, as used for example in high-throughput construction of gene deletion cassettes (10) and in construction of a synthetic bacterial genome (15). A genomic CRISPR cut in a *C. albicans* mutant allele can enable integration of wild-type allele, as shown by Nguyen et al. (25). We view CARMA as an amalgam of the principles demonstrated in these prior studies.

CARMA requires some attention to the extent of functional regulatory regions. Integration of the multi-gene concatemer at one of the mutant loci reconstitutes that gene with its native genomic 5' flanking region. We chose to integrate specifically at the

*ume6*Δ::*r1* locus because *UME6* is known to have a large 5' regulatory region that includes 7 Kbp or more upstream of the open reading frame (7, 8). Long 5' regulatory regions are also found among other genes connected to filamentation, including *NRG1*, *HWP1*, and *ALS3* (1, 7, 18). The *BRG1* 5' regulatory region has not been subjected to detailed analysis, and we included 1.6 Kbp of 5' flanking sequence in our genetic rescue segment. For many genes, we have found that ~1.6 Kbp is sufficient for full rescue. However, we note that the *BRG1* 5' region extends for over 10 Kbp before the next open reading frame, so it seems possible that it is yet another filamentation-related gene with a long 5' regulatory region. We infer that 1.6 Kbp of 5' flanking sequence is not sufficient for full *BRG1* expression, because *BRG1* RNA levels were lower in the validation strain than in the wild-type strain. The reduced *BRG1* expression was nonetheless sufficient for substantial rescue of the double mutant phenotypes. Perhaps the best quantitative measure of function was the expression levels of *UME6*- and *BRG1*-dependent target genes, which were significantly greater than observed in either single mutant or the double mutant.

Note that, in all experiments reported here, the target integration loci are deletion alleles with fairly small relics of the initial marker; they are *yfg1*Δ*::dpl200* or *yfg1*Δ*::r1* alleles. Our preliminary studies suggest that a CARMA-like approach is less efficient with insertion-deletion alleles of structure *yfg1*Δ*::ARG4*. In addition, it is noteworthy that many mutant strains currently in use have different markers at the two alleles (of genotype *yfg1*Δ*::ARG4/ yfg1*Δ*::URA3*, for example). The recipient strains in the present study were all homozygous for identical mutant alleles at the target integration locus.

Therefore, for existing *yfg1Δ::ARG4/yfg1Δ::URA3* strains, it may be necessary to use two sgRNA genes in order to direct integration to both alleles.

A few features of CARMA seem especially useful. First, all of the DNA molecules in the transformation reaction are PCR products, including the DNA segments destined to be assembled into the integrated concatemer, and the genes specifying Cas9 and the sgRNA, which are the original Vyas constructs (31) used with our transient CRISPR protocol (24). This feature saves some time because the investigator circumvents cloning and screening for appropriate plasmids. This feature also overcomes the potential problem that a DNA fragment may be unstable in the intermediate host used for cloning. A second useful feature is that homozygous integrants can be recovered. This feature can eliminate partial phenotypic rescue due to gene dosage effects. Third, while in CARMA each DNA segment corresponded to a complete coding region and flanking regions, concatenating homology of PCR products could also be used for construction of fusion genes, localized mutagenesis, mapping of functional promoter elements, or introduction of allelic variants.

METHODS

Strains, Media, and Transformations

C. albicans strains were all archived in 15% glycerol stocks stored at -80°C. Strains were grown on solid YPD media for 2 days at 30°C, and then in liquid YPD media overnight at 30°C with shaking prior to all assays and transformations. Transformations were performed using the lithium acetate transformation method supplemented with transient CRISPR-Cas9 system components (24). Transformants were selected on CSM medium lacking specified auxotrophic supplements. All strains are listed in Supplemental File S1.

Preparation of DNA Cassettes and genotyping

In preparation for the amplification of wild type alleles, genomic DNA was prepared from the SC5314 strain according to a modified version of the Hoffman and Winston procedure (16).

All PCR reactions were performed at 30 cycles according to the manufacturer's protocols using TaKaRa Ex Taq™ DNA Polymerase. All primer sequences are listed in Supplemental File S2.

The *C. albicans ARG4* gene was amplified from SC5314 genomic DNA using primers “CaARG4 MGC/F” and “CaARG4 3’->CaLEU2 5’/R”. The *C. albicans LEU2* gene was amplified from SC5314 genomic DNA using primers “CaLEU2 MGC/F” and “CaLEU2 3’-

>CdHIS1 for/R". The *C. dubliniensis HIS1* cassette was amplified from plasmid pSN52 (27) using primers "CdHIS1 for->CaLEU2 3'/F" and "CdHIS1 rev->CaARG4 3'/R" (27).

The "CaARG4 3'->CaLEU2 5'/R" primer contains 80 bp of homology to the upstream region of the *C. albicans LEU2* gene. The "CaLEU2 3'->CdHIS1 for/R" primer contains 80 bp of homology to the upstream region of the *C. dubliniensis HIS1* cassette. The "CdHIS1 for->CaLEU2 3'/F" primer contains 80 bp of homology to the downstream region of the *C. albicans LEU2* gene. The "CdHIS1 rev->CaARG4 3'/R" primer contains 80 bp of homology to the downstream region of the *ARG4* locus.

The *C. albicans UME6* gene was PCR amplified from SC5314 genomic DNA using primers "UME6 MGC/F" and "UME6 3'->BRG1 5'/R". The *C. albicans BRG1* gene was PCR amplified from SC5314 genomic DNA using primers "BRG1 MGC/F" and "BRG1 3'->CdHIS1 for/R". The *C. dubliniensis HIS1* cassette was amplified from plasmid pSN52 (27) using primers "CdHIS1 for->BRG1 3'/F" and "CdHIS1 rev->UME6 3'/R" (27).

The "UME6 3'->BRG1 5'/R" primer contains 80 bp of homology to the upstream region of the *C. albicans BRG1* gene. The "BRG1 3'->CdHIS1 for/R" primer contains 80 bp of homology to the upstream region of the *C. dubliniensis HIS1* cassette. The "CdHIS1 for->BRG1 3'/F" primer contains 80 bp of homology to the downstream region of the *C. albicans BRG1* gene. The "CdHIS1 rev->UME6 3'/R" primer contains 80 bp of homology to the downstream region of the *UME6* locus.

PCR Genotyping

Genotyping of the *UME6:BRG1* synthetic locus was performed using four different pairs of primers. Amplification using these primer pairs yielding a DNA segment of the correct size is indicative of the expected homologous integration event. Integration of *UME6* at the *UME6* locus was confirmed using primers “UME6 FarUP/F” and “UME6 Int/R”. Integration of *BRG1* downstream of *UME6* was confirmed using primers “UME6 2346 Int/F” and “BRG1 Int/R”. Integration of the *C. dubliniensis HIS1* cassette was confirmed using primers “BRG1 Int/F” and “HIS1 Check Int/R”, as well as “HIS1 CRIME/F” and “UME6 FarDOWN/R”.

Homozygosity was confirmed using primers “UME6 MGC/F” and “UME6 downstream/R”, which yields alternatively either a 1.1 Kbp product if a *ume6*Δ::*r1* allele is present, or a 3.3 Kbp product if a *UME6* allele is present.

Single-guide RNA Cassette Construction

Fusion PCR was used to assemble sgRNA cassettes according to previously described protocols (17, 24). The arg4::dpl200 sgRNA cassette was prepared using primers “sgRNA/F arg4::dpl200” and “SNR52/R arg4::dpl200”, and targets the sequence “TCAAGAATAATCccgctgct”. The ume6::r sgRNA cassette was prepared using the primers “sgRNA/F ume6::r” and “SNR52/R ume6::r”, and targets the sequence “TTAATCCACTGTATAACTCG”.

RNA preparation and NanoString analysis

Overnight cultures of the relevant strains were inoculated into 25 mL of YPD to a final OD_{600} of 0.2 in 250 mL Erlenmeyer flask. Cultures were then allowed to grow at 4 hours with vigorous shaking (200 rpm) at 37°C, then collected by vacuum filtration and stored at -80°C until ready for extraction. RNA was extracted from frozen cell cultures using the Qiagen RNeasy Mini Kit (cat#74104) with modifications as described previously (33).

Nanostring analysis of RNA samples was performed using previously described methods (33). The Nanostring codeset contained 166 gene targets associated with environmental response pathways and 15 gene targets of interest for a variety of reasons (33).

Biofilm and Filamentation Assay Imaging

Biofilms were grown according to previously described protocols with the following modifications (33). Overnight cultures grown at 30°C were inoculated into 2 ml YPD in wells containing a 1.5cm x 1.5cm silicone square (Bentec Medical Inc.) and allowed to adhere to the silicone square for 90 minutes in an incubator-shaker at 37°C with shaking at 60 rpm. Squares were then washed by brief immersion into a well containing 2 ml PBS, then placed into a fresh well containing 2 ml YPD. After 24 hours, biofilms were fixed using 2 mls of 4% formaldehyde and 2.5% glutaraldehyde in 1x PBS on an orbital shaker for 1 hour at 60 rpm. The biofilms were then stained for 24 hours with concanavalin A, Alexa Fluor 594 conjugate (Life Technologies) diluted to 25 µg/ml in PBS, before being imaged using a slit-scan confocal optical unit on a Zeiss Axiovert 200

microscope with an 40x 0.85 NA oil immersion objective. The imaging processes are described in detail elsewhere (19).

The filamentation assay was performed using strains inoculated into 5 ml YPD in glass test tubes and grown for 4 hours at 37°C in a roller drum at maximum speed. Following growth, the cells were fixed and stained with Calcafluor white. An aliquot of cells was then imaged using conventional fluorescence microscopy on a Zeiss Axio Observer Z.1 fluorescence microscope and a 20x 0.8 NA objective.

Software

Colony counts were collected using OpenCFU software (14). Microscopy images were compiled with ImageJ (29), Photoshop CS6, and Microsoft Powerpoint 2010. RNA counts were analyzed in Microsoft Excel 2010 and Graphpad Prism 7. Heat maps were generated using MeV software (28), then arranged in Photoshop.

Data Availability

Nanostring transcription profiling data may be found in Supplemental File S3.

Acknowledgements

We sincerely thank Dr. Frederick Lanni and Katherine Lagree for their invaluable advice and contributions, and in their support and assistance. We would also like to thank Tatyana Aleynikov and Megan Bean for their extensive technical support, as well as Surya D. Aggarwal and Dr. N. Luisa Hiller for assistance in statistical analysis.

These studies were supported by NIH/NIAID grant 1R21AI135178 (APM).

LITERATURE CITED, CHAPTER 3

1. Argimon, S., J. A. Wishart, R. Leng, S. Macaskill, A. Mavor, T. Alexandris, S. Nicholls, A. W. Knight, B. Enjalbert, R. Walmsley, F. C. Odds, N. A. Gow, and A. J. Brown. 2007. Developmental regulation of an adhesin gene during cellular morphogenesis in the fungal pathogen Candida albicans. Eukaryot Cell 6:682-692.
2. Banerjee, M., D. S. Thompson, A. Lazzell, P. L. Carlisle, C. Pierce, C. Monteagudo, J. L. Lopez-Ribot, and D. Kadosh. 2008. UME6, a novel filament-specific regulator of Candida albicans hyphal extension and virulence. Mol Biol Cell 19:1354-1365.
3. Baryshnikova, A., M. Costanzo, C. L. Myers, B. Andrews, and C. Boone. 2013. Genetic interaction networks: toward an understanding of heritability. Annu Rev Genomics Hum Genet 14:111-133.
4. Bharucha, N., Y. Chabrier-Rosello, T. Xu, C. Johnson, S. Sobczynski, Q. Song, C. J. Dobry, M. J. Eckwahl, C. P. Anderson, A. J. Benjamin, A. Kumar, and D. J. Krysan. 2011. A large-scale complex haploinsufficiency-based genetic interaction screen in Candida albicans: analysis of the RAM network during morphogenesis. PLoS Genet 7:e1002058.
5. Bockmuhl, D. P., S. Krishnamurthy, M. Gerads, A. Sonneborn, and J. F. Ernst. 2001. Distinct and redundant roles of the two protein kinase A isoforms Tpk1p and Tpk2p in morphogenesis and growth of Candida albicans. Mol Microbiol 42:1243-1257.
6. Braun, B. R., and A. D. Johnson. 2000. TUP1, CPH1 and EFG1 make independent contributions to filamentation in candida albicans. Genetics 155:57-67.
7. Childers, D. S., and D. Kadosh. 2015. Filament condition-specific response elements control the expression of NRG1 and UME6, key transcriptional regulators of morphology and virulence in Candida albicans. PLoS One 10:e0122775.
8. Childers, D. S., V. Mundodi, M. Banerjee, and D. Kadosh. 2014. A 5' UTR-mediated translational efficiency mechanism inhibits the Candida albicans morphological transition. Mol Microbiol 92:570-585.
9. Cleary, I. A., A. L. Lazzell, C. Monteagudo, D. P. Thomas, and S. P. Saville. 2012. BRG1 and NRG1 form a novel feedback circuit regulating Candida albicans hypha formation and virulence. Mol Microbiol 85:557-573.
10. Colot, H. V., G. Park, G. E. Turner, C. Ringelberg, C. M. Crew, L. Litvinkova, R. L. Weiss, K. A. Borkovich, and J. C. Dunlap. 2006. A high-throughput gene knockout procedure for Neurospora reveals functions for multiple transcription factors. Proc Natl Acad Sci U S A 103:10352-10357.
11. Davis, D., R. B. Wilson, and A. P. Mitchell. 2000. RIM101-dependent and-independent pathways govern pH responses in Candida albicans. Mol Cell Biol 20:971-978.
12. Du, H., G. Guan, J. Xie, Y. Sun, Y. Tong, L. Zhang, and G. Huang. 2012. Roles of Candida albicans Gat2, a GATA-type zinc finger transcription factor, in biofilm formation, filamentous growth and virulence. PLoS One 7:e29707.
13. Falkow, S. 1988. Molecular Koch's postulates applied to microbial pathogenicity. Rev Infect Dis 10 Suppl 2:S274-276.
14. Geissmann, Q. 2013. OpenCFU, a new free and open-source software to count cell colonies and other circular objects. PLoS One 8:e54072.

15. Gibson, D. G., G. A. Benders, C. Andrews-Pfannkoch, E. A. Denisova, H. Baden-Tillson, J. Zaveri, T. B. Stockwell, A. Brownley, D. W. Thomas, M. A. Algire, C. Merryman, L. Young, V. N. Noskov, J. I. Glass, J. C. Venter, C. A. Hutchison, 3rd, and H. O. Smith. 2008. Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. Science 319:1215-1220.
16. Hoffman, C. S., and F. Winston. 1987. A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of Escherichia coli. Gene 57:267-272.
17. Huang, M. Y., and A. P. Mitchell. 2017. Marker Recycling in Candida albicans through CRISPR-Cas9-Induced Marker Excision. mSphere 2.
18. Kim, S., Q. B. Nguyen, M. J. Wolyniak, G. Frechette, C. R. Lehman, B. K. Fox, and P. Sundstrom. 2018. Release of transcriptional repression through the HCR promoter region confers uniform expression of HWP1 on surfaces of Candida albicans germ tubes. PLoS One 13:e0192260.
19. Lagree, K., J. V. Desai, J. S. Finkel, and F. Lanni. 2018. Microscopy of fungal biofilms. Curr Opin Microbiol 43:100-107.
20. Lermann, U., and J. Morschhauser. 2008. Secreted aspartic proteases are not required for invasion of reconstituted human epithelia by Candida albicans. Microbiology 154:3281-3295.
21. Li, X., N. Robbins, T. R. O'Meara, and L. E. Cowen. 2017. Extensive functional redundancy in the regulation of Candida albicans drug resistance and morphogenesis by lysine deacetylases Hos2, Hda1, Rpd3 and Rpd31. Mol Microbiol 103:635-656.
22. Lo, H. J., J. R. Kohler, B. DiDomenico, D. Loebenberg, A. Cacciapuoti, and G. R. Fink. 1997. Nonfilamentous C. albicans mutants are avirulent. Cell 90:939-949.
23. Ma, H., S. Kunes, P. J. Schatz, and D. Botstein. 1987. Plasmid construction by homologous recombination in yeast. Gene 58:201-216.
24. Min, K., Y. Ichikawa, C. A. Woolford, and A. P. Mitchell. 2016. Candida albicans Gene Deletion with a Transient CRISPR-Cas9 System. mSphere 1:00130-00116.
25. Nguyen, N., M. M. F. Quail, and A. D. Hernday. 2017. An Efficient, Rapid, and Recyclable System for CRISPR-Mediated Genome Editing in Candida albicans. mSphere 2.
26. Nobile, C. J., E. P. Fox, J. E. Nett, T. R. Sorrells, Q. M. Mitrovich, A. D. Hernday, B. B. Tuch, D. R. Andes, and A. D. Johnson. 2012. A recently evolved transcriptional network controls biofilm development in Candida albicans. Cell 148:126-138.
27. Noble, S. M., and A. D. Johnson. 2005. Strains and strategies for large-scale gene deletion studies of the diploid human fungal pathogen Candida albicans. Eukaryot Cell 4:298-309.
28. Saeed, A. I., V. Sharov, J. White, J. Li, W. Liang, N. Bhagabati, J. Braisted, M. Klapa, T. Currier, M. Thiagarajan, A. Sturn, M. Snuffin, A. Rezantsev, D. Popov, A. Ryltsov, E. Kostukovich, I. Borisovsky, Z. Liu, A. Vinsavich, V. Trush, and J. Quackenbush. 2003. TM4: a free, open-source system for microarray data management and analysis. Biotechniques 34:374-378.
29. Schindelin, J., I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J. Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona. 2012. Fiji: an open-source platform for biological-image analysis. Nat Methods 9:676-682.

30. Shapiro, R. S., A. Chavez, C. B. M. Porter, M. Hamblin, C. S. Kaas, J. E. DiCarlo, G. Zeng, X. Xu, A. V. Revtovich, N. V. Kirienko, Y. Wang, G. M. Church, and J. J. Collins. 2018. A CRISPR-Cas9-based gene drive platform for genetic interaction analysis in Candida albicans. Nat Microbiol 3:73-82.
31. Vyas, V. K., M. I. Barrasa, and G. R. Fink. 2015. A Candida albicans CRISPR system permits genetic engineering of essential genes and gene families. Sci Adv 1:e1500248.
32. Wilson, R. B., D. Davis, and A. P. Mitchell. 1999. Rapid hypothesis testing with Candida albicans through gene disruption with short homology regions. J Bacteriol 181:1868-1874.
33. Woolford, C. A., K. Lagree, W. Xu, T. Aleynikov, H. Adhikari, H. Sanchez, P. J. Cullen, F. Lanni, D. R. Andes, and A. P. Mitchell. 2016. Bypass of Candida albicans Filamentation/Biofilm Regulators through Diminished Expression of Protein Kinase Cak1. PLoS Genet 12:e1006487.

FIGURE LEGENDS, CHAPTER 3

Figure 1. Strategy for Concatemer Assembly for Rescue of Mutant Abilities (CARMA). Given a double mutant strain with the genotype *yfg1*Δ/Δ *yfg2*Δ/Δ with an available selectable marker *M1*, both *YFG1* and *YFG2* may conceptually be reintroduced at either the *yfg1*Δ (depicted in Line 2 and Line 3) or *yfg2*Δ locus. Cassettes containing *YFG1* and *YFG2* are generated by PCR from wild-type genomic DNA (Line 1). The *YFG1* cassette with sequences shown in blue is amplified with the primers indicated at A and BC. It contains a segment of the *YFG1* promoter, the *YFG1* ORF, and terminator. Sequence C, containing homology to the *YFG2* cassette, is introduced into the *YFG1* cassette using the long primer indicated at BC. The *YFG2* cassette with sequences shown in gold is amplified with the primers indicated at C and DE. It contains a segment of the *YFG2* promoter, the *YFG2* ORF, and terminator. Sequence E, containing homology to the *M1* marker cassette, is introduced into the *YFG2* cassette using the long primer indicated at DE. The *M1* marker cassette with sequences shown in green is amplified from a plasmid containing the *M1* marker with the primers indicated at DE and F. Sequence D, containing homology to the terminator region of *YFG2* is introduced into the *M1* cassette using the long primer indicated at DE. Sequence F, containing homology to a region downstream of *YFG1* is introduced into the *M1* cassette using the long primer indicated at F. All three cassettes are transformed into the *yfg1*Δ/Δ *yfg2*Δ/Δ *m1*Δ/Δ strain under M1 selection, along with DNA cassettes expressing *CAS9* and a single guide RNA targeting the *yfg1*Δ locus. The double strand break introduced by Cas9 complexed with the *yfg1*Δ targeting sgRNA is

indicated by two black triangles. Expected homologous recombination events are depicted as single crosses, and together should yield a concatenated locus containing *YFG1*, *YFG2*, and *M1* (Line 4). Boxes C, D, E, and F each indicate segments of 80 bp. Where possible, homology between PCR cassettes is maximized. Sequences D and E together provide 160 base pairs of homology between the *YFG2* and *M1* cassettes. (In pilot experiments, we could not generate integrated concatemers with only 80 base pairs of homology between the *YFG2* and *M1* cassettes.) Sequences C and F provide 80 base pairs of homology to the *YFG1* cassette and *yfg1*Δ locus respectively. For construction of the *ARG4:LEU2:HIS1* allele, the A-*YFG1* interval provided 419 base pairs of homology. For construction of the *UME6:BRG1:HIS1* allele, the A-*YFG1* interval provided 403 base pairs of homology. In our experiments, we avoided extending the homology between the *YFG1* and *YFG2* cassettes to 160 bp in length by adding 80 bp of homology at Sequence B. We were concerned that presence of Sequence B in the *YFG2* cassette could allow integration of only *YFG2* and *M1* downstream of the *yfg1*Δ deletion scar in the B-F interval.

Figure 2. Genotyping *UME6:BRG1:HIS1* candidates. A. Primer pairs for detection of *UME6*, *BRG1*, and *HIS1*. The *UME6:BRG1* locus is depicted in the first line with the expected location and integration of *UME6*, *BRG1*, and *HIS1* depicted. Primers 1 + 2 are used for detection of *UME6* at the *ume6*Δ::*r1* locus. Primer 1 anneals upstream of any predicted recombinational junction. Primers 3 + 4 are used to detect *BRG1* integration downstream of *UME6*. Primers 5 + 6 are used to detect *HIS1* integration

downstream of *BRG1*. Primers 7 + 8 are used to detect *HIS1* at the *ume6*Δ::*r1* locus. Primer 8 sits downstream of any predicted recombinational junction.

The *ume6*Δ::*r1* locus is depicted on the second line, with the genetic scar consisting of a single repeat from the *r1HIS1r1* cassette indicated by a small white box. Primers 9 + 10 bind upstream and downstream of the *ume6*Δ::*r1* locus and can be used for detection of the *ume6*Δ::*r1* allele. **B.** PCR products for detection of *UME6* at its native locus using primer pair 1 + 2. Amplification of a 1.8 Kbp band is consistent with the presence of *UME6* at its native locus. **C.** PCR products for detection of the *UME6:BRG1* junction using primer pair 3 + 4. Amplification of a 2.7 Kbp band is consistent with the presence of *BRG1* downstream of *UME6.* **D.** PCR products for detection of the *BRG1:HIS1* junction using primer pair 5 + 6. Amplification of a 1.3 Kbp band is consistent with the presence of *HIS1* downstream of *BRG1*. **E.** PCR products for detection of integration of *HIS1* at the *ume6*Δ::*r1* locus using primer pair 7 + 8. Amplification of a 2.4 Kbp band is consistent with the presence of *HIS1* at the *ume6*Δ::*r1* locus. **F.** PCR products for detection of homozygous integration using primer pair 9 + 10. Amplification of a 1.1 Kbp band suggests the presence of a *ume6*Δ::*r1* allele (see section A, lower line). Amplification of a 3.3 Kbp band is consistent with the presence of either a WT *UME6* or concatenated *UME6:BRG1* allele (see section A, upper line). A 1.1 Kbp is observed in the parental strain and in strain 2, indicating the presence of at least one copy of the *ume6*Δ::*r1* allele. The absence of a 1.1 Kbp band in strains 1, 3-10 suggests homozygous integration of the concatenated allele. **G.** Expanded insert of panel F for detection of homozygous integration. Lane 2 contains a

3.3 Kbp band which is absent in the parental strain. The 3.3 Kbp band in lane 2 is consistent with integration of at least one *UME6* allele, which suggests this strain underwent heterozygous integration of the *UME6:BRG1* concatemer.

Figure 3. Filamentation phenotypes of mutant and validation strains. Fluorescence images of cells with Calcafluor White staining after 4 hours of growth in YPD + serum. The SN250 WT strain, *ume6*Δ/Δ mutant, *brg1*Δ/Δ mutant *ume6*Δ/Δ *brg1*Δ/Δ double mutant, and *UME6:BRG1* validation strains were assayed for phenotypic rescue of filamentation under planktonic growth conditions. The wild type strain forms long filamentous hyphae, whereas the *ume6*Δ/Δ mutant, *brg1*Δ/Δ mutant, and *ume6*Δ/Δ *brg1*Δ/Δ double mutant grow primarily as yeast-form cells. The *UME6:BRG1* validation strain forms long filamentous hyphae comparable to WT cells.

Figure 4. Side view and apical projections of mutant and validation strain biofilms. The SN250 WT strain, *ume6*Δ/Δ mutant, *brg1*Δ/Δ mutant *ume6*Δ/Δ *brg1*Δ/Δ double mutant, and *UME6:BRG1* validation strains were assayed for phenotypic rescue of biofilm formation. **A.** Lateral projections of *C. albicans* WT, mutant, and validation strain biofilms. A wild type strain forms a large robust biofilm, while the *ume6*Δ/Δ, *brg1*Δ/Δ, and *ume6*Δ/Δ *brg1*Δ/Δ form defective biofilms. The *UME6:BRG1* validation strain forms a biofilm with more than half the thickness of the WT biofilm. **B.** Apical projections of representative sections of each biofilm. A wild type strain forms a robustly filamentous biofilm, whereas *ume6*Δ/Δ, *brg1*Δ/Δ, and *ume6*Δ/Δ *brg1*Δ/Δ mutant strains form biofilms with largely reduced filamentation. The *UME6:BRG1* strain forms a biofilm with visibly filamentous cells.

Figure 5. RNA expression levels in mutant and validation strains. RNA was extracted from SN250 WT, *ume6*Δ/Δ mutant, *ume6*Δ/Δ *brg1*Δ/Δ double mutant, and *UME6:BRG1* validation strains grown for 4 hours in YPD at 37°C, then analyzed with a NanoString nCounter. Three biological replicates were assayed, and the average of all replicates was used for further analysis. *brg1*Δ/Δ mutant gene expression was drawn from Woolford et. al, assayed under the same conditions as this experiment (4 hr YPD, 37°C) and analyzed using the same nanoString codeset (33). **A.** A heat map shows $\log_2$ fold expression of environmentally responsive genes between strains normalized against expression in a *ume6*Δ/Δ *brg1*Δ/Δ double mutant. Log_2 fold expression is depicted on a color gradient ranging between -3.0 (blue) and +3.0 (yellow). Tree depicts clustering of strains according to an algorithm for minimum Manhattan distance (MeV software). The *ume6*Δ/Δ mutant, *brg1*Δ/Δ mutant, and *ume6*Δ/Δ *brg1*Δ/Δ double mutant cluster, showing similar profiles with reduced expression of environmentally responsive genes. The *UME6:BRG1* validation strain clusters with the wildtype, showing similar expression profiles. **B.** Mean fold change in expression of hyphal associated genes *ALS3*, *ECE1*, *HWP1*, *IHD1*, and *RBT1* between mutant strains and validation strain normalized against wild type expression. Values are mean (SD). ** indicates significant difference in mean fold change in comparison to the mutant strains (Tukey-Kramer test, $p < 0.01$) **C.** Mean normalized counts of *UME6* and *BRG1* mRNA between WT, mutant, and validation strains. Values are mean (SD). ** indicates significant difference in mean RNA counts in comparison to all other strains (Tukey-Kramer test, $p < 0.01$).

Supplemental Files

Supplemental file S1. Table of *Candida albicans* strains.

Supplemental file S2. Table of oligonucleotide primer sequences.

Supplemental file S3. Table of Nanostring measurements of RNA levels. Raw data, normalized probe counts, and gene expression ratios are given in three different tabs.

FIGURES, CHAPTER 3

Figure 1

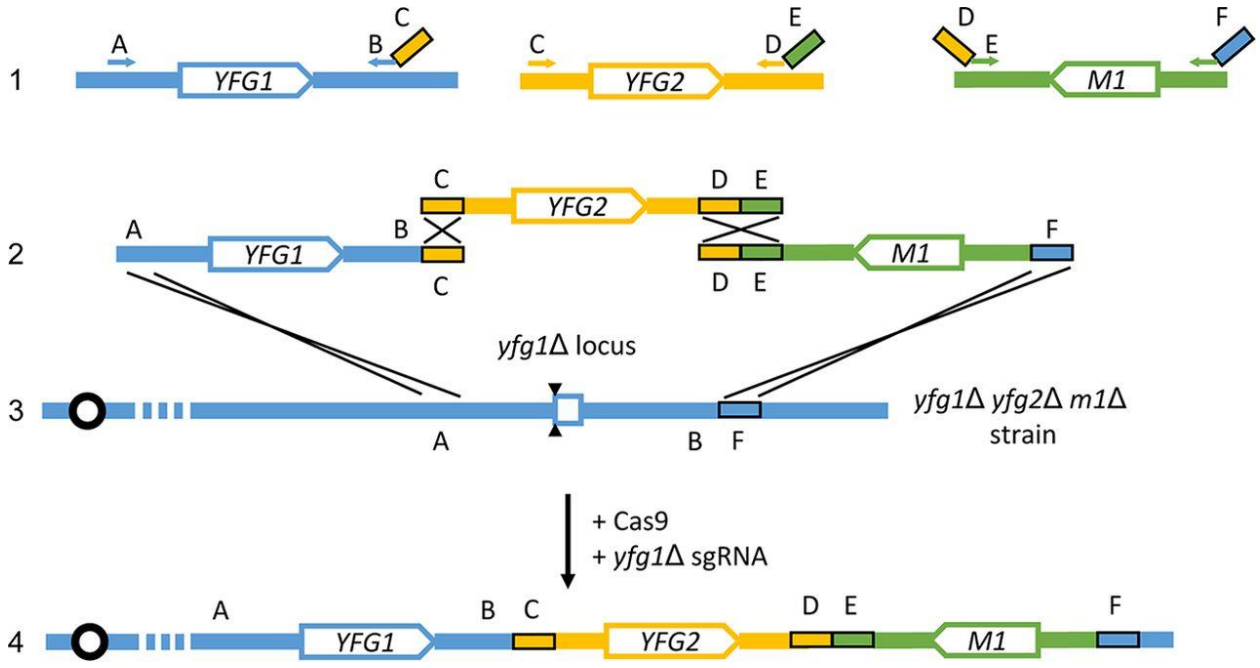


Figure 2

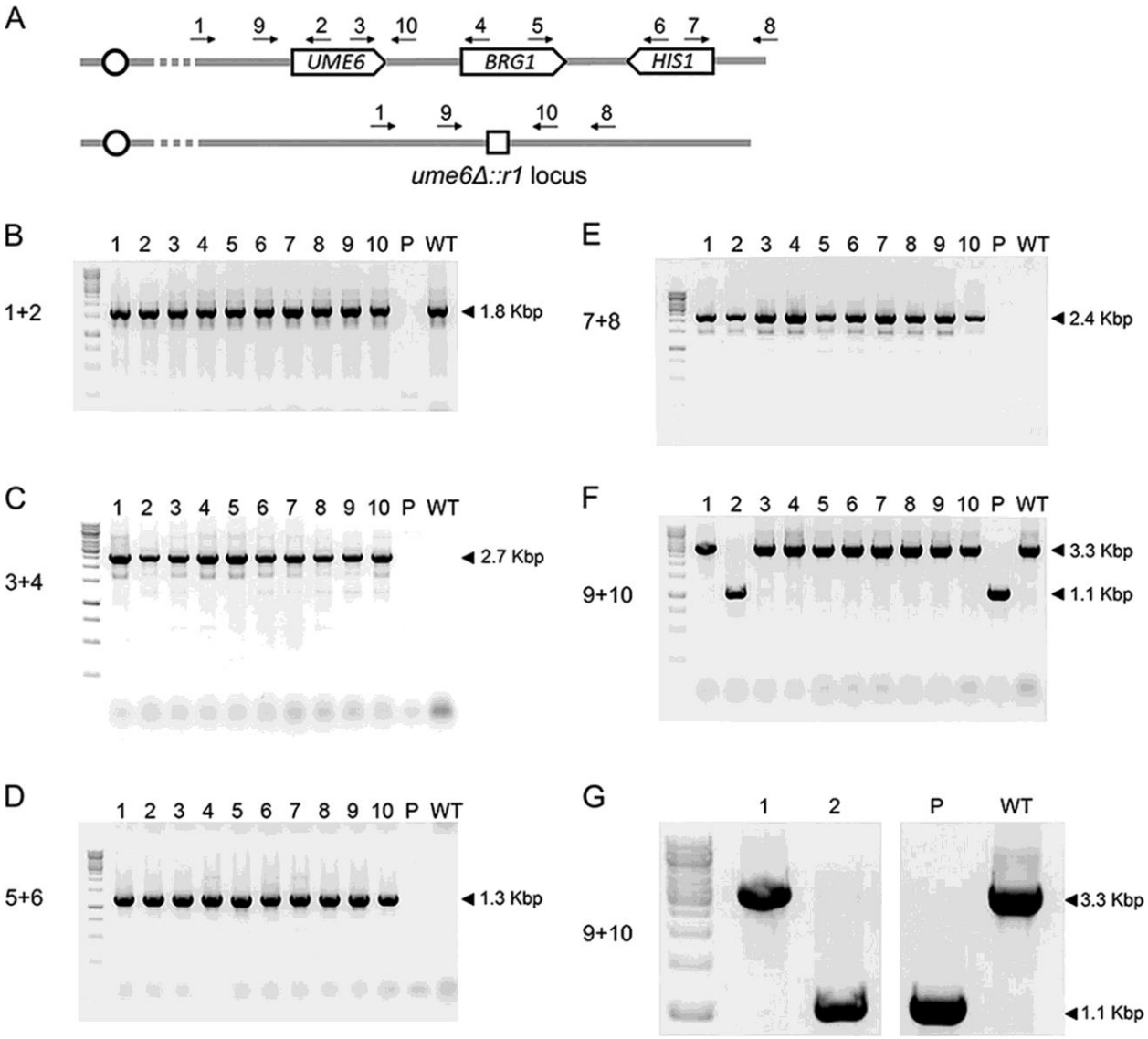


Figure 3

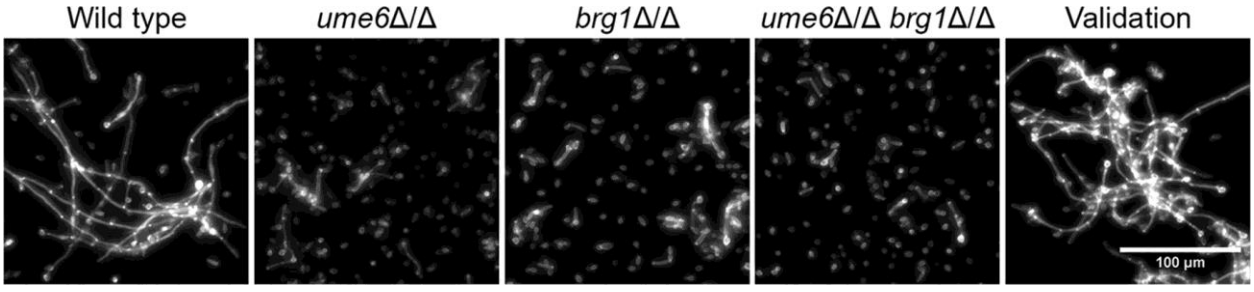


Figure 4

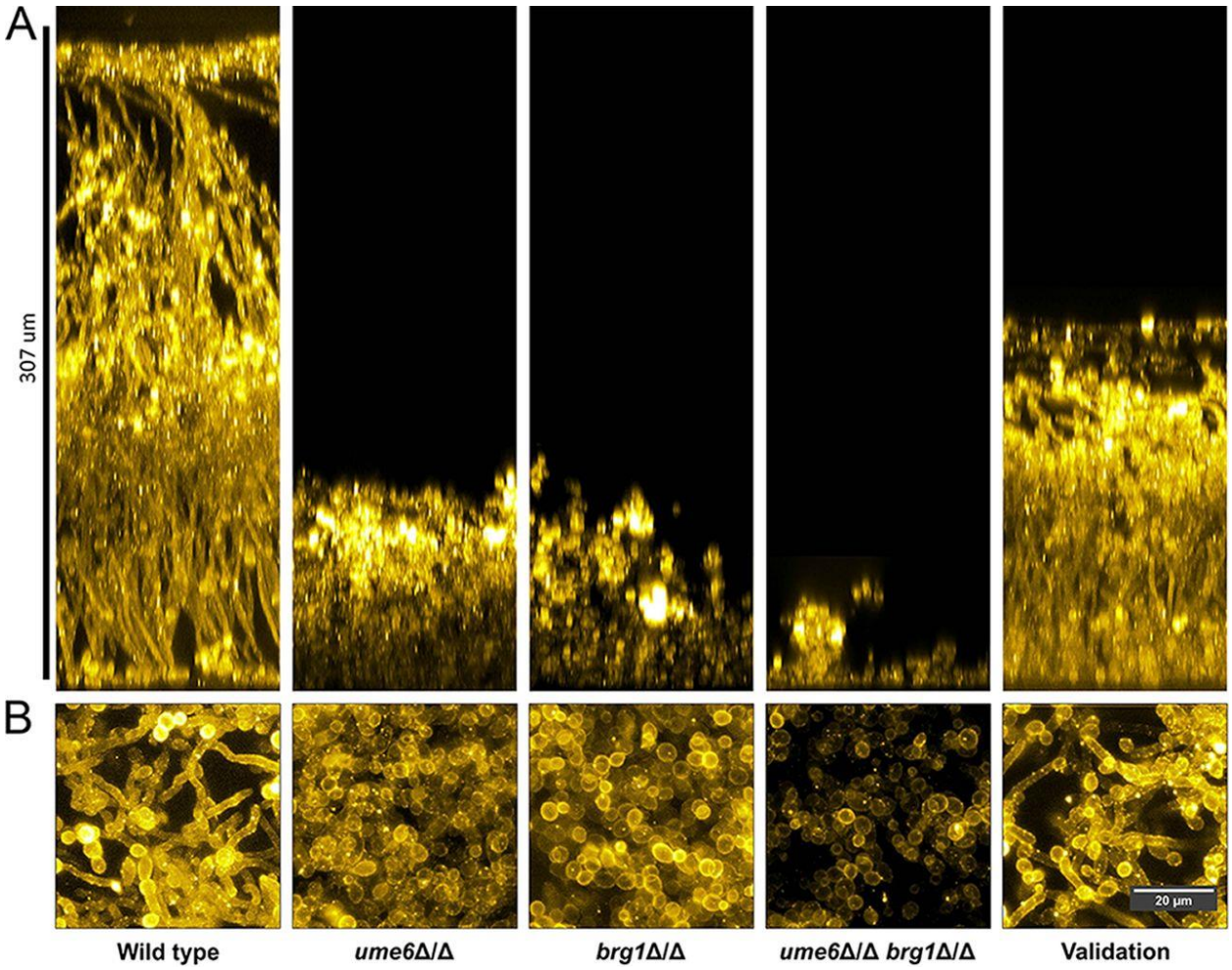
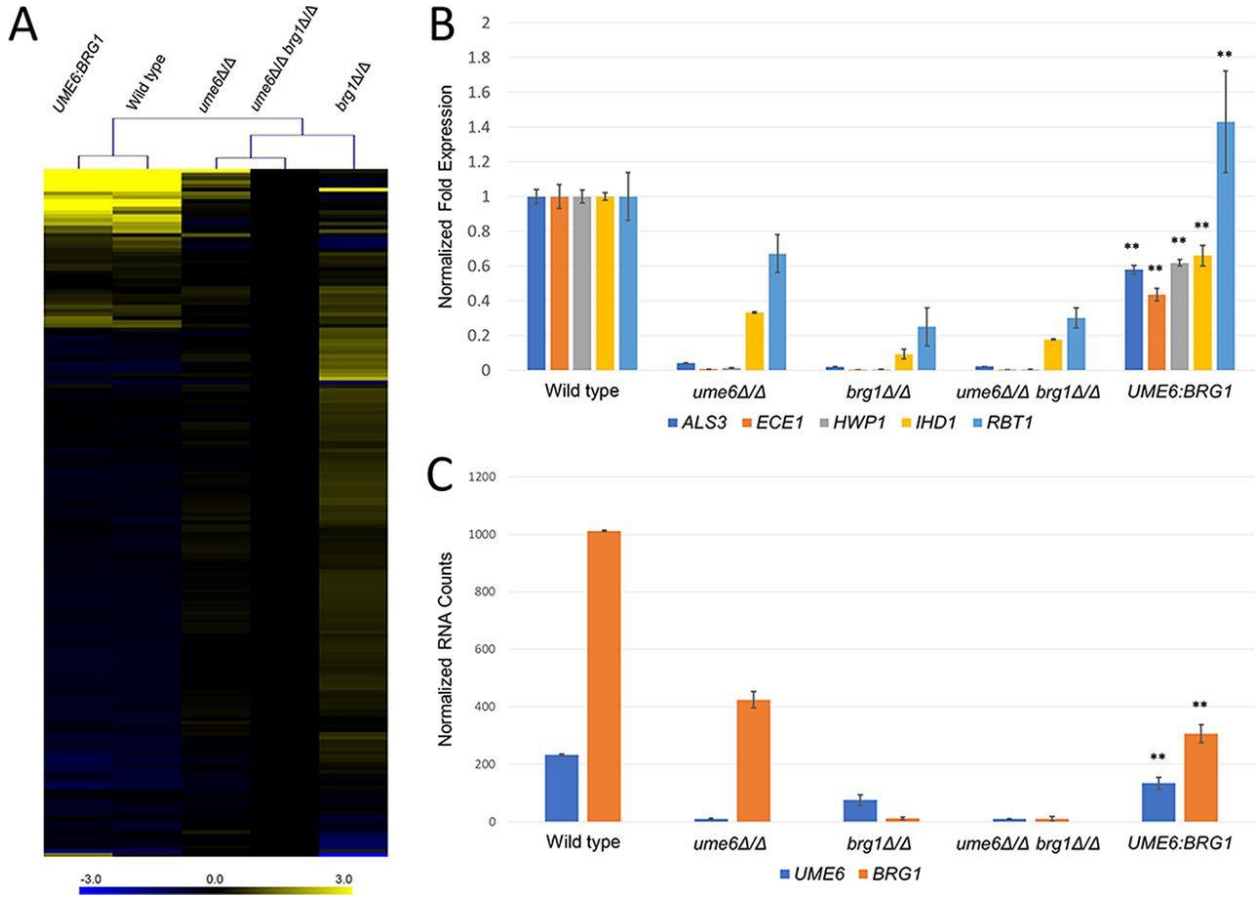

A
307 um
B
20 µm
Wild type
ume6Δ/Δ
brg1Δ/Δ
ume6Δ/Δ brg1Δ/Δ
Validation

Figure 5


A
UME6:BRG1
Wild type
ume6Δ/Δ
ume6Δ/Δ brg1Δ/Δ
brg1Δ/Δ
-3.0
0.0
3.0
B
Normalized Fold Expression
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Wild type
ume6Δ/Δ
brg1Δ/Δ
ume6Δ/Δ brg1Δ/Δ
UME6:BRG1
ALS3
ECE1
HWP1
IHD1
RBT1
C
Normalized RNA Counts
1200
1000
800
600
400
200
0
Wild type
ume6Δ/Δ
brg1Δ/Δ
ume6Δ/Δ brg1Δ/Δ
UME6:BRG1
UME6
BRG1

TABLES, CHAPTER 3

Table 1

SN152 His^+ Leu^+ Arg^+ transformation outcomes

Expt or parameter	No. of His^+ colonies with no sgRNA	No. of His^+ colonies with sgRNA
1	4	831
2	3	716
3	0	402
Avg ± SD	2.33 ± 2.08	650 ± 222
% prototrophic	83	94 ± 2.7

Chapter 4: Circuit diversification in a biofilm regulatory network

Manning Y. Huang, Carol A. Woolford, Gemma May, C. Joel McManus, and Aaron P. Mitchell

FOREWORD

This chapter is lightly modified from the publication version as appeared in PLOS Pathogens, May 2019. Dr. Carol Woolford carried out the Nanostring experiments described in this chapter, Gemma May prepared RNA-seq libraries, and Dr. C. Joel McManus and I analyzed the data. I also carried out the rest of the work described in this chapter. All referenced figures, including supplemental figures are available at the end of the chapter. Supplemental tables have not been included, but are available online at: https://doi.org/10.1371/journal.ppat.1007787

ABSTRACT

Genotype-phenotype relationships can vary extensively among members of a species. One cause of this variation is circuit diversification, the alteration of gene regulatory relationships among members of a species. Circuit diversification is thought to be a starting point for the circuit divergence or rewiring that occurs during speciation. How widespread is circuit diversification? Here we address this question with the fungal pathogen *Candida albicans,* which forms biofilms rich in distinctive hyphal cells as a prelude to infection. Our understanding of the biofilm/hyphal regulatory network comes primarily from studies of one clinical isolate, strain SC5314, and its marked derivatives. We used CRISPR-based methods to create mutations of four key biofilm transcription factor genes – *BCR1, UME6, BRG1,* and *EFG1* – in SC5314 and four additional clinical isolates. Phenotypic analysis revealed that mutations in *BCR1* or *UME6* have variable impact across strains, while mutations in *BRG1* or *EFG1* had uniformly severe impact. Gene expression, sampled with Nanostring probes and examined comprehensively for *EFG1* via RNA-Seq, indicates that regulatory relationships are highly variable among isolates. Our results suggest that genotype-phenotype relationships vary in this strain panel in part because of differences in control of *BRG1* by *BCR1,* a hypothesis that is supported through engineered constitutive expression of *BRG1*. Overall, the data show that circuit diversification is the rule, not the exception, in this biofilm/hyphal regulatory network.

INTRODUCTION

Each species has broad properties that define its members, yet individuals present diversity that reflects the events of their lineage. Although some phenotypic differences between individuals arise from single allelic differences or gene acquisitions, the vast majority represent the interplay of multiple genetic and epigenetic differences [1,2,3]. Natural variation has been measured through assays of biological phenotypes such as fitness, disease susceptibility, or cellular differentiation, and through molecular phenotypes such as the expression of sets of genes. The impact of natural variation is also manifested in genetic background effects on the phenotypes of defined mutations. Comparison of large scale gene knock-out or knock-down collections in pairs of *Saccharomyces cerevisiae* [4,5] and *Caenorhabditis elegans* [6] strains has shown that genetic background effects are widespread, affecting single gene loss-of-function phenotypes for up to 20% of genes. The implication of such studies, as proposed by Gasch and colleagues, is that network relationships between genes may vary considerably among representatives of the same species [7].

The clearest example to date in fungi of species-level natural variation in network architecture comes from Chin et al., who studied adherence in two strains of *S. cerevisiae* [8]. A MAP Kinase pathway (fMAPK) is required for adherence and expression of the adhesin gene *FLO11* in strain Σ1278b but not in strain S288c. Crosses between the strains indicated that the regulation of fMAPK-dependence is genetically complex, though a cloning-based rescue strategy defined one modifier

locus, *RPI1*, that specifies a transcription factor. Rpi1 can bypass the fMAPK pathway through its ability to bind to the 5' region of *FLO11,* an ability enabled by the *RPI1* allele of S288c and abolished by the *RPI1* allele of Σ1278b [8]. Therefore, these two representatives of the same species rely upon distinct signaling pathways - either an fMAPK-dependent pathway or an Rpi1-dependent pathway - to control expression of *FLO11* and, ultimately, adherence [8]. Chin et al. hypothesized that the natural variation in regulatory relationships that they observed within a species, which they call "circuit diversification," is a precursor to the evolutionary rewiring and circuit divergence that is observed between species.

How prevalent is circuit diversification among members of a species? What is the extent of its impact? Here we use the fungal pathogen *Candida albicans* to address these questions. We focus on two well characterized virulence traits: its ability to grow as hyphae and to produce a biofilm [9,10]. Hyphae are tubular arrays of cells that can be hundreds of microns in length, and hypha-associated genes specify adhesins, hydrolases, and the toxin Candidalysin that together cause tissue damage [10,11]. Biofilms are multicellular surface-bound communities that produce an extracellular matrix and are recalcitrant to antimicrobial treatment [12]. Biofilms of *C. albicans* are rich in hyphae, and genetic studies indicate that biofilm production depends upon hyphae in vitro and in animal infection models [12]. Biofilm formation is connected to virulence because biofilm on implanted medical devices is a major source of infection [12].

Our understanding of *C. albicans* biofilm formation comes primarily from studies of one clinical isolate, strain SC5314, and its derivatives CAI-4, BWP17, and SN152, whose markers facilitate genetic manipulation. Among the most well characterized biofilm regulators are the transcription factors (TFs) Efg1, Bcr1, Ume6, and Brg1 ([13,14,15,16,17,18]; reviewed in [19]). A deletion mutation affecting any one can cause a biofilm defect, depending upon the precise growth conditions. All four TFs are also required under many conditions for normal hyphal formation, expression of hypha-associated genes, and virulence in animal models. These TFs are interconnected through their control of overlapping sets of target genes and of one another's expression [19].

Because biofilm production and hyphal formation have been extensively characterized, this network provides a valuable starting point for an appraisal of natural variation. Uniform network architecture may prevail among *C. albicans* isolates, or circuit diversification may prevail. We test these possibilities through analysis of four different single gene deletion mutations in five different *C. albicans* clinical isolates. Our results show that the gene expression impact of regulatory network defects is highly variable among strains, and thus argue that circuit diversification is widespread.

RESULTS

Natural variation in biofilm production

Our studies employed five *C. albicans* clinical isolates: SC5314 (clade 1), P76067 (clade 2), P57055 (clade 3), P87 (clade 4), and P75010 (clade 11) [20,21]. SC5314 is a dermatological isolate and is the standard laboratory strain for most molecular and genetic studies; P76067, P57055, and P75010 are bloodstream isolates; P87 is an oral isolate. These strains were chosen to represent the major clades of clinical isolates and thus to capture the range of genetic diversity.

Biofilm production was assayed at the end of a 24 hr incubation in RPMI+serum medium at 37 degrees. These conditions induce biofilm formation strongly in strain SC5314 (Figure 1, left column, and S1 Figure). Biofilm depth, visualized by confocal microscopy, was substantial for strains SC5314 and P76067, intermediate for strains P57055 and P87, and minimal for strain P75010 (Figure 1, left column, and S1 Figure). These results indicate that biofilm production varies among this set of isolates.

Confocal imaging was used to assay for presence of hyphae in biofilms. Side-view (Figure 1, left column) and apical (Figure 2, left column) confocal projections revealed the presence of abundant hyphae in the four strong and intermediate biofilms. No hyphae were evident in the minimal biofilm produced by strain P75010. These results

are consistent with the conclusion from extensive mutant analysis in the strain SC5314 background that hyphal formation is required for biofilm formation [19].

We also assayed hyphal formation by each strain under planktonic growth conditions (4 hr, RPMI+serum medium, 37 degrees). The strong and intermediate biofilm formers produced abundant long hyphae (Figure 3, left column). The intermediate biofilm forming strain P57055 produced slightly unusual hyphae; many had bends at ~20 micron intervals. The minimal biofilm forming strain P75010 yielded infrequent hyphae under these conditions. Quantitative measurements confirmed these qualitative impressions: hyphae were less abundant, and hyphal unit cell lengths were smaller, in strain P75010 than in the strong and intermediate biofilm formers (S2 Figure). These assay results indicate that production of planktonic hyphae correlates with production of biofilm hyphae in this panel of strains.

Genotype-phenotype relationships

To assess natural variation in genetic control over biofilm production, we created deletion mutations for each of the biofilm regulatory genes *BCR1, UME6, BRG1,* and *EFG1* in all five strains. Mutants were assayed for biofilm production under RPMI+serum growth conditions. A *bcr1*Δ/Δ mutation had little impact under these conditions on biofilm production by the two strong biofilm formers, strains SC5314 and P76067: mutant biofilm depth (Figure 1) and hyphal content (Figure 2) were comparable to those of the respective wild-type strains. However, we noted regional separation of

the basal and upper biofilm layers in these mutants (Figure 1). In contrast, a *bcr1*Δ/Δ mutation impaired biofilm production by the two intermediate biofilm formers, strains P57055 and P87: biofilm depth and hyphal content were severely reduced (Figures 1, 2). A *bcr1*Δ/Δ mutation had little effect on the weak biofilm former, strain P75010 (Figures 1, 2). Pannanusorn et al., in pioneering studies of a set of *Candida parapsilosis* clinical isolates, also observed that impact of *bcr1*Δ/Δ mutations was highly strain-dependent in that species [22]. Our results indicate that *BCR1* is dispensable for biofilm production in some *C. albicans* strain backgrounds and essential for biofilm production in others.

A *ume6*Δ/Δ mutation had broad effects on biofilm production: it caused a partial or severe impairment in all of the strong and intermediate biofilm former backgrounds. Biofilm depth (Figure 1) and hyphal content (Figure 2) were reduced. Biofilm disruption by the *ume6*Δ/Δ mutation was particularly severe in the intermediate biofilm former P57055, perhaps due to the absence of biofilm hyphae (Figure 2). A *ume6*Δ/Δ mutation had little measurable effect on the weak biofilm former P75010 (Figures 1, 2). These results show that Ume6 functional impact varies with strain background, as is the case with Bcr1.

Both *brg1*Δ/Δ and *efg1*Δ/Δ mutations caused severe impairment of biofilm production in the strong and intermediate biofilm formers. Biofilm depth was reduced to ~20 microns (Figure 1), and hyphal content was nearly or entirely eliminated (Figure 2). The mutations had little effect on the already weak biofilms formed in the strain P75010

genetic background. These results show that Brg1 and Efg1 have broad functional impact on phenotype that varies minimally with strain background.

We also assayed the effect of each mutation on production of planktonic hyphae. The results (Figure 3 and S3 Figure) correlated generally with production of biofilm hyphae (Figure 2). Reconstituted derivatives of all mutants, in which one or two copies of the deleted gene were re-introduced, regained hyphal formation ability comparable to the respective wild-type strains (S4 Figure). Interestingly, P75010 derivatives that carried the *BRG1* and *EFG1* alleles from SC5314 displayed increased hyphal production compared to P75010 (S4 and S5 Figures). These results support the conclusion that the magnitude of impact on phenotype of several of the transcription factors varies with strain background.

Natural variation in network architecture

Results above indicate that several biofilm regulatory mutations vary in phenotypic severity among the clinical isolates. To explore this conclusion at the level of gene expression, we conducted Nanostring profiling of each wild-type and regulatory mutant strain. Growth conditions were identical to those for the hyphal induction assays. RNA levels were measured for 181 genes, including 60 genes that have been connected through function or expression to hyphae or biofilms. RNA levels in each mutant were compared to the respective wild type in order to calculate fold changes (S1 Table). The

results revealed that gene regulatory relationships are strongly contingent upon strain background.

One indication of regulatory variation across strains comes from a count of the number of significantly up- or down-regulated genes in each mutant strain compared to their respective wild-type strains (Table 1). For example, in the SC5314 strain background, there were 23 genes whose RNA levels were altered significantly (≥2-fold, FDR=0.1) by a *bcr1*Δ/Δ mutation. In the P57055 background, there were 58 genes whose RNA levels were altered significantly by a *bcr1*Δ/Δ mutation. Across all backgrounds, only 12 genes responded consistently to a *bcr1*Δ/Δ mutation ("Common" column, Table 1). The overall lack of concordance presented by *bcr1*Δ/Δ mutations was recapitulated by the other mutations: the number of responsive genes varied by a factor of 2 among strain backgrounds, and the number of shared responsive genes ("Common") was fewer than half of the number of responsive genes in any background. A similar outcome was observed if only the criterion of an FDR=0.1 was applied without a fold-change requirement (S2 Table). These results indicate that there is substantial variation in regulatory relationships within the *C. albicans* species.

Extensive variation is also seen in the architecture of the biofilm/hyphal regulatory network defined by the mutants and their gene expression impact (Figure 4). Some genes that are annotated to hyphal formation, such as *CHT2* and *SOD5*, varied considerably with respect to strain background in their dependence on specific TFs (Figure 5). The variable response of *CHT2* was particularly noteworthy because its 5'

region is bound by Bcr1 and Brg1 as shown by overlapping binding peaks centered approximately 2250 bp upstream of the start codon [16], an indication that it is a direct target of those two TFs. Interestingly, among all five isolates, no SNPs were identified in Bcr1 and Brg1 motifs in this region. An additional illustration of network variation comes from the regulation of the TF genes *BRG1* and *UME6* (Figures 4 and 5). In the two strong biofilm formers, SC5314 and P76067, Bcr1 is not required for expression of *BRG1* and *UME6*. In the intermediate and weak biofilm formers, Bcr1 is required for expression of both *BRG1* and *UME6.* These dependency relationships provide a possible explanation for the greater impact of the *bcr1*Δ/Δ mutation on gene expression and biological phenotypes in the intermediate biofilm formers than in the strong biofilm formers (Table 1; Figures 1-3). Overall, these results indicate that regulatory network architecture is strongly contingent upon strain background. In addition, the observation that many genes are dependent upon a TF in one strain background but not another is evidence for circuit diversification among these strains.

We distilled gene expression changes into a common network of regulatory relationships that are shared among all five strains (Figure 4; S1G Table). In the 181 assayed genes, 60 were annotated to GO terms related to hyphae or biofilm. Compared to regulatory relationships determined solely in SC5314, a larger proportion of the relationships defined by this common network were with these hyphae or biofilm annotated genes (p=0.014, Fisher's exact test). The common network also trended toward enrichment for direct targets of the TFs (p=0.083 compared to SC5314, Fisher's exact test) as defined by ChIP-Seq experiments [16,23]. These observations suggest

that the common target genes found in diverse strains may give clearer functional insight into their regulators than the target genes found in any one strain.

Genome-wide Efg1 regulon analysis

For a genome-wide view of regulatory relationships among strains, we carried out RNA-Seq analysis of the five clinical isolates and their *efg1*Δ/Δ derivatives (S3 Table). Each clinical isolate was compared with its corresponding *efg1*Δ/Δ mutant in order to define Efg1-responsive genes. The gene expression impact of the *efg1*Δ/Δ mutation varied considerably among clinical isolates (Table 2). The number of Efg1-responsive genes ranged from 523 (P76067) to 864 (SC5314). Approximately 15-27% of the genes that responded to Efg1 in any one strain did not respond in any of the other four strains (Figure 6B). Many additional genes were Efg1-responsive only in a subset of genetic backgrounds (Figure 6B). Overall, these genome-wide data support the concept that gene expression targets vary considerably among *C. albicans* species representatives, and indicate that circuit diversification frequently affects Efg1 target genes.

Gene expression profiles converged on 177 core Efg1-responsive genes (21-34% of total) that were up- or down-regulated in *efg1*Δ/Δ mutants of every strain background (Figure 6B). These core Efg1-responsive genes included 138 Efg1-activated genes (i.e., down-regulated in *efg1*Δ/Δ mutants) and 39 Efg1-repressed genes (i.e., up-regulated in *efg1*Δ/Δ mutants). Core Efg1-activated genes were enriched for the GO term biofilm formation (p=4.86e-07) (Figure 6A,C; S4 Table). This enrichment was

greater than observed with the Efg1-activated genes of any individual strain (Figure 6A,C; S4 Table). Core Efg1-repressed genes were enriched for GO terms that include cell surface (p=6.94e-07) and cell wall (p=6.78e-07) (S4 Table). This enrichment was comparable to that observed with individual strains. We found that 24% of core Efg1-responsive genes were direct Efg1 targets, based on chromatin immunoprecipitation data [16,23], whereas 16-22% of Efg1-responsive genes in individual strains were direct targets (Table 2). Compared to the proportion of direct targets among all SC5314 Efg1-responsive genes, the proportion of direct targets among core Efg1-responsive genes was greater (p = 0.036, Fisher's exact test), though it only trended toward greater in comparisons to some other strains. Overall, these observations indicate that core Efg1-responsive genes align well with what is known about Efg1 function.

The Efg1-activated genes of several strains were enriched for carbohydrate metabolic functions (Figure 6A,C, S6 Figure), which are mainly glycolytic genes, as expected from prior studies [24]. However, there was no enrichment for these functions in the core Efg1-activated gene set. Their exclusion from core genes is based on properties of one strain, P76067. Examination of individual gene expression responses shows that these genes display less dependence on Efg1 for expression in strain P76067 compared to the other strains (S3 Table). Therefore, the impact of Efg1 on carbohydrate metabolic genes behaves as a quantitative trait among *C. albicans* isolates.

The unique Efg1-responsive genes in each strain (S3 Table) ranged from 97 (strain P57055) to 234 (strain SC5314). They were roughly split between Efg1-activated and -

repressed genes (Figure 6B). We found only minor enrichments for GO terms among most of these gene sets, and no significant enrichment at all among the SC5314 strain-specific Efg1-responsive genes. Although these genes do not share distinguishing GO assignments, there are prospective functionally relevant genes among them. For example, the SC5314 *efg1*Δ/Δ mutation leads to significantly reduced expression of *SUN41,* which is required for biofilm formation [25,26]. Therefore, strain-specific Efg1-responsive genes may contribute to the mutant phenotype, but they do not reveal broad pathways that respond to Efg1 in a strain-specific manner.

Functional impact of circuit diversification

The *BCR1-BRG1* relationship provides a simple illustration of circuit diversification: *BCR1* is required for *BRG1* expression in intermediate but not strong biofilm formers (Figures 4 and 5). We hypothesized that this regulatory difference was the reason that *BCR1* is required for biofilm production by intermediate but not strong biofilm formers under our assay conditions. Specifically, reduced *BRG1* expression may contribute to the biofilm defect of *bcr1*Δ/Δ mutants in intermediate biofilm formers, while constitutive *BRG1* expression may permit biofilm production by *bcr1*Δ/Δ mutants in strong biofilm formers.

This hypothesis predicts that constitutive *BRG1* expression will permit biofilm production in a *bcr1*Δ/Δ mutant in an intermediate biofilm former. We tested this hypothesis with strain P57055, an intermediate biofilm former that transforms efficiently. We fused one

allele of *BRG1* with the *TDH3* promoter in P57055 *BCR1/BCR1 BRG1/BRG1* and *bcr1*Δ/Δ *BRG1/BRG1* strains to create *BRG1/TDH3-BRG1* derivatives. We then compared four strains of genotypes *BCR1/BCR1 BRG1/BRG1, bcr1*Δ/Δ *BRG1/BRG1, BCR1/BCR1 BRG1/TDH3-BRG1,* and *bcr1*Δ/Δ *BRG1/TDH3-BRG1.* Using Nanostring, we confirmed that *BRG1* RNA levels were Bcr1-dependent in the *BRG1/BRG1* strains and Bcr1-independent in the *BRG1/TDH3-BRG1* strains (Figure 7A). Although the *TDH3* promoter is often used for overexpression, in this case it did not yield greatly elevated *BRG1* expression. As predicted by the hypothesis, biofilm production was also Bcr1-dependent in the *BRG1/BRG1* strains and Bcr1-independent in the *BRG1/TDH3-BRG1* strains (Figure 7C). As a further functional test of the hypothesis, we examined planktonic hyphal formation. In the P57055 background, the *bcr1*Δ/Δ mutant had reduced length of hyphal cell compartments and a reduced ratio of hyphae to yeast cells (Figure 7B,D). In the P57055 *BRG1/TDH3-BRG1* derivatives, the *bcr1*Δ/Δ mutant did not display these phenotypes (Figure 7B,D). Therefore, the phenotypic impact of a *bcr1*Δ/Δ mutation in the P57055 background depends upon the *BCR1-BRG1* regulatory relationship.

What is the mechanism behind divergent dependence of *BRG1* expression on Bcr1? One hypothesis is that the cis regulatory elements of *BRG1* alleles may contain SNPs that allow Bcr1-independent *BRG1* expression in some strains but not others. We tested this hypothesis by constructing P57055 *bcr1*Δ/Δ mutant strains carrying $BRG1^{SC5314}$ or $BRG1^{P57055}$ alleles at the *MDR1* locus. These alleles contained 1642 bp of the *BRG1* upstream region and 712 bp of the *BRG1* downstream region. Ectopic expression of

BRG1^{SC5314} and *BRG1*^{P57055} in this manner complemented the hyphal formation defect of a P57055 *brg1*Δ/Δ mutant strain, demonstrating that the cis contexts captured in these regions were sufficient for *BRG1* expression and function (S7AFigure). However, ectopic expression of *BRG1*^{SC5314} or *BRG1*^{P57055} in the P57055 *bcr1*Δ/Δ mutant failed to rescue hyphal formation (S7B Figure and S5 Table). Furthermore, strains expressing *BRG1*^{SC5314} were not significantly different from strains expressing *BRG1*^{P57055} in hyphal formation capacity. We conclude then that the cause of Bcr1-independent *BRG1* expression in SC5314 does not lie solely in cis regulatory element SNPs carried in these allelic segments.

DISCUSSION

Our studies address whether genetic regulatory relationships are uniform within the species *C. albicans*. We approached the problem through measurement of biological phenotypes and gene expression changes that result from mutations in each of four TF genes in the biofilm/hyphal regulatory network. Two of the TF gene mutations, *bcr1*Δ/Δ and *ume6*Δ/Δ, had variable phenotypic impact among the strains. These mutations also had variable gene expression impact, an outcome that might have been predicted from phenotypic variation. The other two TF mutations, *brg1*Δ/Δ and *efg1*Δ/Δ, had uniform phenotypic impact, yet still had highly variable gene expression impact. These observations argue that circuit diversification – variation in regulator-target relationships within a species – is prevalent for this biofilm/hyphal regulatory network.

The traits we examined, biofilm production and hyphal formation, are known to vary quantitatively among *C. albicans* isolates [20,21,27,28,29]. Hence it seemed reasonable that gene expression impact of key biofilm/hyphal regulators would vary as well. We were nonetheless struck by the extent of strain-specific gene expression changes we observed; only about half of the gene expression response to a mutation in any one strain was shared among the other four strains. The fact that even the mutation with the strongest and most uniform biological phenotypes, *efg1*Δ/Δ, caused variable gene expression impact across strains is especially noteworthy, because large scale dual-strain comparisons of loss-of-function defects have relied on biological

phenotypes [4,5,6]. Our results argue that biological phenotype measurements may underestimate the difference in impact of a mutation in two different strains.

What sorts of variation do we see in TF-target gene relationships? Regulation of *SOD5, BRG1,* and *UME6* (Figure 5) represents one frequent pattern: expression of each is down-regulated in a mutant, such as *bcr1*Δ/Δ, in some strains but not others. Analogous observations were made with expression of *S. cerevisiae FLO11* and its control by the fMAPK pathway by Chin and colleagues [8]. The regulation of *CHT2* is more complex, for example in its response to Efg1. It is up-regulated in an *efg1*Δ/Δ mutant in SC5314, as shown previously [30], but it is down-regulated in *efg1*Δ/Δ mutants of other strains. Efg1 is known to function as an activator at some promoters and a repressor at others [31,32] in strains derived from SC5314. However, our results raise the possibility that Efg1 may function as a repressor or an activator at a single promoter, depending upon the strain background. These examples illustrate strain-dependent differences in TF-target gene relationships that are indicative of circuit diversification.

Variation in biofilm/hyphal network architecture has clear functional impact, as illustrated by strain differences in the *BCR1-BRG1* relationship. A *bcr1*Δ/Δ mutation had little effect on biofilm production or *BRG1* expression in two strong biofilm formers, and caused a severe defect in both biofilm production and *BRG1* expression in two intermediate biofilm formers. Because *BRG1* was required for biofilm production in all strains, we considered that differences in *BCR1*-dependence of biofilm production may

arise from differences in *BCR1*-dependence of *BRG1* expression. This hypothesis was supported by the finding that constitutive *BRG1* expression eliminated *BCR1*-dependence of biofilm production in an intermediate biofilm former. Prior studies have shown that Bcr1 and Brg1 have considerable functional overlap: among 252 direct Bcr1 target genes identified by ChIP-seq, 194 are Brg1 direct targets as well [16]. Overlap of target genes may be the reason that *BCR1* is required for biofilm formation only when *BRG1* levels are low. We cannot find strain differences in the Bcr1 binding sites upstream of *BRG1*. Furthermore, cis-regulatory elements of the *BRG1*^{SC5314} allele were not sufficient for *BRG1* function in the P57055 *bcr1*Δ/Δ mutant strain. We infer that variation in the *BCR1-BRG1* relationship arises from differences in trans-acting factors that can compensate for absence of Bcr1. This inference is consistent with the conclusion from many studies the bulk of gene expression variance between individuals arises from differences in trans-acting gene products [1].

Glycolytic genes provide an example of a functionally related group of genes that vary in strength of connection to biofilm regulator Efg1 (Figure 6, S6 Figure). In most strains, the *efg1*Δ/Δ mutation caused a severe reduction in glycolytic gene expression. In contrast, in strain P76067 the *efg1*Δ/Δ mutation caused a mild reduction in glycolytic gene expression. Inspection of the RNA-Seq data shows that expression of *GAL4*, an activator of glycolytic genes [33], is strongly reduced in most *efg1*Δ/Δ mutants but only mildly reduced in the P76067-derived *efg1*Δ/Δ mutant. Efg1 does not bind directly to the *GAL4* upstream region [16,23]. Therefore, this example of circuit diversification also

seems to arise from differences in activity of trans-acting factors that, in this case, compensate for absence of Efg1.

Our data provide the first view of *C. albicans* natural variation from the perspective of gene expression profiles, and several manifestations of strain variation are evident. For example, compared to SC5314, all isolates had significantly increased RNA levels for various cell wall-related genes (S6C Table). Also, higher *BCR1* and *BRG1* RNA levels among isolates correlate with lower RNA levels for ribosome-related genes (S6A Table). These correlations may reflect natural variation in TOR pathway activity, which is known to promote ribosome biogenesis and inhibit Bcr1-dependent adhesin expression [34]. Although high resolution trait mapping is not yet feasible for *C. albicans*, a candidate gene-based approach could unravel the causes for these strain differences and their functional consequences.

A valuable practical application of multi-strain analysis is the distillation of a common set of genetic regulatory relationships. This outcome was suggested by our small-scale Nanostring profiling, but was most clearly documented through genome-wide analysis of Efg1-responsive genes. Specifically, the common Efg1-activated gene set was significantly enriched for biofilm-related genes, and trended toward enrichment for direct Efg1 target genes, compared to any individual strain's Efg1-activated genes. The common Efg1-repressed gene set was enriched for cell-surface related genes, an enrichment that was not found among Efg1-repressed gene sets for individual strains. These outcomes argue that multi-strain analysis of mutants is significant both for the

validation of conclusions across multiple species representatives, and for its ability to narrow a panel of responsive genes to those with a strong connection to relevant biological processes.

METHODS

Strains and Media

The following *C. albicans* clinical isolate strains were obtained through BEI Resources, NIAID, NIH: *Candida albicans*, Strain P76067, NR-29442; *Candida albicans*, Strain P57055, NR-29439; *Candida albicans*, Strain P87, NR-29453; *Candida albicans*, Strain P75010, NR-29437. All strains and mutants were maintained in 15% glycerol stocks stored at -80°C. Prior to all experiments, strains were grown on YPD (2% Bacto Peptone, 2% dextrose, 1% yeast extract) for 2 days at 30°C, and then cultured overnight in liquid YPD at 30°C with shaking. Transformants were selected on YPD + 400 μg/ml nourseothricin or complete synthetic media (CSM) (2% dextrose, 1.7% Difco yeast nitrogen base with ammonium sulfate and auxotrophic supplements). For phenotypic assays, strains were grown in liquid RPMI-1640 Media (Sigma-Aldrich, Inc., St. Louis) adjusted to pH 7.4 and supplemented with 10% fetal bovine serum (Atlanta Biologicals, Inc., Flowery Branch). A full list of the strains used in this study is provided under supplemental files (S7 Table).

Primers and Plasmid Construction

All primers and plasmids used in this study are provided under supplemental files (S8 Table).

We previously demonstrated that the use of repeat flanked selectable markers allowed for CRISPR-Cas9 induced marker excision in subsequent manipulations [35]. To adapt the *NAT1* marker for marker recycling with this method, we generated vectors containing *NAT1* inserted respectively at the BamHI (pMH05) and XmaI (pMH06) restriction sites in the plasmid YEp24 backbone [36].

To generate plasmid pMH05, the *NAT1* marker was amplified from plasmid pNAT [37] using primers “BamHI_YEp24_H+AdapN/F” and “BamHI_YEp24_H+AdapN/R”. An aliquot of plasmid YEp24 was then digested with BamHI, and digest products were transformed alongside the *NAT1* PCR product into the *Saccharomyces cerevisiae* strain BJ8918 with selection on synthetic media lacking uracil to allow gap repair of the digested YEp24 vector with the *NAT1* PCR product [38]. The resulting vector was recovered from Ura+ transformants using a Zymoprep Yeast Plasmid Miniprep II Kit (Zymo Research, Irvine) and correct integration of *NAT1* at the BamHI restriction site was verified by PCR.

To generate plasmid pMH06, the *NAT1* marker was amplified from plasmid pNAT using primers “XmaI_YEp24_H+AdapN/F” and “XmaI_YEp24_H+AdapN/R”. An aliquot of plasmid YEp24 was then digested with XmaI, and digest products were transformed alongside the *NAT1* PCR product into the *Saccharomyces cerevisiae* strain BJ8918 with selection on synthetic media lacking uracil to allow for gap repair of the digested YEp24 vector with the *NAT1* PCR product. The resulting vector was recovered from Ura+

transformants using a Zymoprep Yeast Plasmid Miniprep II Kit and correct integration of *NAT1* at the XmaI restriction site was verified by PCR.

Auxotrophic Strain Construction

To increase the number of available markers, *HIS1* was deleted in strains P76067, P57055, P87, P75010, and SC5314 using the transient CRISPR-Cas9 system [37]. Each strain was transformed with approximately 1 µg Cas9 DNA cassette, 1 µg CaHIS1 sgRNA DNA cassette, and 3 µg *his1*∆::*r3NAT1r3* repair template. The Cas9 DNA cassette was amplified by PCR from plasmid pV1093 as previously described [37,39]. The CaHIS1 sgRNA DNA cassette was generated using split-joint PCR using previously described protocols with the primers “CaHIS1 sgRNA/F” and “CaHIS1 SNR52/R” [37]. The *his1*∆*::r3NAT1r3* repair template was constructed in two sections using previously described protocols [35]. The first section was amplified from plasmid pMH05 using primers “HIS1 del rNATrBamHI/F” and “NAT1 CRIME/R”. The second section was amplified from plasmid pMH06 using primers “NAT1 CRIME/F” and “HIS1 del rNATrXmaI/R”. Recombination between these two sections yields the full length *his1*∆::*r3NAT1r3* repair template following transformation.

Transformants were selected for nourseothricin resistance, and subsequently replica plated onto CSM lacking histidine to screen for a His- phenotype. Deletion of *HIS1* in candidate transformants was verified by PCR from genomic DNA using primers “CaHIS1 Check/F” and “CaHIS1 Check Int/R” for absence of the *HIS1* ORF, and using

primers “CaHIS1 Check/F” and “NAT1 Check/R” for presence of the *NAT1* marker at the *his1*Δ locus.

Transcription Factor Mutant Strain Construction

To delete *BCR1*, the *his1*Δ strains of each background were transformed with approximately 1 µg Cas9 DNA cassette, 1 µg BCR1-2 sgRNA DNA cassette, 1 µg NAT1-2 sgRNA DNA cassette, and 3 µg *bcr1*Δ*::r1HIS1r1* repair template. Inclusion of the NAT1-2 sgRNA DNA cassette targets a Cas9 mediated double stranded break to the repeat flanked *NAT1* marker at the *his1*Δ::*r3NAT1r3* locus. The segment of vector YEp24 backbone between BamHI and XmaI constitutes the repeats flanking the NAT1 marker. We refer to these repeats as “r3”. The BCR1-2 sgRNA DNA cassette was generated using split-joint PCR with the primers “sgRNA/F BCR1-2” and “SNR52/R BCR1-2”. The NAT1-2 sgRNA DNA cassette was generated using split-joint PCR with the primers “sgRNA/F NAT1-2” and “SNR52/R NAT1-2”. The *bcr1*Δ*::r1HIS1r1* repair template was generated in two parts. The first was amplified from plasmid pMH01 using primers “HIS1 CRIME/F” and “BCR1 del KpnI-rHIS1r/R”, and the second was amplified from plasmid pMH02 using primers “BCR1 del SapI-rHIS1r/F” and “HIS1 CRIME/R”.

Recombination between the direct repeats excises the marker, rendering the strain nourseothricin sensitive and leaving only a single copy of the repeat (*r3*) at the recycled locus [35]. Transformants were selected on CSM medium lacking histidine, and replica plated onto YPD + nourseothricin plates to screen for nourseothricin sensitivity.

Candidate colonies were further genotyped by PCR using primers “BCR1 check up/F” and “BCR1 check int/R” for absence of the *BCR1* ORF, and using primers “BCR1 check up/F” and “CdHIS1 Check Int/R” for presence of the *HIS1* marker at the *bcr1*Δ locus.

To delete *UME6*, the *his1*Δ strains of each background were transformed with approximately 1 µg Cas9 DNA cassette, 1 µg UME6 sgRNA DNA cassette, 1 µg NAT1-2 sgRNA DNA cassette, and 3 µg *ume6*Δ*::r1HIS1r1* repair template. The UME6 sgRNA DNA cassette was generated using split-joint PCR with the primers “sgRNA/F UME6” and “SNR52/R UME6”. The *ume6*Δ*::r1HIS1r1* repair template was generated in two parts. The first was amplified from plasmid pMH01 using primers “HIS1 CRIME/F” and “UME6 del KpnI-rHIS1r/R”, and the second was amplified from plasmid pMH02 using primers “UME6 del SapI-rHIS1r/F” and “HIS1 CRIME/R”. Transformants were selected on CSM media lacking histidine, and replica plated onto YPD + nourseothricin plates to screen for nourseothricin sensitivity. Candidate colonies were further genotyped by PCR using primers “UME6 check up/F” and “UME6 check int/R” for absence of the *UME6* ORF, and using primers “UME6 check up/F” and “CdHIS1 Check Int/R” for presence of the *HIS1* marker at the *ume6*Δ locus.

To delete *BRG1*, the *his1*Δ strains of each background were transformed with approximately 1 µg Cas9 DNA cassette, 1 µg BRG1 sgRNA DNA cassette, 1 µg NAT1-2 sgRNA DNA cassette, and 3 µg *brg1*Δ*::r1HIS1r1* repair template. The BRG1 sgRNA DNA cassette was generated using split-joint PCR with the primers “sgRNA/F BRG1” and “SNR52/R BRG1”. The *brg1*Δ*::r1HIS1r1* repair template was generated in two

parts. The first was amplified from plasmid pMH01 using primers “HIS1 CRIME/F” and “BRG1 del rHISr-KpnI/R”, and the second was amplified from plasmid pMH02 using primers “BRG1 del rHISr-SapI/F” and “HIS1 CRIME/R”. Transformants were selected on CSM media lacking histidine, and replica plated onto YPD + nourseothricin plates to screen for nourseothricin sensitivity. Candidate colonies were further genotyped by PCR using primers “BRG1 check up/F” and “BRG1 check int/R” for absence of the *BRG1* ORF, and using primers “BRG1 check up/F” and “CdHIS1 Check Int/R” for presence of the *HIS1* marker at the *brg1*Δ locus.

Transformations to delete *BRG1* yielded no colonies in the P87 background using this method. To isolate *brg1*Δ mutants in this background, a repair template with extended homology was employed. This cassette was generated in two pieces, using PCR from the genomic DNA of an SC5314 *brg1*Δ::*r1HIS1r1* strain. Primers “BRG1 FarUp/F” with “HIS1 CRIME/R” were used for the first piece, and “HIS1 CRIME/F” and “BRG1 FarDown/R” were used for the second piece.

To delete *EFG1*, the *his1*Δ strains of each background were transformed with approximately 1 µg Cas9 DNA cassette, 1 µg EFG1-2 sgRNA DNA cassette, 1 µg NAT1-2 sgRNA DNA cassette, and 3 µg *efg1*Δ*::r1HIS1r1* repair template. The EFG1 sgRNA DNA cassette was generated using split-joint PCR with the primers “sgRNA/F EFG1” and “SNR52/R EFG1”. The *efg1*Δ*::r1HIS1r1* repair template was generated in two parts. The first was amplified from plasmid pMH01 using primers “HIS1 CRIME/F” and “EFG1 del rHIS1r-KpnI/R”, and the second was amplified from plasmid pMH02

using primers “EFG1 del rHIS1r-SapI/F” and “HIS1 CRIME/R”. Transformants were selected on CSM media lacking histidine, and replica plated onto YPD + nourseothricin plates to screen for nourseothricin sensitivity. Candidate colonies were further genotyped by PCR using primers “EFG1 check up/F” and “EFG1 check int/R” for absence of the *EFG1* ORF, and using primers “EFG1 check up/F” and “CdHIS1 Check Int/R” for presence of the *HIS1* marker at the *efg1*Δ locus.

To generate strains overexpressing *BRG1*, a *NAT1-pTDH3* cassette containing flanking homology to the *BRG1* upstream region was amplified using primers “BRG1 OE/F” and “BRG1 OE/R” from plasmid CJN542 [40]. The P57055 WT and P57055 *bcr1*Δ mutant were then transformed with 3 µg of this *NAT1-pTDH3* cassette, 1 µg of Cas9, and 1 µg of P-BRG1 sgRNA DNA cassette. The P-BRG1 sgRNA cassette was generated using split-joint PCR with primers “sgRNA/F P-BRG1” and “SNR52/R P-BRG1”. Transformants were selected on YPD + nourseothricin plates for the resistant phenotype, and were genotyped by PCR using primers “BRG1 Check Up/F” and “BRG1 Check Int/R” for the presence of one copy of the native *BRG1* promoter, and “NAT1 CRIME/F” and “BRG1 Check Int/R” for presence of the *NAT1-pTDH3* cassette in the *BRG1* promoter region.

Reconstituted Strain Construction

To validate the construction of our TF deletion mutants, we reintroduced a copy of the SC5314 allele of each TF at the TF deletion locus using our concatemer assembly method [41].

A *BCR1* cassette was amplified from SC5314 genomic DNA using primers “BCR1 check up/F” and “BCR1 3’R->pNAT 5’/R”, containing concatenating homology to a *NAT1* marker. The SC5314 *BCR1* allelic segment amplified by these primers contains 277 bp of the *BCR1* upstream region and 399 bp of the *BCR1* downstream region. A *NAT1* marker was then amplified from pNAT using “pNAT for adap/F” and “pNAT 3’R->BCR1down/R”. As no colonies were recovered from the P75010 using these cassettes, A *NAT1* marker with extended homology was amplified from strain MH351 gDNA using “pNAT for adap/F” and “BCR1 fardown/R”.

A *UME6* cassette was amplified from plasmid pSG1-UME6 (provided by K. Lagree) containing a SC5314 *UME6* allele using primers “UME6 Check Up/F” and “UME6 3’R->pNAT 5’/R”, containing concatenating homology to a *NAT1* marker. The SC5314 *UME6* allelic segment amplified by these primers contains 403 bp of the *UME6* upstream region and 399 of the *UME6* downstream region. A *NAT1* marker was then amplified from pNAT using “pNAT for adap/F” and “pNAT 3’R->UME6down/R”.

A *BRG1* cassette was amplified from plasmid pCW1071 containing a SC5314 *BRG1* allele using primers “BRG1 Check Up/F” and “BRG1 3’R->pNAT 5’/R”, containing concatenating homology to a *NAT1* marker. The SC5314 *BRG1* allelic segment

amplified by these primers contains 407 bp of the *BRG1* upstream region and 400 bp of the *BRG1* downstream region. A *NAT1* marker was then amplified from pNAT using “pNAT for adap/F” and “pNAT 3’R->BRG1down/R”.

An *EFG1* cassette was amplified from plasmid pCW861 containing a SC5314 *EFG1* allele using primers “EFG1 Check Up/F” and “EFG1 3’R->pNAT 5’/R”, containing concatenating homology to a *NAT1* marker. The SC5314 *EFG1* allelic segment amplified by these primers contains 153 bp of the *EFG1* upstream region and 401 bp of the *EFG1* downstream region. A *NAT1* marker was then amplified from pNAT using “pNAT for adap/F” and “pNAT 3’R->EFG1down/R”.

The TF-containing cassette and corresponding *NAT1* marker were transformed into the respective TF deletion mutant in all clinical isolate backgrounds, with approximately 2 µg of the TF-containing cassette, 2 µg of the *NAT1* marker cassette, 1 µg of Cas9, and 1 µg of r1 sgRNA DNA cassette. The r1 sgRNA DNA cassette was generated using split-joint PCR with primers “sgRNA/F r1” and “SNR52/R r1”. Heterozygosity or homozygosity at the edited TF locus was determined using the presence or absence of an *r1* scar [35,41] using PCR genotyping with the corresponding “TF Check Up/F” and “r1 check int/R” primers.

BRG1 **Ectopic Expression Strain Construction**

To construct *BRG1* ectopic expression strains, we replaced the *MDR1* ORF with varying *BRG1* alleles using our concatemer assembly method [41].

A cassette containing 1642 bp of *BRG1* upstream sequence, the *BRG1* ORF and 712 bp of *BRG1* downstream sequence was amplified from SC5314 genomic DNA using primers “BRG1 1641 5’F->MDR1 up/F” and “BRG1 712 3’R->pNAT 5’/R”, containing concatenating homology to a *NAT1* marker. A *NAT1* marker was then amplified from pNAT using “pNAT for adap/F” and “pNAT 3’R->MDR1 down/R”. The same process was performed with P57055 genomic DNA.

The *BRG1*^{SC5314} or *BRG1*^{P57055} containing cassettes and *NAT1* marker cassette were transformed alongside Cas9 and MDR1 sgRNA DNA cassettes into the P57055 *bcr1*Δ/Δ mutant and P57055 *brg1*Δ/Δ mutant strains. Approximately 2 µg of the *BRG1* containing cassette, 2 µg of the *NAT1* marker cassette, 1 µg of Cas9, and 1 µg of MDR1 sgRNA DNA cassette were included in each transformation mix. The MDR1 sgRNA DNA cassette was generated using split-joint PCR with primers “sgRNA/F MDR1-5” and “SNR52/R MDR1-5”. Integration of either *BRG1* allele at the *MDR1* locus was determined using PCR genotyping with the primers “MDR1 check up/F” and “BRG1 check int/R”. Heterozygosity or homozygosity of *BRG1* integration was determined using PCR genotyping with the primers with the primers “MDR1 check up/F” and “MDR1 check int/R”.

Biofilm Growth and Imaging

To assay biofilm formation, strains were inoculated to an OD_{600} of 0.5 from overnight cultures into 2 ml of RPMI + 10% serum containing a 1.5 cm x 1.5 cm silicone square (Bentec Medical Inc., Woodland) in the wells of an untreated 12 well plate. The cells were then incubated in an incubator shaker at 37°C for 90 minutes with mild shaking (60 rpm) to allow for adherence to the silicone square, and following initial adhesion, were washed of non-adherent cells by brief immersion in 2 ml PBS then reintroduced into a new well containing fresh 2 ml of RPMI + 10% serum. Biofilms were then allowed to grow for 24 hours in an incubator shaker at 37°C with mild shaking (60 rpm), before being washed of media and fixed for one hour using a solution of 4% formaldehyde and 2.5% glutaraldehyde in PBS.

Silicone squares from biofilm assays that were not fixed for confocal imaging were soaked in distilled water and agitated to remove the bulk of any adherent biofilm material. Several passes of scrubbing then rinsing in distilled water were then used to remove any remaining adherent material. Silicone squares were then subsequently dried and autoclaved for re-use. To ensure reproducibility, recycled squares were used in all assays in the P57055 background.

Fixed biofilms were stained overnight with Concanavalin A, Alexa Fluor 594 conjugate (Life Technologies) diluted to 25 μg/ml in PBS. Biofilms were then washed once more in PBS to remove any excess dye, then transferred to glass scintillation vials and index matched through subsequent passages through 100% methanol, 50:50 methanol and

methyl salicylate solution, and 100% methyl salicylate. Biofilms were then imaged using a slit-scan confocal optical unit on a Zeiss Axiovert 200 microscope with a Zeiss 40x/0.85 NA oil immersion objective. The index matching and imaging are described in greater detail by Lagree et al. [42].

Hyphal Induction Assays and Imaging

To assay hyphal formation, strains were inoculated to an OD_{600} of 0.5 from overnight cultures into 5 ml of RPMI + 10% serum in glass test tubes. Cells were then grown for 4 hours at 37°C in a roller drum for vigorous agitation. Cells were then collected by centrifugation and fixed with 4% formaldehyde for 15 minutes. Fixed cells were then washed twice in PBS and stained with Calcofluor-white. Stained cells were then imaged using a slit-scan confocal optical unit on a Zeiss Axiovert 200 microscope with a Zeiss C-Apochromat 40x/1.2 NA water immersion objective. Results were then quantified using two metrics: length of hyphal units and ratio of hyphal units to yeast cells. To quantify the length of hyphal units, the distances between septa on hyphae were measured using ImageJ. At least 50 inter-septal distance measurements were taken from 3 separate 112 µm x 83.5 µm fields of view. Hyphal units and yeast cells were then counted using the same fields of view to obtain the ratio of hyphal units to yeast cells.

RNA extraction and Nanostring

For all RNA extractions, strains were inoculated from overnight cultures into 25 ml of RPMI + 10% serum to an OD_{600} of 0.2. Cells were then grown for 4 hours with vigorous shaking (225 rpm) in an incubator shaker then harvested by vacuum filtration and quickly frozen at -80°C until RNA extraction. Three cultures of each strain were grown to provide three biological replicates for Nanostring and RNA-Seq experiments.

RNA extraction and NanoString analysis was performed according to previously published methods [43]. Cell disruption was achieved mechanically using Zirconia beads (Ambion, Fisher Scientific, Waltham), and extraction was performed using a 25:24:1 phenol:chloroform:isoamyl alcohol method combined with a Qiagen RNeasy Mini Kit (Qiagen, Venlo, Netherlands). 25 ng of extracted RNA was added to a nanoString codeset mix and incubated at 65°C for 18 hours, before further binding and washing on a nanoString nCounter Prep Station and scanning on an nCounter digital analyzer. Raw counts were normalized against average total counts with background subtraction. Statistical significance in differential expression was assessed using the Benjamini-Hochberg procedure at a FDR of 0.1.

RNA-Seq

RNA-Seq was performed on the same RNA samples prepared for Nanostring. Five micrograms of total RNA was incubated with 2 units of TurboDNAse (Invitrogen) in a 50 ul reaction for 15 minutes at 37 degrees C. The RNA was purified acid phenol-chloroform extraction, and the supernatant containing the RNA was purified over a

column and eluted into 15 ul of nuclease free water. Two micrograms of total RNA was used as input for the Lexogen mRNA sense kit v2. The kit was used according to the manufacturer's instructions for shorter amplicons. Eleven cycles of PCR were performed, incorporating unique barcode indices on each library. The resulting thirty libraries were pooled evenly and subjected to one lane of Illumina sequencing (Novogene), resulting in an average of 16 million reads per library.

Raw fastq reads were trimmed using cutadapt (v 1.9.1) (DOI: https://doi.org/10.14806/ej.17.1.200), with options "-m 42 -a AGATCGGAAGAGC" to remove Illumina 3' adapter sequence and "-u 10 -u -6" to remove the Lexogen random priming sequences, according to the Lexogen's instructions. Trimmed reads were mapped using tophat (v 2.0.8) [44] with options "–no-novel-juncs" and "-G" to align to the *C. albicans* SC5314 reference genome assembly 22 annotation gff file. Primary alignments were selected using samtools (v 0.1.18) [45] with options "view -h -F 256". Gene counts were created using "coverageBed" from bedtools (v 2.17.0) [46] with option "-S" to count stranded alignments (as Lexogen reads are reverse complement). The SC5314 release 22 is a phased diploid assembly. RNA-Seq reads mapped to the two alleles of each gene were combined for further analysis. Differential expression was assessed using DEseq2 (v 1.22.1) [47] in R (v 3.5.1) using default options (alpha = 0.05).

Software

Images were compiled and any adjustments were performed in ImageJ [48]. Single guide RNA sequences were checked for specificity using Cas-OFFinder software [49]. Network graphs were constructed using Cytoscape software [50]. Analyses were performed with Graphpad Prism version 8.00 (Graphpad Software, Inc., La Jolla). Venn diagrams were constructed using Venn Diagrams software (http://bioinformatics.psb.ugent.be/webtools/Venn/).

ACKNOWLEDGEMENTS

We are grateful to Julia Carter for construction of the *bcr1*Δ/Δ mutants used in this study. We thank Katherine Lagree, Frederick Lanni, Max Cravener, Luisa Hiller, Dannie Durand, and Veronica Hinman for many helpful discussions and for comments on this manuscript.

LITERATURE CITED, CHAPTER 4

1. Albert FW, Kruglyak L (2015) The role of regulatory variation in complex traits and disease. Nat Rev Genet 16: 197-212.
2. Sardi M, Gasch AP (2018) Genetic background effects in quantitative genetics: gene-by-system interactions. Curr Genet 64: 1173-1176.
3. Schacherer J (2016) Beyond the simplicity of Mendelian inheritance. C R Biol 339: 284-288.
4. Dowell RD, Ryan O, Jansen A, Cheung D, Agarwala S, et al. (2010) Genotype to phenotype: a complex problem. Science 328: 469.
5. Ryan O, Shapiro RS, Kurat CF, Mayhew D, Baryshnikova A, et al. (2012) Global gene deletion analysis exploring yeast filamentous growth. Science 337: 1353-1356.
6. Vu V, Verster AJ, Schertzberg M, Chuluunbaatar T, Spensley M, et al. (2015) Natural Variation in Gene Expression Modulates the Severity of Mutant Phenotypes. Cell 162: 391-402.
7. Gasch AP, Payseur BA, Pool JE (2016) The Power of Natural Variation for Model Organism Biology. Trends Genet 32: 147-154.
8. Chin BL, Ryan O, Lewitter F, Boone C, Fink GR (2012) Genetic variation in Saccharomyces cerevisiae: circuit diversification in a signal transduction network. Genetics 192: 1523-1532.
9. Hall RA, Noverr MC (2017) Fungal interactions with the human host: exploring the spectrum of symbiosis. Curr Opin Microbiol 40: 58-64.
10. Mayer FL, Wilson D, Hube B (2013) Candida albicans pathogenicity mechanisms. Virulence 4: 119-128.
11. Sudbery PE (2011) Growth of Candida albicans hyphae. Nat Rev Microbiol 9: 737-748.
12. Desai JV, Mitchell AP, Andes DR (2014) Fungal biofilms, drug resistance, and recurrent infection. Cold Spring Harb Perspect Med 4.
13. Banerjee M, Uppuluri P, Zhao XR, Carlisle PL, Vipulanandan G, et al. (2013) Expression of UME6, a key regulator of Candida albicans hyphal development, enhances biofilm formation via Hgc1- and Sun41-dependent mechanisms. Eukaryot Cell 12: 224-232.
14. Cleary IA, Lazzell AL, Monteagudo C, Thomas DP, Saville SP (2012) BRG1 and NRG1 form a novel feedback circuit regulating Candida albicans hypha formation and virulence. Mol Microbiol 85: 557-573.
15. Du H, Guan G, Xie J, Sun Y, Tong Y, et al. (2012) Roles of Candida albicans Gat2, a GATA-type zinc finger transcription factor, in biofilm formation, filamentous growth and virulence. PLoS One 7: e29707.
16. Nobile CJ, Fox EP, Nett JE, Sorrells TR, Mitrovich QM, et al. (2012) A recently evolved transcriptional network controls biofilm development in Candida albicans. Cell 148: 126-138.
17. Nobile CJ, Mitchell AP (2005) Regulation of cell-surface genes and biofilm formation by the C. albicans transcription factor Bcr1p. Curr Biol 15: 1150-1155.

18. Ramage G, VandeWalle K, Lopez-Ribot JL, Wickes BL (2002) The filamentation pathway controlled by the Efg1 regulator protein is required for normal biofilm formation and development in Candida albicans. FEMS Microbiol Lett 214: 95-100.
19. Lohse MB, Gulati M, Johnson AD, Nobile CJ (2018) Development and regulation of single- and multi-species Candida albicans biofilms. Nat Rev Microbiol 16: 19-31.
20. Wu W, Lockhart SR, Pujol C, Srikantha T, Soll DR (2007) Heterozygosity of genes on the sex chromosome regulates Candida albicans virulence. Mol Microbiol 64: 1587-1604.
21. Hirakawa MP, Martinez DA, Sakthikumar S, Anderson MZ, Berlin A, et al. (2015) Genetic and phenotypic intra-species variation in Candida albicans. Genome Res 25: 413-425.
22. Pannanusorn S, Ramirez-Zavala B, Lunsdorf H, Agerberth B, Morschhauser J, et al. (2014) Characterization of biofilm formation and the role of BCR1 in clinical isolates of Candida parapsilosis. Eukaryot Cell 13: 438-451.
23. Lassak T, Schneider E, Bussmann M, Kurtz D, Manak JR, et al. (2011) Target specificity of the Candida albicans Efg1 regulator. Mol Microbiol 82: 602-618.
24. Doedt T, Krishnamurthy S, Bockmuhl DP, Tebarth B, Stempel C, et al. (2004) APSES proteins regulate morphogenesis and metabolism in Candida albicans. Mol Biol Cell 15: 3167-3180.
25. Norice CT, Smith FJ, Jr., Solis N, Filler SG, Mitchell AP (2007) Requirement for Candida albicans Sun41 in biofilm formation and virulence. Eukaryot Cell 6: 2046-2055.
26. Hiller E, Heine S, Brunner H, Rupp S (2007) Candida albicans Sun41p, a putative glycosidase, is involved in morphogenesis, cell wall biogenesis, and biofilm formation. Eukaryot Cell 6: 2056-2065.
27. Li X, Yan Z, Xu J (2003) Quantitative variation of biofilms among strains in natural populations of Candida albicans. Microbiology 149: 353-362.
28. MacCallum DM, Castillo L, Nather K, Munro CA, Brown AJ, et al. (2009) Property differences among the four major Candida albicans strain clades. Eukaryot Cell 8: 373-387.
29. Ropars J, Maufrais C, Diogo D, Marcet-Houben M, Perin A, et al. (2018) Gene flow contributes to diversification of the major fungal pathogen Candida albicans. Nat Commun 9: 2253.
30. Harcus D, Nantel A, Marcil A, Rigby T, Whiteway M (2004) Transcription profiling of cyclic AMP signaling in Candida albicans. Mol Biol Cell 15: 4490-4499.
31. Zordan RE, Miller MG, Galgoczy DJ, Tuch BB, Johnson AD (2007) Interlocking transcriptional feedback loops control white-opaque switching in Candida albicans. PLoS Biol 5: e256.
32. Sohn K, Urban C, Brunner H, Rupp S (2003) EFG1 is a major regulator of cell wall dynamics in Candida albicans as revealed by DNA microarrays. Mol Microbiol 47: 89-102.
33. Martchenko M, Levitin A, Hogues H, Nantel A, Whiteway M (2007) Transcriptional rewiring of fungal galactose-metabolism circuitry. Curr Biol 17: 1007-1013.
34. Bastidas RJ, Heitman J, Cardenas ME (2009) The protein kinase Tor1 regulates adhesin gene expression in Candida albicans. PLoS Pathog 5: e1000294.

35. Huang MY, Mitchell AP (2017) Marker Recycling in Candida albicans through CRISPR-Cas9-Induced Marker Excision. mSphere 2: 00050-00017.
36. Botstein D, Falco SC, Stewart SE, Brennan M, Scherer S, et al. (1979) Sterile host yeasts (SHY): a eukaryotic system of biological containment for recombinant DNA experiments. Gene 8: 17-24.
37. Min K, Ichikawa Y, Woolford CA, Mitchell AP (2016) Candida albicans Gene Deletion with a Transient CRISPR-Cas9 System. mSphere 1.
38. Jones EW, Berget PB, Burnette JM, 3rd, Anderson C, Asafu-Adjei D, et al. (2008) The spectrum of Trp- mutants isolated as 5-fluoroanthranilate-resistant clones in Saccharomyces bayanus, S. mikatae and S. paradoxus. Yeast 25: 41-46.
39. Vyas VK, Barrasa MI, Fink GR (2015) A Candida albicans CRISPR system permits genetic engineering of essential genes and gene families. Sci Adv 1: e1500248.
40. Nobile CJ, Solis N, Myers CL, Fay AJ, Deneault JS, et al. (2008) Candida albicans transcription factor Rim101 mediates pathogenic interactions through cell wall functions. Cell Microbiol 10: 2180-2196.
41. Huang MY, Woolford CA, Mitchell AP (2018) Rapid Gene Concatenation for Genetic Rescue of Multigene Mutants in Candida albicans. mSphere 3: 00169-00118.
42. Lagree K, Desai JV, Finkel JS, Lanni F (2018) Microscopy of fungal biofilms. Curr Opin Microbiol 43: 100-107.
43. Woolford CA, Lagree K, Xu W, Aleynikov T, Adhikari H, et al. (2016) Bypass of Candida albicans Filamentation/Biofilm Regulators through Diminished Expression of Protein Kinase Cak1. PLoS Genet 12: e1006487.
44. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25: 1105-1111.
45. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078-2079.
46. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841-842.
47. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15: 550.
48. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, et al. (2012) Fiji: an open-source platform for biological-image analysis. Nat Methods 9: 676-682.
49. Bae S, Park J, Kim JS (2014) Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics 30: 1473-1475.
50. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13: 2498-2504.

FIGURE LEGENDS, CHAPTER 4

Figure 1. Biofilm side-view projections. Wild-type and mutant strains in each clinical isolate background were assayed for biofilm formation under *in vitro* conditions. All strains were grown on silicone squares in RPMI + 10% serum at 37°C for 24 hours. Fixed biofilms were stained using Concanavalin A, Alexa Fluor 594 conjugate, then imaged by confocal microscopy. Representative sections from each biofilm are shown; relevant genotypes are given beneath each column. Scale bars indicate the depth of the corresponding wild-type biofilm. Strain backgrounds: **A.** SC5314. **B.** P76067. **C.** P57055. **D.** P87. **E.** P75010.

Figure 2. Biofilm apical-view projections. Apical views of representative sections from each clinical isolate and mutant biofilm are shown. Relevant genotypes are given beneath each column. White scale bars in each panel are 20μm in length. Projections were generated with the same datasets used in Figure 1. Strain backgrounds: **A.** SC5314. **B.** P76067. **C.** P57055. **D.** P87. **E.** P75010.

Figure 3. Filamentation assays. Wild-type and mutant strains of each background were assayed for filamentation under planktonic growth conditions. Strains were grown in RPMI + 10% serum at 37°C for 4 hours with shaking. Fixed cells were stained with Calcofluor-white for confocal microscopy. White scale bars in each panel are 20μm in length. Strain backgrounds: **A.** SC5314. **B.** P76067. **C.** P57055. **D.** P87. **E.** P75010.

Figure 4. Variation in the *C. albicans* biofilm/hyphal regulatory network. Network diagrams are presented for each clinical isolate as well as for features shared among them ("Common"). Nodes represent genes analyzed by Nanostring, with white denoting the four TF genes, and blue and teal denoting prospective target genes. Node positions are identical across network graphs. The teal color indicates that the gene is annotated for function in biofilm or hyphal formation. A significant gene expression alteration by a TF gene mutation is denoted by an edge between two nodes; a dot on an edge indicates the connected TF was reported to bind in the upstream region of the target gene [16,23]. A significant gene expression alteration was defined as a two-fold difference in mRNA level difference between mutant and wild type, and a significant difference in mean mRNA Nanostring counts between mutant and wild type (Benjamini-Hochberg step-up procedure, FDR = 0.1). Three biological replicates were analyzed using Nanostring for all strains.

Figure 5. Range of TF mutant gene expression impact. Fold-change values are plotted for RNAs from the *CHT2*, *SOD5, BRG1*, and *UME6* genes in each TF mutant in all five strain backgrounds. Three biological replicates were analyzed using Nanostring for all strains. Data are extracted from S1 Table.

Figure 6. Genome-wide Efg1 regulons. Global expression was assayed using RNA-Seq. Three biological replicates were analyzed for each *efg1*Δ/Δ mutant and clinical isolate. Fold change values were determined using DeSeq2. **A.** Heatmap depicting log2 fold change in gene expression. Upper (Yellow) and lower bounds (Blue) correspond to

a log2 fold change value of 2 and -2 respectively. Sections labeled glycolysis and biofilm are enriched for genes annotated for roles in glycolysis and biofilm formation respectively. **B.** Venn diagrams depicting intersection of genes dependent upon *EFG1* in each clinical isolate background. We considered all genes that were significantly differentially expressed ($p<0.05$, Benjamini-Hochberg adjustment), and had at least a 2 fold difference in expression between *efg1*Δ/Δ mutant and wild type. **C.** Heatmap depicting p-values from GO term analysis of sets of genes that had significantly lower expression in the *efg1*Δ/Δ mutant vs matched wild type**.** The analyzed sets were the set of genes dependent upon SC5134, P76067, P57055, P87, and P75010, the set of genes common to all 5 clinical isolates (SC5134 ∩ P76067 ∩ P57055 ∩ P87 ∩ P75010), and the set of genes common to all clinical isolates except P76067 (SC5134 ∩ P57055 ∩ P87 ∩ P75010 - P76067). Upper (white) and lower bounds (dark blue) corresponding to a log10 P-value of 0 and -8 respectively.

Figure 7. Impact of constitutive *BRG1* expression in a Bcr1-dependent strain background. Parental *BRG1/BRG1* strains and derived *BRG1/TDH3-BRG1* were assayed for planktonic hyphal formation and biofilm production in RPMI + 10% serum at 37°C. Planktonic cultures were grown for 4 hours, and biofilm cultures were grown for 24 hours. **A.** Fold change in expression of *BRG1* mRNA analyzed by Nanostring. Values shown are mean (SD). Significance is indicated above horizontal bars (Tukey-Kramer test; "****", $P < 0.0001$). Three biological replicates were analyzed. **B.** Hyphal length and hypha to yeast ratios were quantified in planktonic culture samples. Values shown are mean (SD). Three technical replicates were performed for each strain. Pairs

of means connected by a horizontal bar are significantly different (Tukey-Kramer test; “*”, P < 0.05; “**”, P < 0.01; “****”, P<0.0001); all unconnected pairs are not significantly different. **C.** Side-view projections of biofilms stained with ConA-Alexafluor 594 conjugate. Scale bar on left indicates depth of wild-type biofilm. **D.** Images of planktonic culture samples stained with Calcofluor-white. White scale bars in each panel are 20μm in length.

SUPPLEMENTARY FIGURES

S1 Figure. Depth of biofilms formed by *C. albicans* clinical isolates.

Biofilm depth was quantified from the indicated clinical isolate strains on silicone squares in RPMI + 10% serum at 37°C for 24 hours. Three biological replicates were analyzed for each clinical isolate. Measurements were taken from several positions on each biofilm by confocal microscopy. Values shown are mean depth (SD). Comparisons between isolate biofilm depths were significant (Tukey-Kramer test, P<0.05) except those indicated by a horizontal bar.

S2 Figure. Quantification of clinical isolate filamentation

Filamentation capacities of clinical isolate wild-type strains were quantified following hyphal induction. Three technical replicates were performed for each strain. A. Boxplots of the distribution of hyphal unit lengths measured from the indicated clinical isolate background. Whiskers are 1.5IQR. Significant differences in mean hyphal unit length between isolates are indicated (Tukey-Kramer test, ***, P<0.001; ****, P<0.0001). B.

Ratio of observed hyphal units to yeast cells in the indicated clinical isolate background. Values are mean (SD). Significant differences in mean hyphal unit : yeast cell ratios are indicated (Tukey-Kramer test, *, P<0.05; **, P<0.01; ****, P<0.0001).

S3 Figure. Quantification of clinical isolate TF mutant filamentation

Filamentation capacities of clinical isolate TF mutant strains were quantified following hyphal induction. Three technical replicates were performed for each strain. Top Panel: Boxplots of the distribution of hyphal unit lengths measured from the indicated mutant and clinical isolate background. Whiskers are 1.5IQR. Significance of the difference in mean hyphal unit length between mutant and wild type of the same clinical isolate background is indicated above each value (Dunnett test; *, P<0.05; **, P<0.01; n.s., not significant). Strains in which hyphae were not detected are marked n/a. Bottom Panel: Ratio of observed hyphal units to yeast cells in the indicated mutant and clinical isolate background. Values are mean (SD). Significance of the differences in mean hyphal unit : yeast cell ratios between mutant and wild type of the same clinical isolate background is indicated above each value (Dunnett test; *, P<0.05; **, P<0.01; n.s., not significant). Strains in which hyphae were not detected are marked n/a.

S4 Figure. Hyphal formation by reconstituted strains.

To validate TF mutant strain filamentation phenotypes, *BCR1, UME6, BRG1, and EFG1* alleles from SC5314 were reconstituted in the corresponding transcription factor mutants in all clinical isolates using our concatemer assembly method [41]. The resultant validation strains were grown in RPMI + 10% serum at 37°C for 4 hours with

shaking alongside wild-type and *efg1*Δ/Δ mutant strains in the corresponding clinical isolate backgrounds. Fixed cells were stained with Calcofluor-white and imaged using confocal microscopy. For *efg1*Δ/Δ and *brg1*Δ/Δ mutant strains, filamentation in homozygous validation strains is shown. For *ume6*Δ/Δ mutant validation strains, filamentation in heterozygous validation strains are depicted. For *bcr1*Δ/Δ mutant strains, filamentation in heterozygous validation strains are depicted, except for P75010 in which only homozygous transformants were recovered. Images for *bcr1*Δ/Δ, *brg1*Δ/Δ, and *ume6*Δ/Δ mutant strains are taken from Figure 3 for visual reference. White scale bars in each panel are 20 μm in length.

S5 Figure. Quantification of filamentation in P75010 expressing EFG1 or BRG1 alleles from SC5314

P75010 wild-type and P75010 background strains expressing *EFG1* or *BRG1* alleles from SC5314 were quantified following hyphal induction. Three technical replicates were performed for each strain. Left Panel: Boxplots of the distribution of hyphal unit lengths. Whiskers are 1.5IQR. Significance of the difference in mean hyphal unit length is indicated for each background (Dunnett test; ns, not significant; *, $P < 0.05$). Bottom Panel: Ratio of observed hyphal units to yeast cells. Values are mean (SD). Significance of the differences in mean hyphal unit : yeast cell ratios are indicated for each background (Bonferroni test; **, $P < 0.01$; ****, $P < 0.0001$).

S6 Figure. Heatmap of glycolytic process gene expression.

Heatmap depicts log2 fold change in expression of genes with “glycolytic process” GO annotation. Sample and gene orders reflect hierarchical clustering of gene expression data, with average linkage clustering based on Manhattan distance. Upper (Yellow) and lower bounds (Blue) correspond to a log2 fold change value of 2 and -2 respectively.

S7 Figure. Hyphal formation of *BRG1* ectopic expression strains.

Wild-type and *BRG1* ectopic expression strains in the P57055 background were assayed for filamentation under planktonic growth conditions. Strains were grown in RPMI + 10% serum at 37°C for 4 hours with shaking. Fixed cells were stained with Calcofluor-white for confocal microscopy. White scale bars in each panel are 20μm in length. **A.** Filamentation in wild-type, *brg1*Δ/Δ, *brg1*Δ/Δ *mdr1*Δ::*BRG1*^{SC5314}/*mdr1*Δ::*BRG1*^{SC5314}, and *brg1*Δ/Δ *mdr1*Δ::*BRG1*^{P57055}/*mdr1*Δ::*BRG1*^{P57055} strains. **B.** Filamentation in *bcr1*Δ/Δ and *bcr1*Δ/Δ strains expressing *BRG1*^{SC5314} or *BRG1*^{P57055}. Strains carrying one (heterozygous expression) or two (homozygous expression) copies of *BRG1* alleles from either background were assayed. Two independent isolates are depicted for each case.

FIGURES, CHAPTER 4

Figure 1

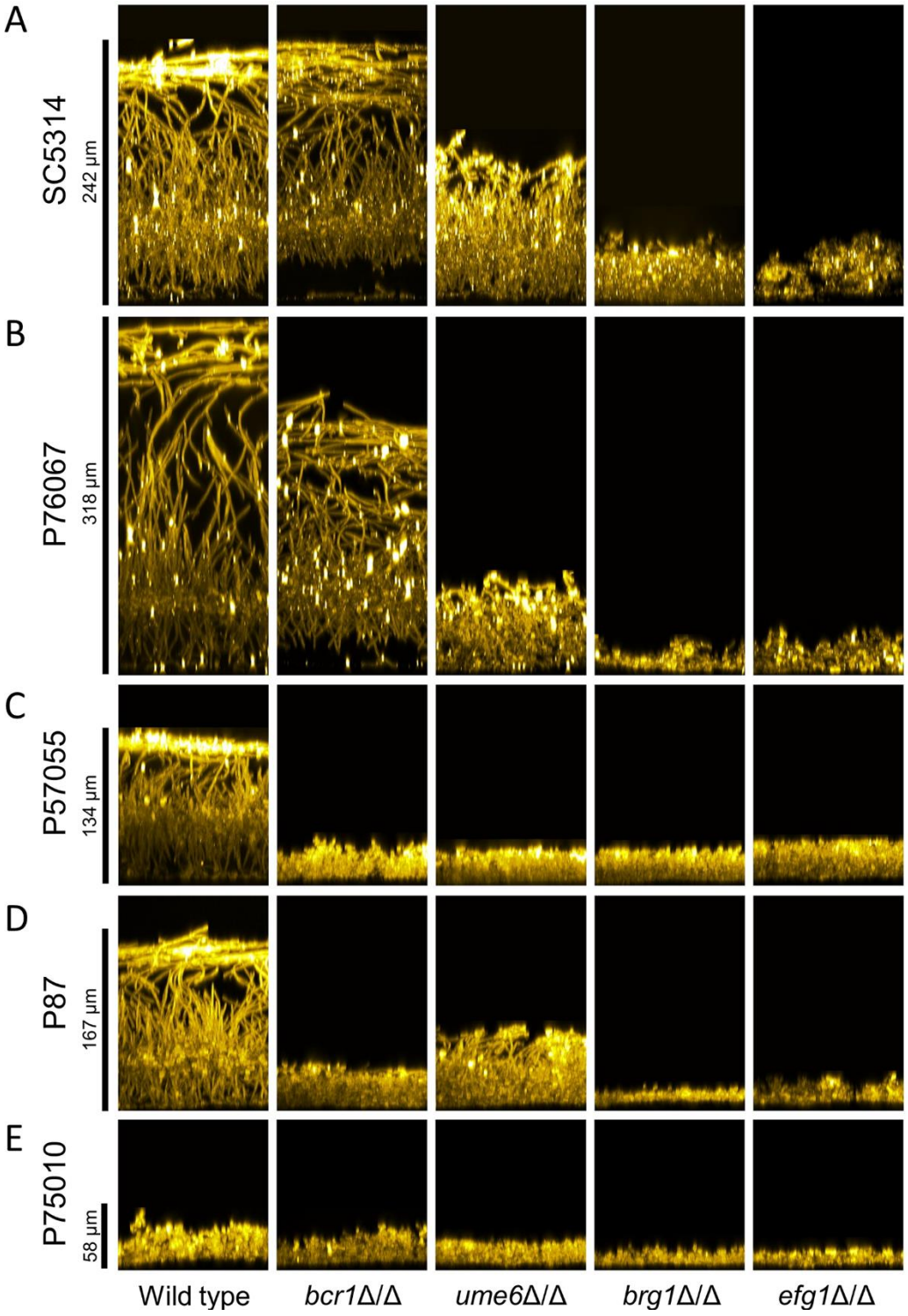


Figure 2

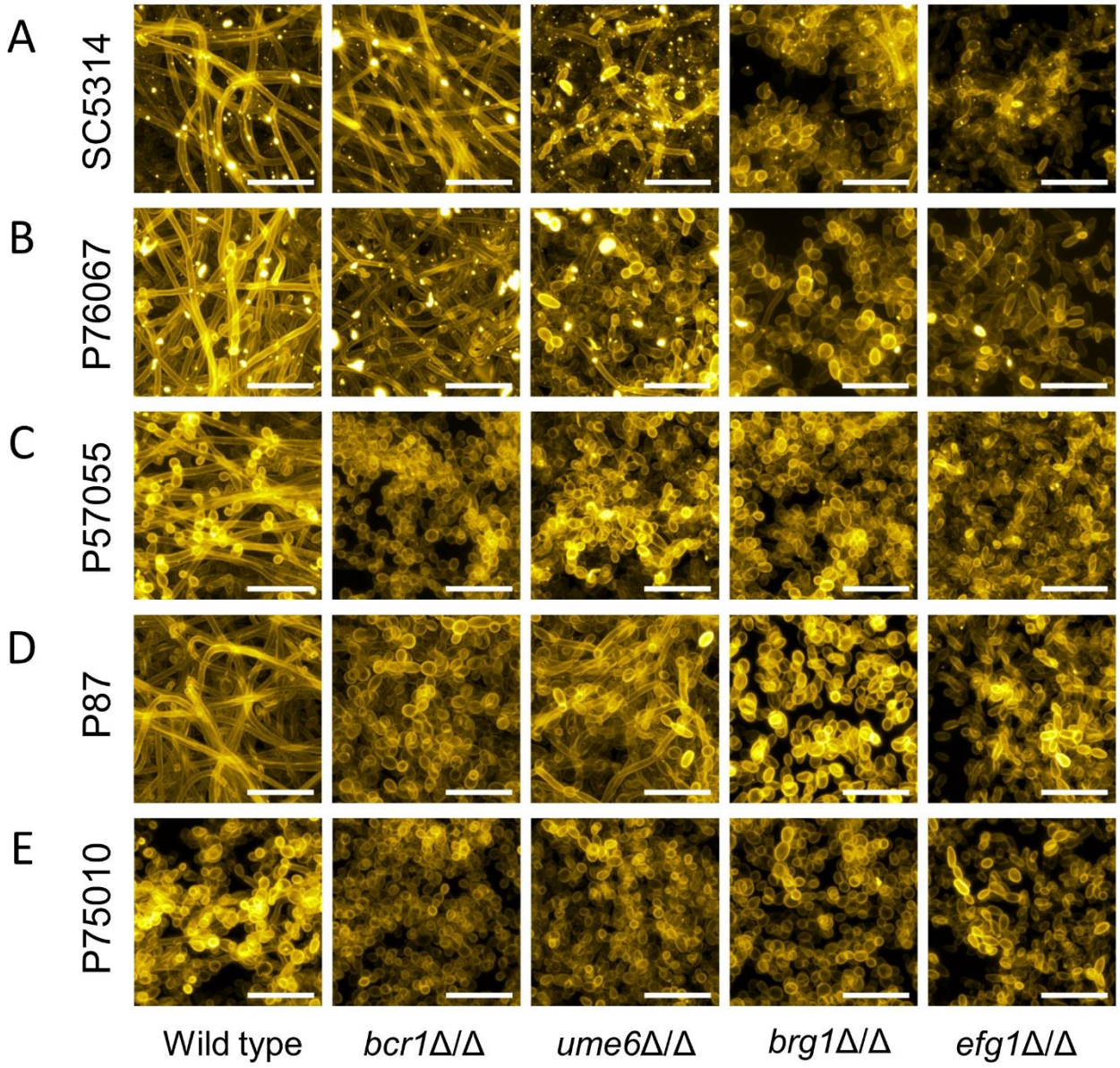

A
SC5314
B
P76067
C
P57055
D
P87
E
P75010
Wild type
bcr1Δ/Δ
ume6Δ/Δ
brg1Δ/Δ
efg1Δ/Δ

Figure 3

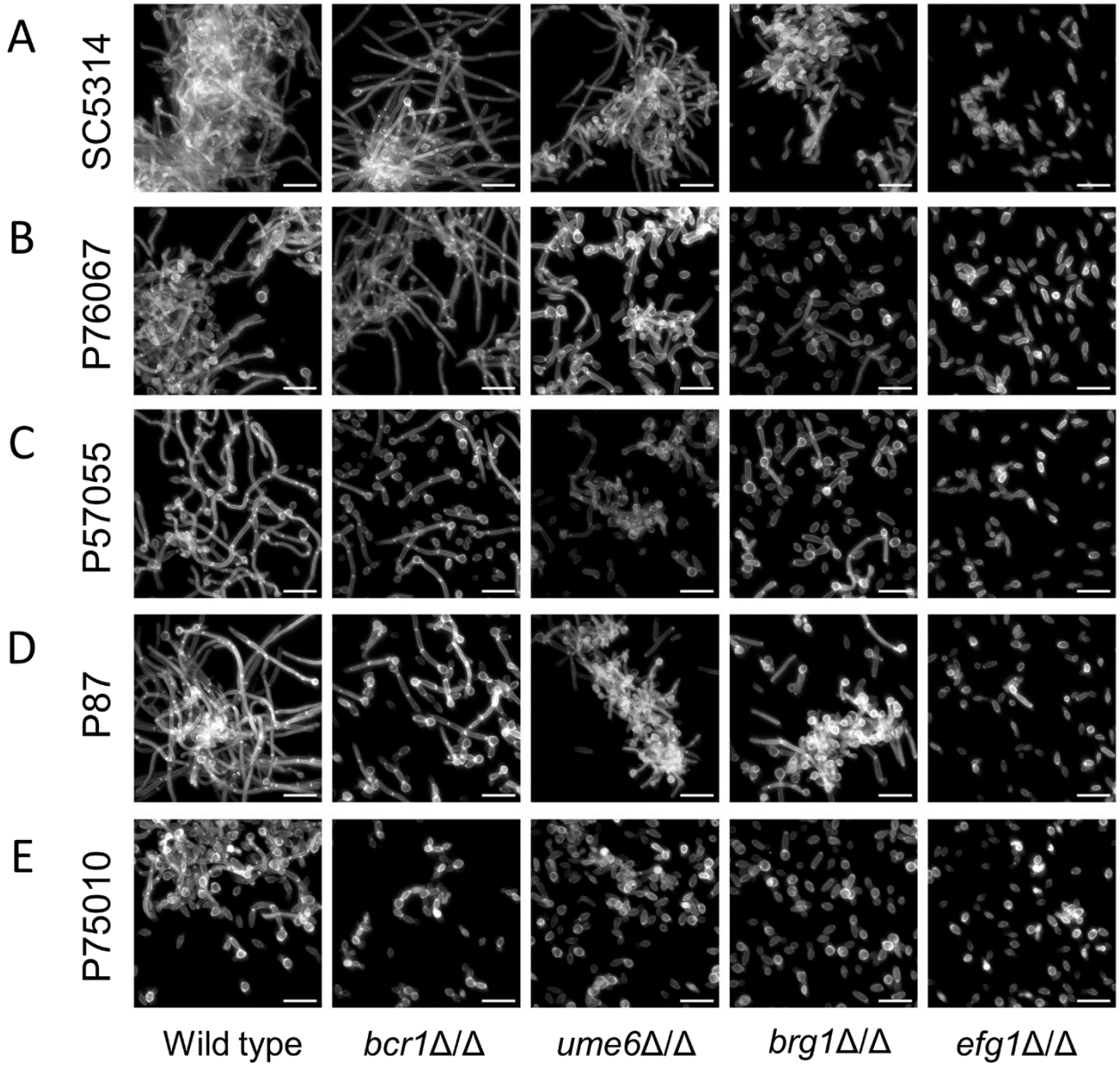

A
SC5314
B
P76067
C
P57055
D
P87
E
P75010
Wild type
bcr1Δ/Δ
ume6Δ/Δ
brg1Δ/Δ
efg1Δ/Δ

Figure 4

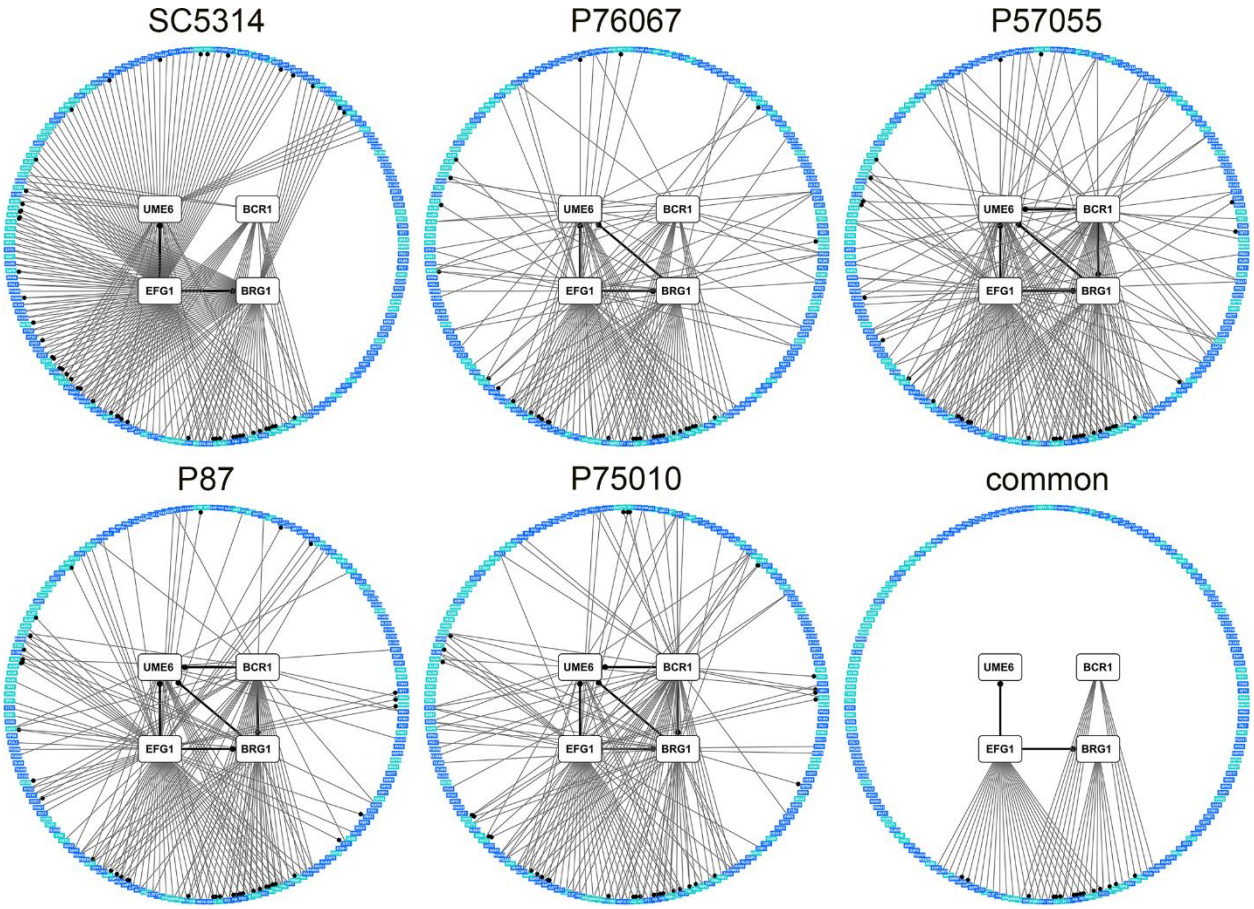


Figure 5

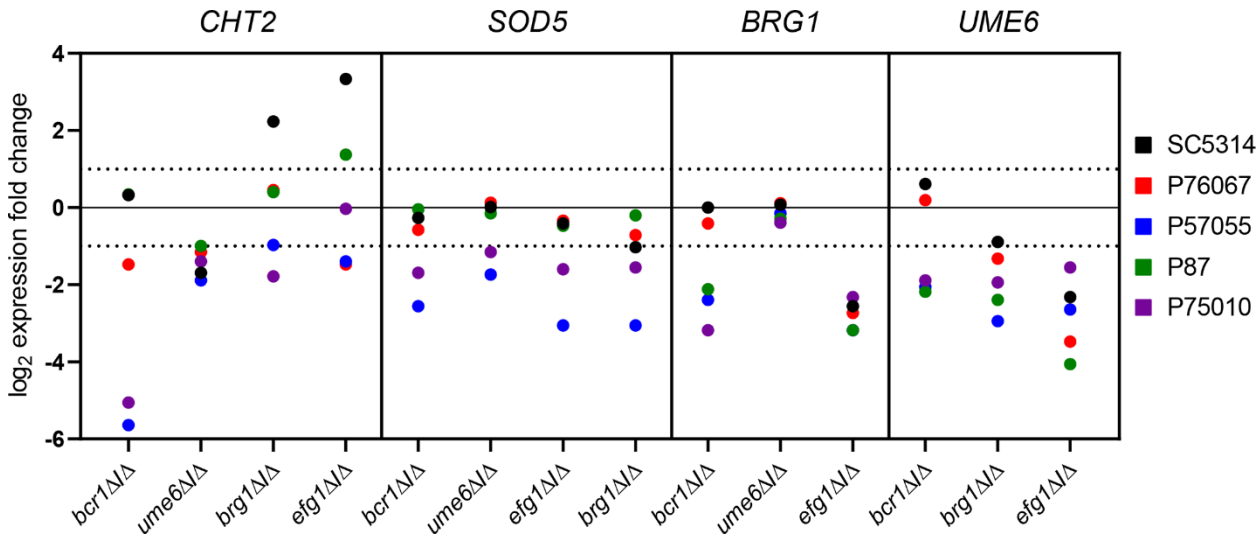


Figure 6

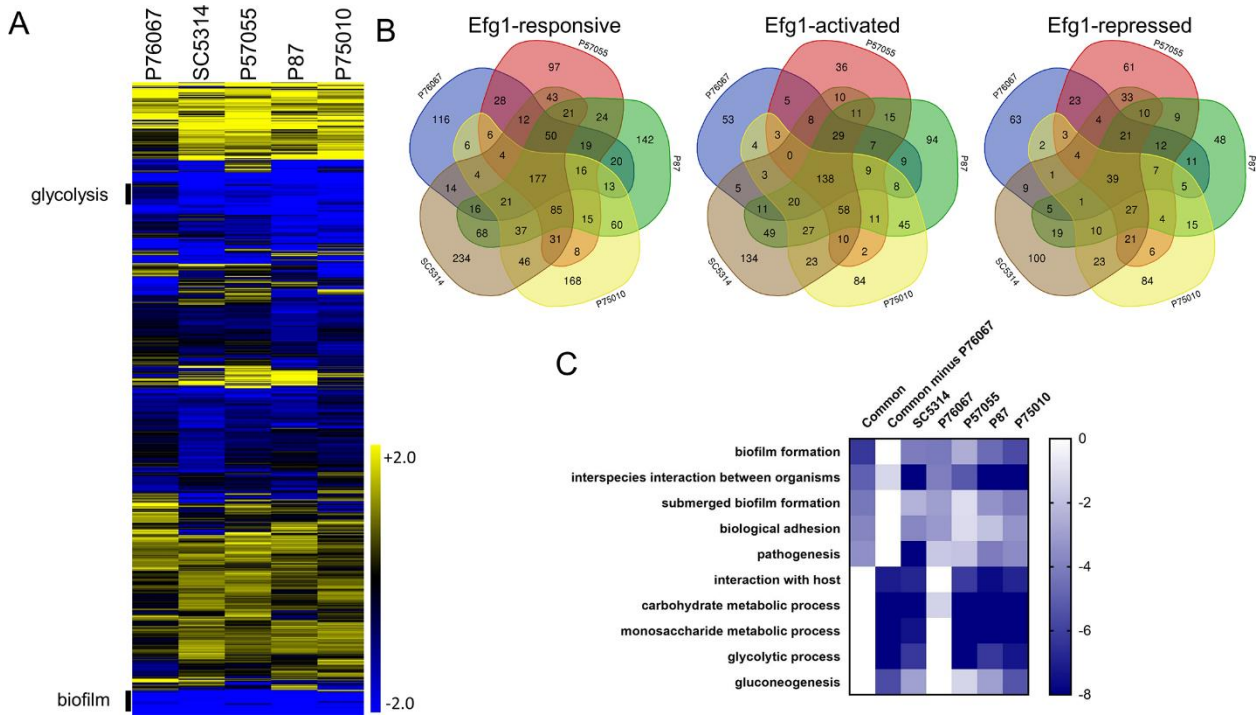
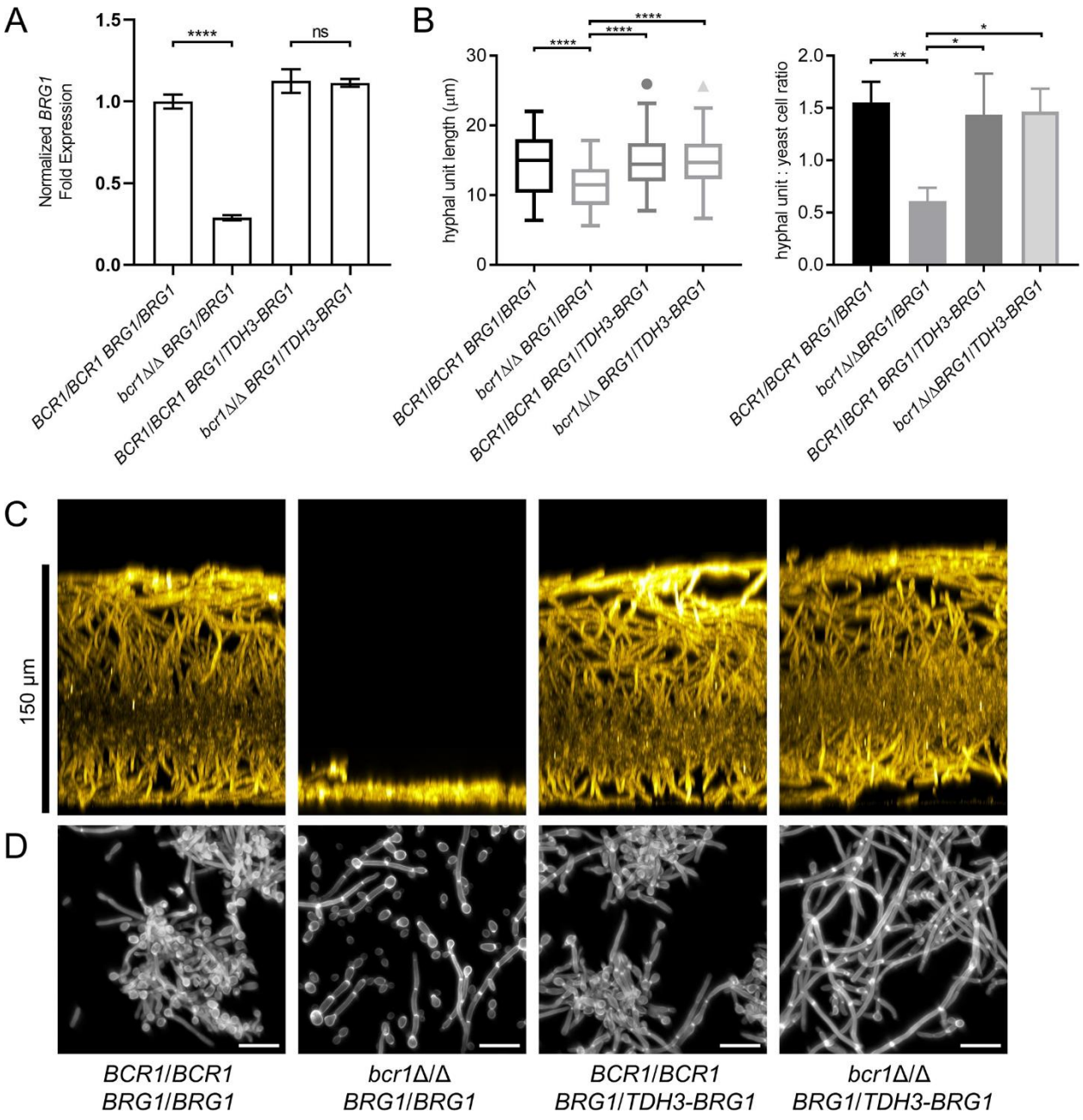
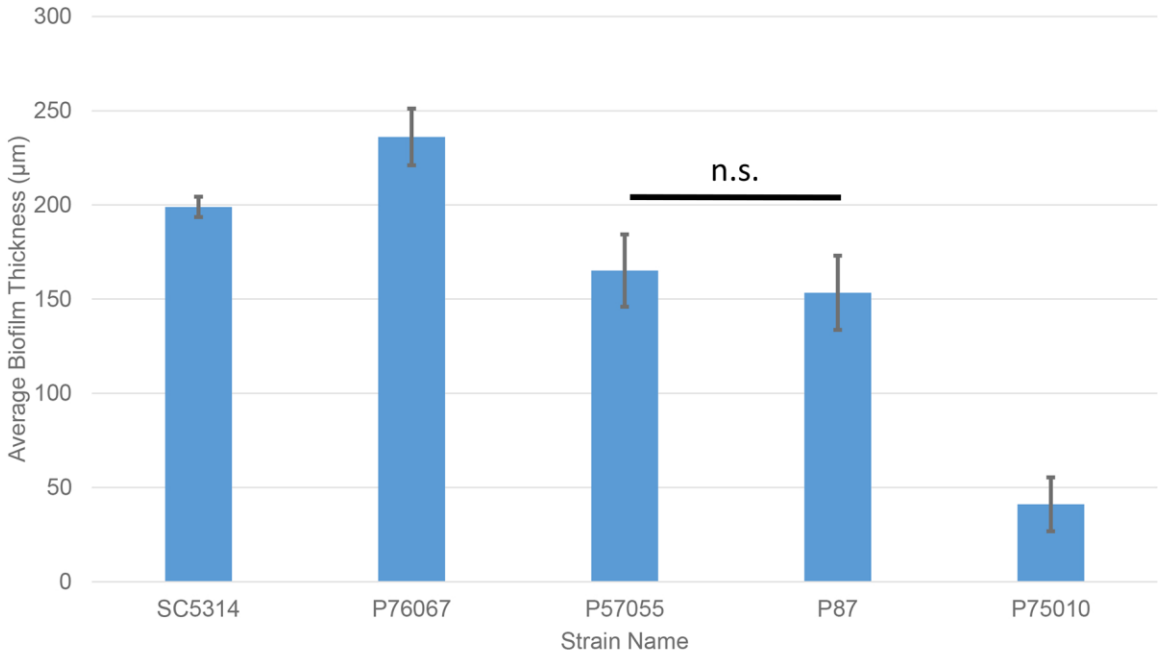


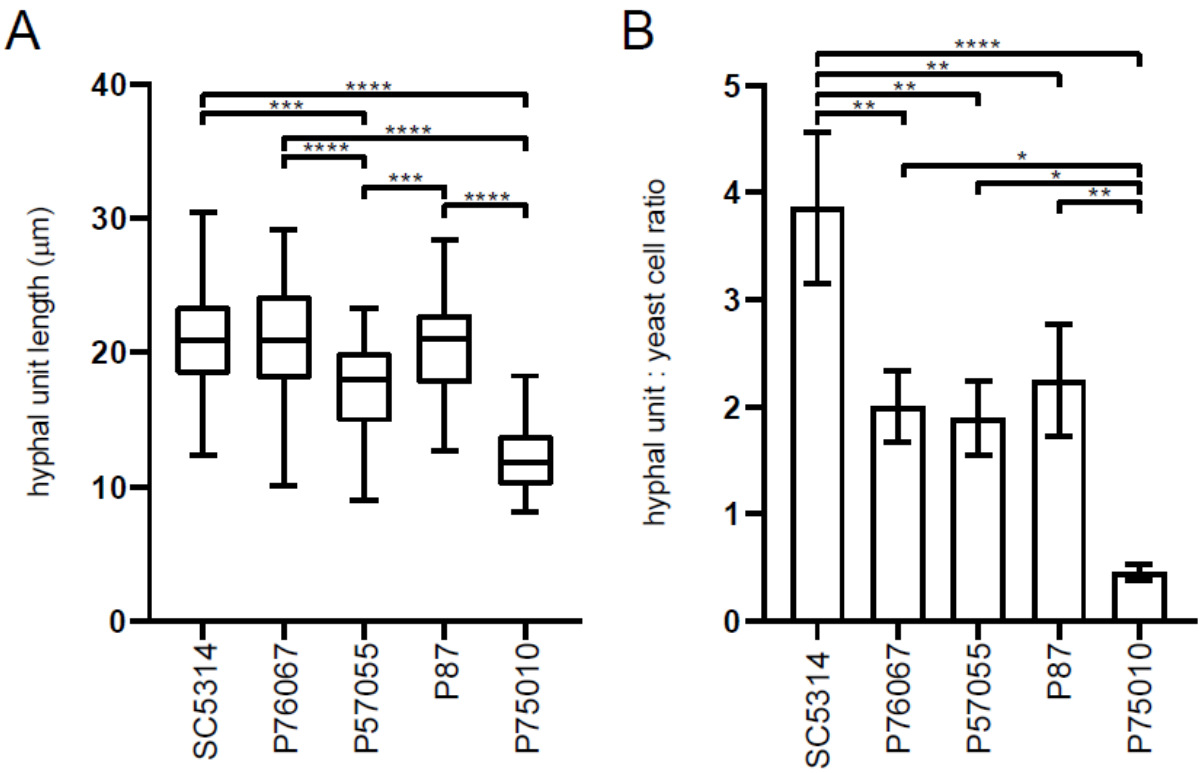
Figure 7


A
Normalized BRG1 Fold Expression
1.5
1.0
0.5
0.0
****
ns
BCR1/BCR1 BRG1/BRG1
bcr1Δ/Δ BRG1/BRG1
BCR1/BCR1 BRG1/TDH3-BRG1
bcr1Δ/Δ BRG1/TDH3-BRG1
B
hyphal unit length (μm)
30
20
10
0
****
****
****
BCR1/BCR1 BRG1/BRG1
bcr1Δ/Δ BRG1/BRG1
BCR1/BCR1 BRG1/TDH3-BRG1
bcr1Δ/Δ BRG1/TDH3-BRG1
hyphal unit : yeast cell ratio
2.0
1.5
1.0
0.5
0.0
**
*
*
BCR1/BCR1 BRG1/BRG1
bcr1Δ/ΔBRG1/BRG1
BCR1/BCR1 BRG1/TDH3-BRG1
bcr1Δ/ΔBRG1/TDH3-BRG1
C
150 μm
D
BCR1/BCR1
BRG1/BRG1
bcr1Δ/Δ
BRG1/BRG1
BCR1/BCR1
BRG1/TDH3-BRG1
bcr1Δ/Δ
BRG1/TDH3-BRG1

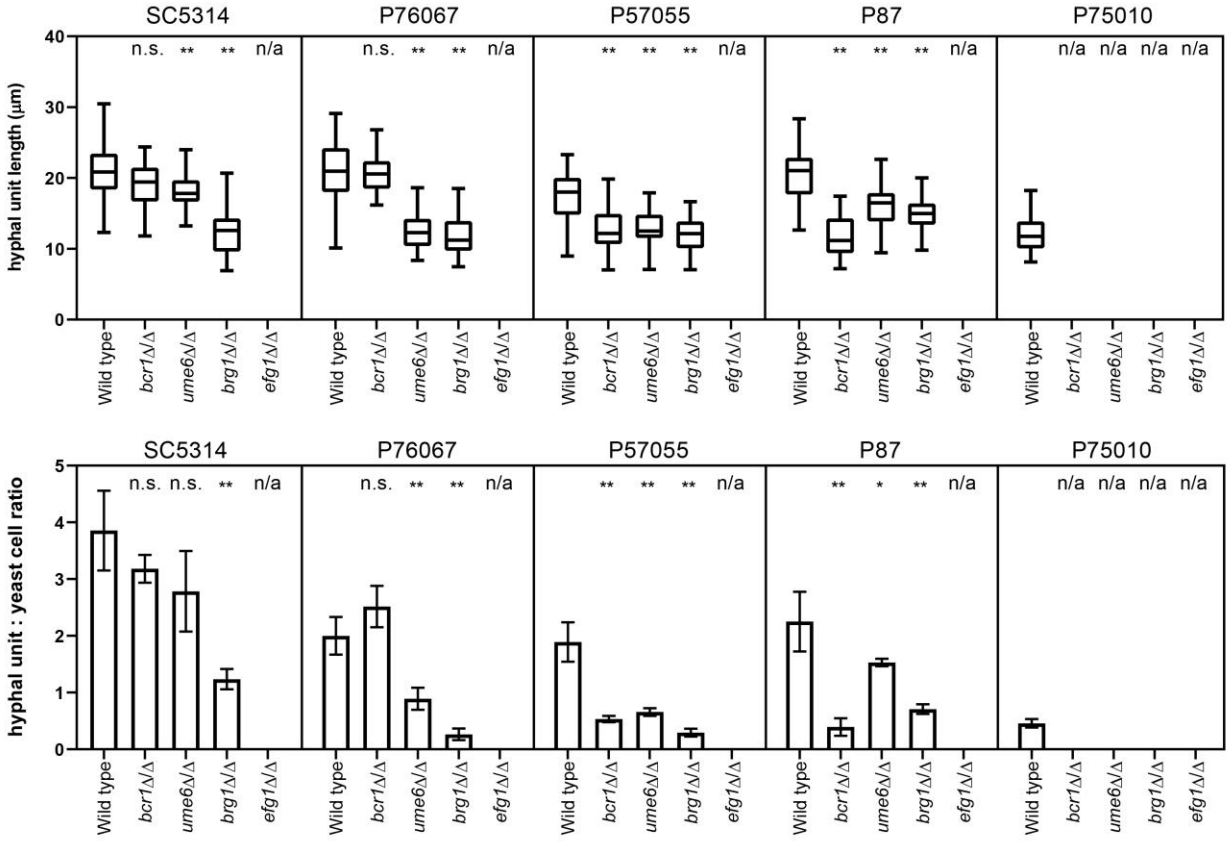
S1 Figure


300
250
200
150
100
50
0
Average Biofilm Thickness (µm)
n.s.
SC5314
P76067
P57055
P87
P75010
Strain Name

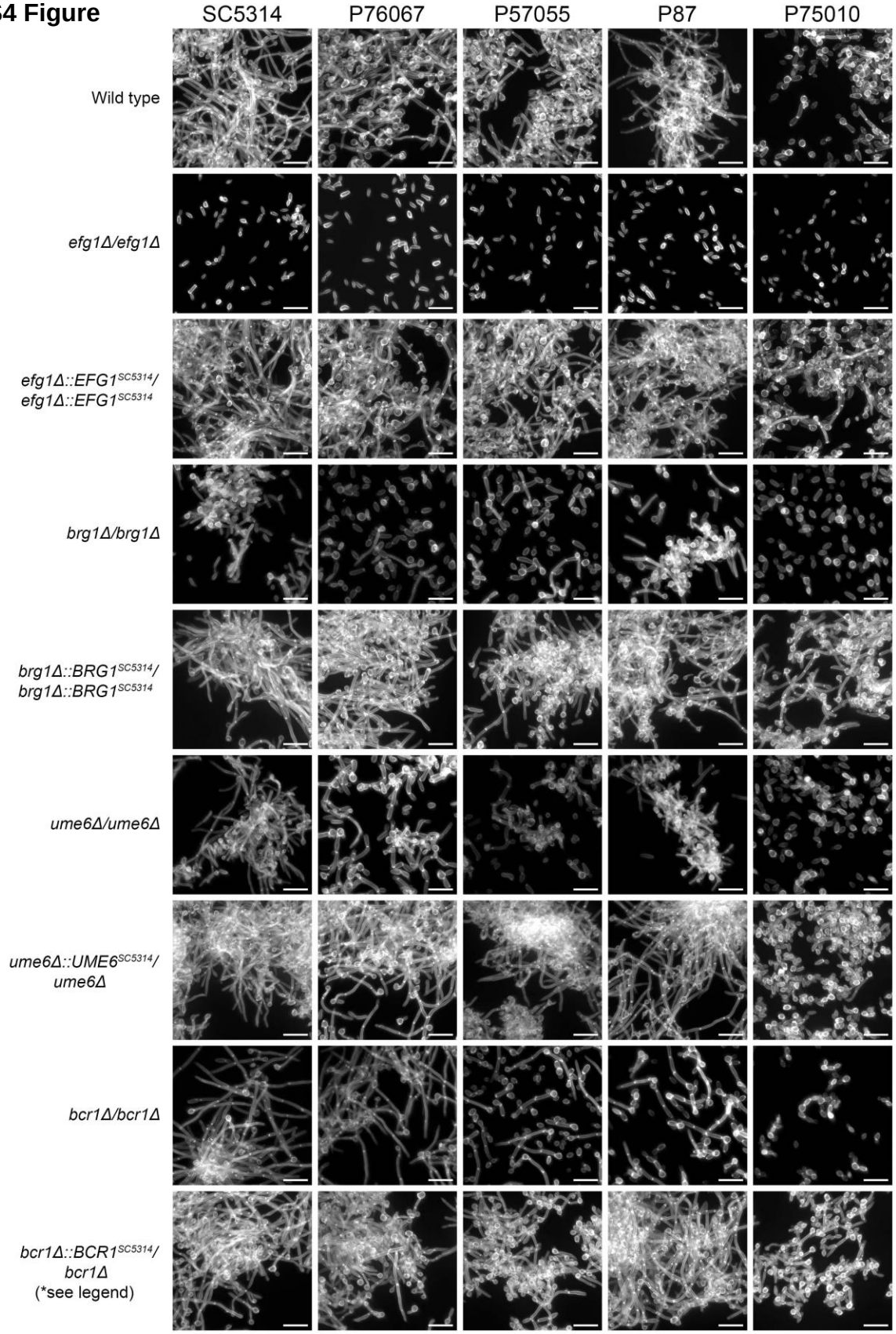
S2 Figure


A
hyphal unit length (µm)
0
10
20
30
40
****
***
****
****
***
****
SC5314
P76067
P57055
P87
P75010
B
hyphal unit : yeast cell ratio
0
1
2
3
4
5
****
**
**
**
*
*
**
SC5314
P76067
P57055
P87
P75010

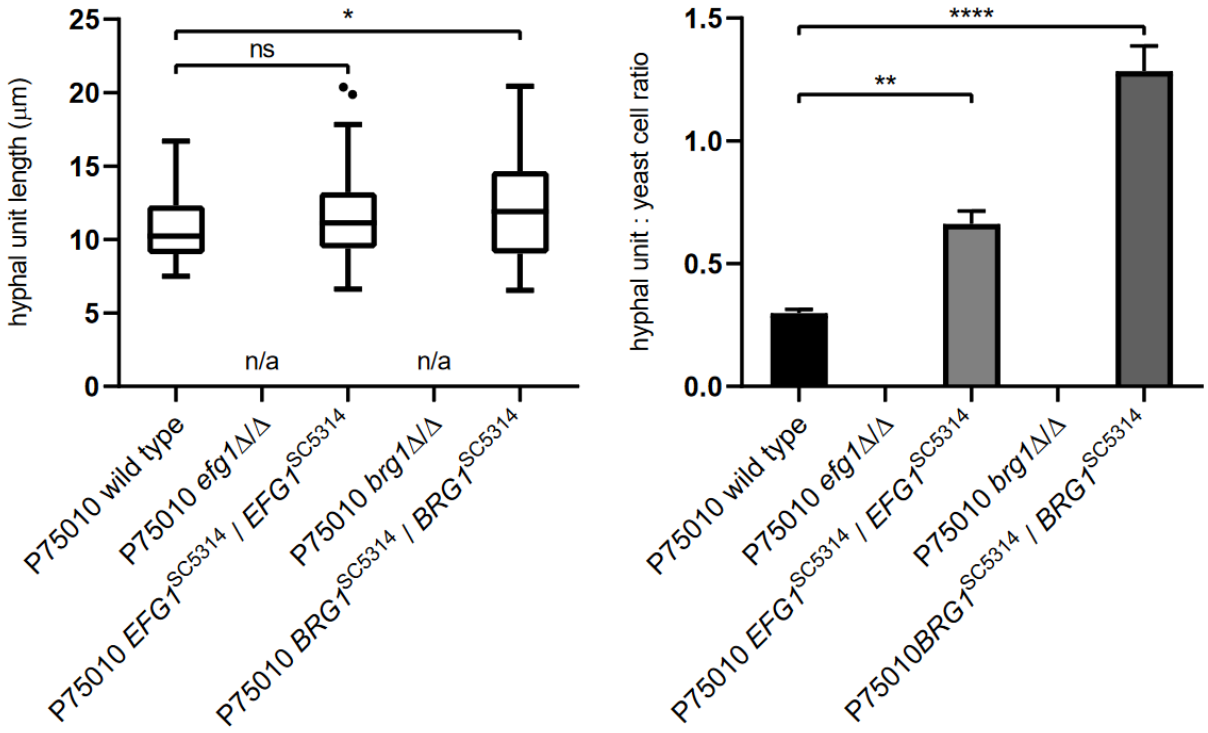
S3 Figure


SC5314
P76067
P57055
P87
P75010
hyphal unit length (μm)
n.s. ** ** n/a
n.s. ** ** n/a
** ** ** n/a
** ** ** n/a
n/a n/a n/a n/a
Wild type
bcr1Δ/Δ
ume6Δ/Δ
brg1Δ/Δ
efg1Δ/Δ
hyphal unit : yeast cell ratio
n.s. n.s. ** n/a
n.s. ** ** n/a
** ** ** n/a
** * ** n/a
n/a n/a n/a n/a

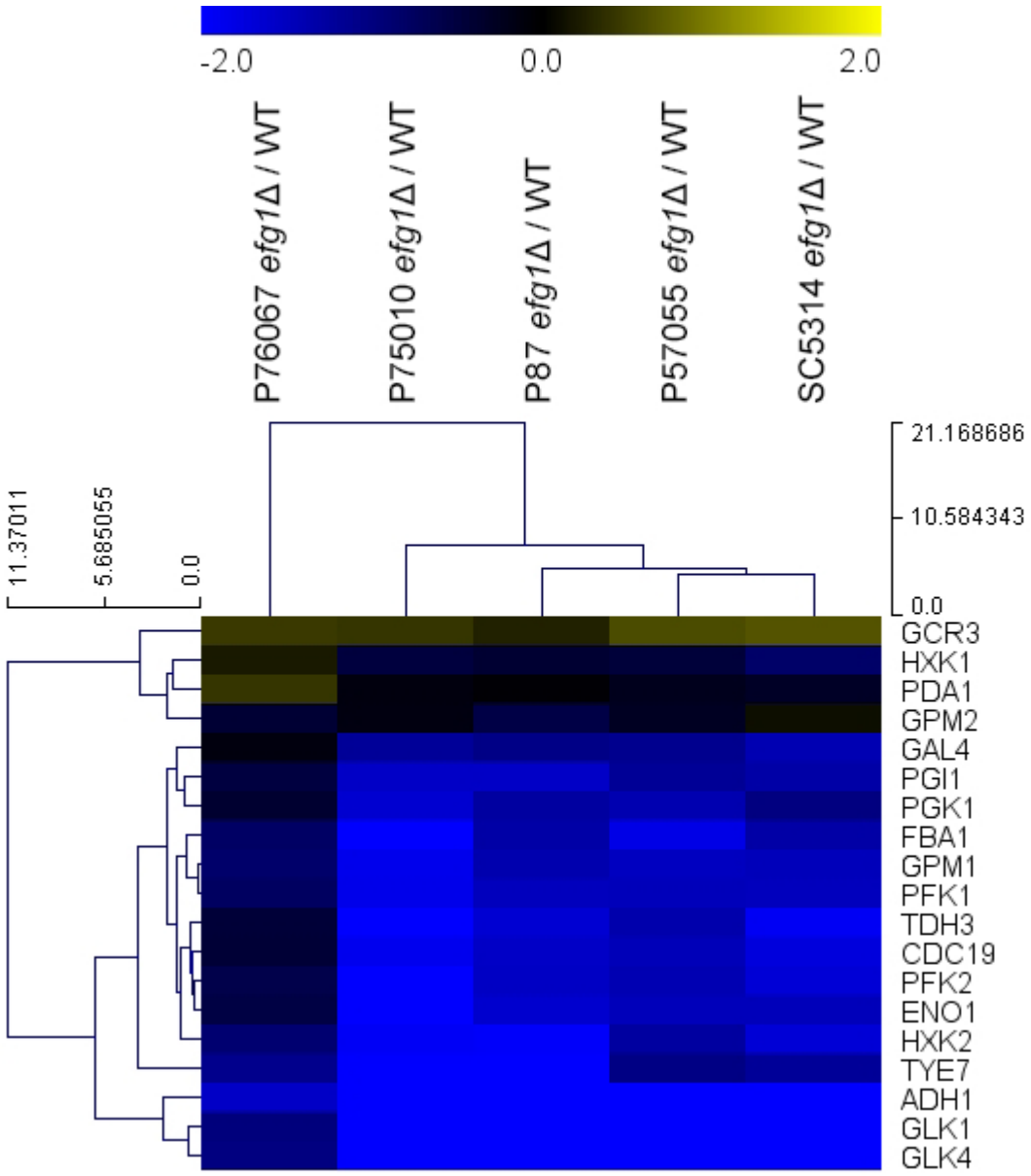
S4 Figure


SC5314
P76067
P57055
P87
P75010
Wild type
efg1Δ/efg1Δ
efg1Δ::EFG1SC5314/ efg1Δ::EFG1SC5314
brg1Δ/brg1Δ
brg1Δ::BRG1SC5314/ brg1Δ::BRG1SC5314
ume6Δ/ume6Δ
ume6Δ::UME6SC5314/ ume6Δ
bcr1Δ/bcr1Δ
bcr1Δ::BCR1SC5314/ bcr1Δ
(*see legend)

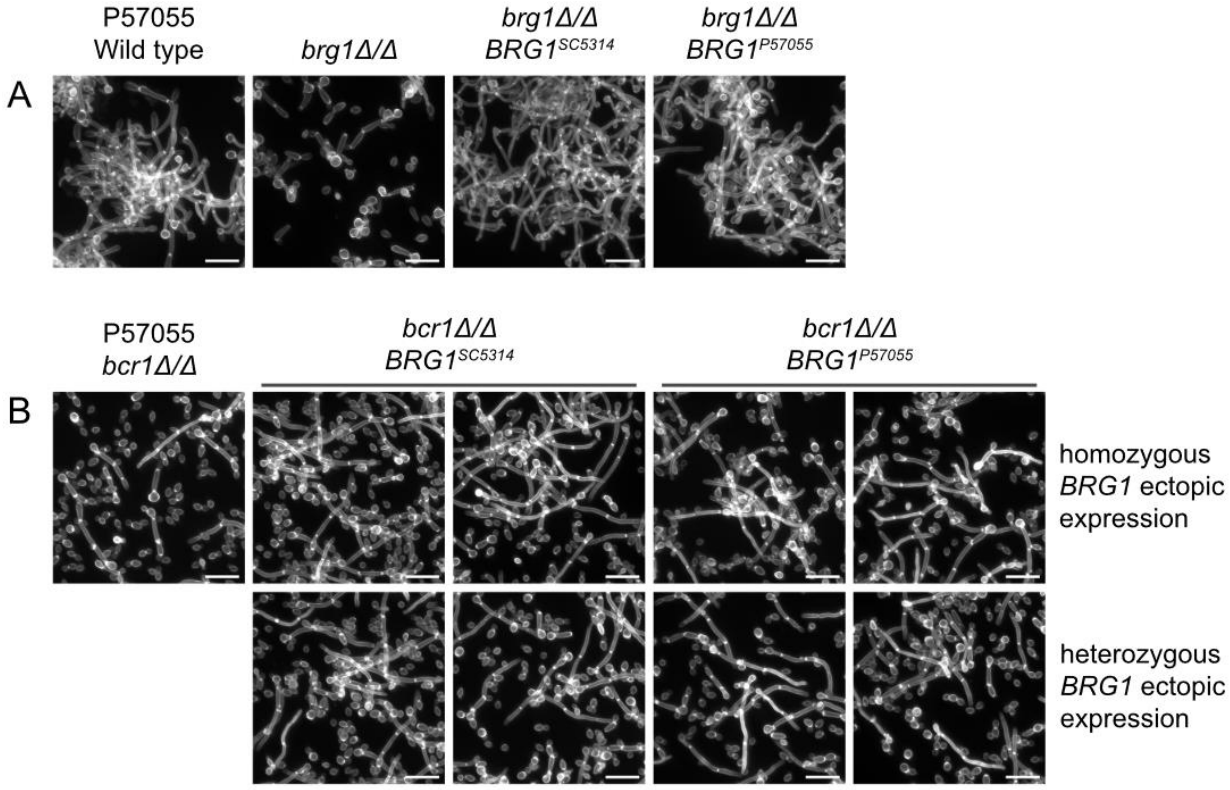
S5 Figure


hyphal unit length (μm)
25
20
15
10
5
0
*
ns
n/a
n/a
P75010 wild type
P75010 efg1Δ/Δ
P75010 EFG1SC5314 / EFG1SC5314
P75010 brg1Δ/Δ
P75010 BRG1SC5314 / BRG1SC5314
hyphal unit : yeast cell ratio
1.5
1.0
0.5
0.0
****
**
P75010 wild type
P75010 efg1Δ/Δ
P75010 EFG1SC5314 / EFG1SC5314
P75010 brg1Δ/Δ
P75010BRG1SC5314 / BRG1SC5314

S6 Figure


-2.0
0.0
2.0
P76067 efg1Δ / WT
P75010 efg1Δ / WT
P87 efg1Δ / WT
P57055 efg1Δ / WT
SC5314 efg1Δ / WT
11.37011
5.685055
0.0
21.168686
10.584343
0.0
GCR3
HXK1
PDA1
GPM2
GAL4
PGI1
PGK1
FBA1
GPM1
PFK1
TDH3
CDC19
PFK2
ENO1
HXK2
TYE7
ADH1
GLK1
GLK4

S7 Figure


P57055
Wild type
brg1Δ/Δ
brg1Δ/Δ
BRG1SC5314
brg1Δ/Δ
BRG1P57055
A
P57055
bcr1Δ/Δ
bcr1Δ/Δ
BRG1SC5314
bcr1Δ/Δ
BRG1P57055
B
homozygous
BRG1 ectopic
expression
heterozygous
BRG1 ectopic
expression

TABLES, CHAPTER 4

Table 1. Affected genes in each strain background

Mutant	SC5314	P76067	P57055	P87	P75010	Common
*bcr1*Δ/Δ	23	33	58	46	57	12
*ume6*Δ/Δ	18	25	34	18	15	0
*brg1*Δ/Δ	69	35	34	27	34	11
*efg1*Δ/Δ	114	56	70	79	56	28

Gene counts are based on Nanostring data (Supplementary Table S1) with a fold-change ≥2 and FDR=0.1. Common genes are those shared among all five strain backgrounds for each TF mutant.

Table 2. Efg1-regulated genes in each strain background

Number of genes	SC5314	P76067	P57055	P87	P75010	Common
Efg1-repressed	327	210	284	243	252	39
Efg1-activated	536	312	352	541	445	138
Total Efg1-responsive	863	522	636	784	697	177
Found by ChIP-ChIP or ChIP-Seq	137	114	117	133	124	42

Gene counts are based on RNA-Seq data (Supplemental Table S3) with a fold-change $\geq$2 and padj < 0.5. Common genes are those shared among all five strain backgrounds.

Chapter 5: Further analyses in circuit diversification and regulatory network core targets

INTRODUCTION

The measure of a transcriptional response or regulatory network's functional relationship with a phenotype is commonly prioritized by investigators into a "core" response or network. These genes may be categorized by effect size on a phenotype or by clearest functional impact, such as in the Boyle "omnigenic" model of complex traits (1). This definition provides an important framework for understanding natural variation. Boyle and colleagues suggest that phenotypic variation of traits coordinated by highly interconnected regulatory networks may occur by variation in peripheral genes that affect regulation or function of core genes (1).

In *C. albicans*, modeling of the filamentation network has previously suggested that regulatory pathways converge on a core set of functional effectors (2,3). Previous analysis of the transcription factors *BCR1, TEC1, NDT80, EFG1, ROB1,* and *BRG1*, identified a set of 19 genes whose expression was regulated by all 6 transcription factors (4). Of these 19 genes, 8 were consistently expressed at lower levels in each of the 6 transcription factor mutants compared to wild type (4). These were *ALS1, TPO4, orf19.3337, orf19.4000, EHT1, HYR1, HWP1,* and *CAN2* (4).

Martin and colleagues defined an environmental condition-independent core transcriptional filamentation response by correlation-based network modeling. They examined transcriptional profiles in three different filament inducing media conditions,

as well as across 4 time points (2). Differentially regulated genes were organized into networks based on highest pearson correlation coefficients, revealing an early and late filamentation network (2). The combination of these two networks revealed an eight member set of genes consistently highly expressed throughout the filamentation process, comprised of *ALS3, DCK1, ECE1, HGT2, IHD1, HWP1, RBT1,* and *orf19.2457* (2). These genes are known to express cell wall or cell membrane proteins, excepting orf19.2457 whose function and cellular component is unknown. Azadmanesh and colleagues also defined a core transcriptional filamentation response, and identified in total 129 genes upregulated and 15 genes downregulated during filamentation across 8 different filament-inducing conditions, including both solid as well as liquid conditions (3).

While these approaches have defined core networks that isolate environmental effects on phenotypic variation, a core transcriptional network has not yet been defined to address genetic background effects on phenotypic variation. We previously examined a 5-strain common set of efg1 regulated targets using RNA-seq and identified 177 target genes which were differentially regulated across all strains. Herein we expand that approach to three additional regulators, and define a set of 28 targets which are differentially expressed in filamentation defective mutant strains. These targets are enriched for biofilm and filamentation process annotated genes, but also contain a large number of uncharacterized targets. We examine these uncharacterized targets for function in filamentation in liquid and solid media conditions across three strains, two

strong biofilm formers and one moderate biofilm former. Our analysis reveals a number of novel functional targets required for filamentation processes.

RESULTS

Circuit Diversification in Biofilm Regulators

Using the clinical isolates P76067, P57055, P87, and P75010, as well as the common lab strain SC5314, we previously identified widespread circuit diversification in the *C. albicans* biofilm regulatory networks under the control of the transcription factors Efg1, Brg1, Ume6, and Bcr1 (See Chapter 4). To identify core targets of these regulatory networks, we extended our analysis genome-wide using RNA-seq. All genes with a fold change in expression greater than 2 and an adjusted p-value less than 0.05 between regulator mutant and associated wild type strain were considered differentially expressed and target genes of that regulator in the given background. Three biological replicates for each transcription factor mutant were analyzed.

The individual target sizes ranged from 327 to 724 for Brg1, 263 to 508 targets for Bcr1, and 152 to 569 targets for Ume6 (Table 1). However, we identified 1578 total unique target genes regulated by Brg1 in at least one of five isolates, 1097 total unique target genes regulated by Ume6 in at least one of five isolates, and 1538 total target genes regulated by Bcr1 in at least one of five isolates (Table 2). For comparison, 1546 total unique target genes were previously found to be regulated by Efg1 in at least one isolate. Furthermore, less than nearly a quarter of targets found in any one clinical isolate background were common across all backgrounds for any transcription factor (Table 1). These observations are consistent with our previous findings for Efg1, and

highlight the ubiquity of circuit diversification in *C. albicans* biofilm regulatory networks (Figure 1).

We previously found that both *bcr1* mutants and *ume6* mutants were defective in biofilm formation only in some clinical isolate backgrounds. Correspondingly, common targets of Ume6 and Bcr1 were not enriched for GO annotated genes involved in biofilm formation processes. Several cell wall genes, including the major adhesin *ALS3* were among core targets of Bcr1, demonstrating a conserved role for Bcr1. Ume6 mutants showed the greatest defect in filamentation formation in the P57055, P76067, P75010 backgrounds, and 6 targets were common to these three strains, but not to SC5314 or P87 (Figure 2). These included *ALS3* in addition to *HGC1*, which has been shown to play critical roles in filamentation and biofilm formation. We previously demonstrated that *brg1* mutants were defective in filamentation and biofilm formation across all strains. As expected, Brg1 common targets were enriched only for GO annotated genes involved in both biofilm formation and adhesion, and included well known genes such as *ALS3, HWP1, HYR1, SAP5, SAP6, HGC1,* and *YWP1*.

Network Modeling

We hypothesized that differential regulation of target genes across backgrounds might occur indirectly through other transcription factors. This was previously shown to underlie differences in *bcr1* mutant phenotypes through *BCR1->BRG1* circuit diversification. *Candida* biofilm regulatory networks show properties characteristic of small-world networks, which are highly connected and efficient in signal propagation. However, the connectivity of such networks also suggests fragility, as perturbations will

similarly propagate efficiently through a network, potentially by connections to other transcription factors.

To estimate the proportion of circuit diversification connected through differential regulation of other transcription factors, we analyzed our RNA-seq using a method modified from the Martin and colleagues network modeling protocol (2). We conservatively identified a total of 150 unique genes annotated as transcription factors manually or by high-throughput experiments on the Candida Genome Database (GOID: 0043565 and 0003700) (5). A total of 55 other transcription factors were differentially expressed across all five isolates in *brg1* mutants, 41 in *efg1* mutants, 52 in *bcr1* mutants, and 31 in *ume6* mutants (Table 3).

We examined the correlation or anticorrelation (referred to simply as correlation) between transcription factor genes and target genes across all 20 conditions (4 transcription factor mutants in 5 backgrounds). For the four master transcription factors (*EFG1, BRG1, UME6, BCR1*), correlation with target genes was only considered in the 15 conditions where that transcription factor had not been deleted. Pairwise Pearson correlation coefficients between all regulated transcription factors and regulated target genes were calculated using the Scipy package (v1.3.2) in Python (v3.7.4). We considered all transcription factor - target gene pairs with a score above 0.75 or below -0.75, corresponding to a maximum p-value of 0.0013. For each transcription factor, a significant proportion of the total unique regulated elements were strongly correlated or anticorrelated with the expression of another regulated transcription factor (Table 2).

More than half of the transcription factors differentially expressed in our master biofilm regulator mutants were annotated for roles in filamentation and biofilm formation. These included well known transcription factors such as *TEC1, NRG1, NDT80, ROB1,* and *CPH1*. Correlations between transcription factors and target genes did not appear to simply reflect an underlying global correlation with filamentation outcomes. While some overlap was observed between genes correlated with biofilm-annotated transcription factors, the correlated gene sets were largely disjoint. For example, the known biofilm regulators *BRG1*, *ROB1*, and *TEC1* were all differentially expressed in some *efg1* mutants. Of all Efg1 regulated target genes, 130 correlated with *BRG1* expression levels, 34 correlated with *ROB1* expression levels, and 79 correlated with *TEC1* expression levels. The overlap between all three sets of correlated genes was limited to 5 genes (Figure 3).

Known targets were found between tightly correlated regulators and target genes. *CPH2* is known to directly regulate *TEC1* expression (6). We observed that *TEC1* was highly correlated with *CPH2* expression, and jointly found among BCR1 and EFG1 regulated targets. Functional relationships were also recapitulated in the correlation data. Of the genes differentially expressed in *bcr1* mutants, 103 were strongly correlated with *BRG1* expression, including many biofilm annotated genes ($p = 0.00019$). We furthermore identified that 64 of these 103 target genes were also differentially expressed in at least one *brg1* mutant. The distribution of target genes correlated with regulated transcription factors for each master regulator are shown in Figure 4. While it remains unclear if or how the correlated targets are truly affected by changes to other transcription factors, these observations illustrate a potential

mechanism for circuit diversification through interactions with other regulatory network members.

Cross-strain Core Targets

We modeled the biofilm regulatory network using the target genes regulated in common across all five isolates by the master regulators, *BCR1, BRG1, UME6,* and *EFG1* (Figure 5). While many targets depended on some combination of multiple master regulators for expression, significant common targets depended only on a single master regulator for expression. Interestingly, we identified no target genes that required all four master regulators for expression, though some became apparent if P75010 was excluded from analysis (data not shown).

We previously observed that *BRG1* and *EFG1* had uniform roles on filamentation and biofilm phenotype, as all *brg1* and *efg1* mutants were defective in both traits across all backgrounds. Given our previous observations with divergent *bcr1* mutant phenotypes, where divergent phenotypic impact arose from divergent regulatory connections, we reasoned that shared phenotypic impact should arise from shared regulatory connections. We identified a core set of 28 genes that required at least *BRG1* and *EFG1* for expression (Table 4). These genes were highly enriched for biofilm and filamentation ($p = 0.0048$), and many of these target genes play known roles in biofilm formation and filamentation, such as *ALS3, HYR1, HWP1, SAP5* and *SAP6* (Table 4). While it is possible that Brg1-Efg1 roles in biofilm and hypha formation are mediated by these known targets genes, several target genes that did not have clear functional roles were also highlighted in this core target set..

We reasoned that these core targets may be enriched for functional relationships. We functionally analysed core targets by constructing parallel deletion mutants of core targets in a panel of 3 isolates, P76067, P57055, and SC5314. Using previously described CRISPR-Cas9 techniques, we successfully deleted 25 of the 28 core target genes in all three clinical isolate backgrounds. Where practicable, our CRISPR induced marker excision method was employed to excise the NAT1 marker resident at the *his1* locus to allow for future manipulations of mutant strains. Otherwise, where possible, two independent isolates were catalogued for subsequent phenotyping. Filamentation phenotypes were examined for planktonic growth in RPMI + serum or Spider media, alongside surface-associated filamentous growth on Spider plates. Mutant phenotypic outcomes are summarized in Table 5.

Many core target mutants did not display aberrant filamentation phenotypes under our assay conditions. Two core targets that did not display filamentation defects were well characterized adhesin genes, *ALS3* and *HWP1*. While *als3* and *hwp1* mutants were not defective in filamentation, mutants in all backgrounds failed to flocculate under planktonic growth conditions, consistent with defects in cell-cell adhesion (data not shown). Other core target mutants also recapitulated known phenotypes. Hgc1 is a G1 cyclin related protein that has been shown to be required for both filamentation and biofilm formation in the SC5314 background (7,8). While *hgc1* mutants were defective in filamentation under both planktonic growth conditions in all backgrounds, *hgc1* mutant severity was greater in the P76067 and P57055 backgrounds (Figure 6). In RPMI + serum, the SC5314 *hgc1* mutant is still capable of forming infrequent true hyphae, but predominantly forms pseudohyphae, whereas in a P76067 *hgc1* mutant, true hyphae

are absent, and the mutant forms only shorter, non-adherent pseudohyphae. A P57055 forms predominantly yeast-form cells, with pseudohyphae infrequently observed. These data suggest that *HGC1* function may be partially bypassed in *SC5314* background, but not in P57055 or P76067.

We identified novel roles in filamentation for several previously uncharacterized genes. Compared to wildtype, *ihd2* mutants form colonies with hyphal portrustions that extend farther into the surrounding media (Figure 7A). This phenotype was clearly evident in the P76067 background, and to a lesser extent observed in the SC5314 background. In contrast, an *ihd1* mutant in the P57055 but not the P76067 background is hyperfilamentous under planktonic growth conditions in liquid Spider media, as well as hyperfilamentous when grown on solid Spider plates (Figure 6, Figure 7A). (Increased filamentation was also observed to a lesser extent in a P76067 *ihd1* mutant under planktonic growth conditions in certain batches of Spider media.) Only a single *ihd1* mutant was obtained in the P57055 background, and additional independent isolates or complement strains will be required. An *erg251* mutant in the P76067 and P57055 backgrounds was defective in filamentation when grown on solid Spider plates (Figure 7B). This phenotype was not readily apparent in the SC5314 background, where an *erg251* mutant only showed aberrant colony morphology.

METHODS

Strains and Media

The following *C*. *albicans* clinical isolate strains were obtained through BEI Resources, NIAID, NIH: Strain P76067, NR-29442; Strain P57055, NR-29439; Strain P87, NR-29453; Strain P75010, NR-29437.

All strains were archived in 15% glycerol stocks stored at -80°C. For non-assay growth, strains were grown on YPD (2% Bacto Peptone, 2% dextrose, 1% yeast extract) for 2 days at 30°C, and then cultured overnight in liquid YPD at 30°C in a roller drum incubator. For all transformations, cells were plated on complete synthetic media lacking histidine (CSM -His) (2% dextrose, 1.7% Difco yeast nitrogen base with ammonium sulfate and auxotrophic supplements without histidine). For phenotypic assays, strains were grown in liquid RPMI-1640 Media (Sigma-Aldrich, Inc., St. Louis) adjusted to pH 7.4 and supplemented with 10% fetal bovine serum (Atlanta Biologicals, Inc., Flowery Branch), Spider Media (1% D-Mannitol, 1% Nutrient Broth, 0.2% Potassium Phosphate Dibasic) with 2% Bacto Agar for solid media. A list of strains used in this study is available upon request (provided under supplemental files (S1 Table)).

Primers, Plasmids, and Strain Construction

A list of primers and plasmids used in this study is available upon request (provided under supplemental files (S2 Table)).

To generate each core target mutant, cells were transformed with 3 ug *yfgXΔ::r1HIS1r1* deletion cassette, 1 ug Cas9 DNA cassette (also referred to as repair template), 1 ug

sgRNA DNA cassette targeting the gene of interest, and optionally 1 ug NAT1-5 sgRNA DNA cassette. All strains marked *his1*Δ::*r3* in the strain list are Nat sensitive and had a resident *NAT1* marker excised from the *his1Δ* locus.

For each *YFGX*, the *yfgXΔ::r1HIS1r1* cassette was generated per previously discussed methods in two pieces. The first piece was amplified from plasmid pMH01 using primers “HIS1 CRIME/F” and “yfgX del rHISr-KpnI/R”, and the second was amplified from plasmid pMH02 using primers “yfgX del rHISr-SapI/F” and “HIS1 CRIME/R”.

For each *YFGX*, the sgRNA DNA cassette was generated per methods discussed by Min and colleagues (9,10). Briefly, sgRNA DNA cassettes were amplified by split-joint PCR using “YFGX sgRNA/F”, “sgRNA/R”, “sgRNA/F” and “YFGX SNR52/R” with pV1093 as a template. When utilized, the NAT1-5 sgRNA DNA cassette was generated by split-joint PCR using “NAT1-5 sgRNA/F”, “sgRNA/R”, “SNR52/F” and “NAT1-5 SNR52/R” with pV1093 as a template as well. sgRNA design principles have been discussed elsewhere, but guides targeting either end of the ORF of the gene of interest have generally proven to be efficient.

The Cas9 DNA cassette was amplified from pV1093 as previously described (9,10).

Hyphal Induction Assays under Planktonic Growth Conditions

In either liquid RPMI + Serum or Spider Media, strains were inoculated from overnight cultures into 5 ml of prewarmed media at an OD600 of 0.5. Cells were grown in a roller drum incubator for 4 hours at 37 degrees celsius at 60 rpm. Cells were collected by centrifugation and fixed with 4% formaldehyde for 15 minutes. Fixed cells were washed twice in PBS and stained with Calcofluor-white. Stained cells were imaged using a slit-scan confocal optical unit on a Zeiss Axiovert 200 microscope with a Zeiss C-Apochromat 40x/1.2 NA water immersion objective.

Colony Morphology Assays

Overnight cultures were diluted in PBS to an OD600 of 0.3, then subsequently serially diluted five-fold six times to an OD600 of 1.9 x 10-5. 50 ul of diluted cell suspension were then plated on Spider plates and grown for 7 days at 37 degrees Celsius. Plates were subsequently photographed and examined for differences in colony morphology.

RNA-seq

RNA-seq was performed on RNA-samples previously harvested, and using the same protocol as previously described in Chapter 4. Differential expression was assessed using DEseq2 (v 1.22.1) (11) in R (v 3.5.1) using default options (alpha = 0.05).

Network Modeling and Software

Network modeling was performed based on methods described by Martin and colleagues (2). Using RNA-seq expression profiles across *brg1, efg1, ume6,* and *bcr1* mutants in all five clinical isolate backgrounds, pairwise Pearson correlation coefficients were calculated between differentially expressed transcription factors and differentially

expressed target genes using the Scipy package (v1.3.2) in Python (v3.7.4). All correlations scoring above 0.75 or below -0.75 were considered. Correlations between differentially expressed transcription factors and target genes per each master regulator are available upon request (provided under supplemental files (S3 Table)).

Graphical representations of networks were created in Cytoscape (12). Some statistical analyses were performed with Graphpad Prism version 8.00 (Graphpad Software, Inc., La Jolla). Venn diagrams were constructed using Venn Diagrams software (http://bioinformatics.psb.ugent.be/webtools/Venn/).

DISCUSSION

In this chapter, we extended our expression profiling of regulatory mutants to include genome-wide transcriptome data. We observed that genome-wide, circuit diversification was recaputiluated at an even greater scale for all transcription factors. Furthermore, we leveraged natural variation across a panel of five isolates to identify a core set of 28 genes regulated by master biofilm regulators and consistently differentially expressed in filamentation defective master regulator mutants. Within these 28 genes, we identified functional relationships connecting three previously uncharacterized genes, *IHD1, IHD2,* and *ERG251* with filamentation phenotypes.

Our analysis of regulatory relationships suggests one mechanistic model for circuit diversification. While circuit diversification may occur at many levels of regulation, the regulatory relationships connecting master regulators and other transcription factors possibly account for a significant proportion of regulatory variation between strains. ChIP-seq experiments may reveal if master regulators are less likely to bind upstream of genes whose expression correlates strongly with secondary transcription factors. The phenotypic consequences of secondary transcription factor gene expression programs also remains to be explored.

Our data suggest the uniform *brg1* mutant filamentation defective phenotype occurs primarily through diminished expression of *HGC1*. The filamentation defect of *hgc1* mutants closely phenocopied the severity of *brg1* filamentation defects. One disadvantage in our approach was that we were limited by the overlap of *brg1* and *efg1* mutant filamentation defects. While *efg1* mutants form only yeast form cells, *brg1*

mutants may still form pseudohyphae. The expression changes that result in the more severe *efg1* mutant may logically be absent from *brg1* mutants.

While our assays primarily examined filamentation capacity, core targets may be involved in separate aspects of functions shared between relevant regulators, as evidenced by *ALS3* and *HWP1*. These two genes were identified as core targets of both Brg1 and Efg1, but did not display a defect in filamentation. Instead, the presence of these genes in the core joint Brg1-Efg1 response appears to reflect the shared biofilm defective phenotype common to *brg1* and *efg1* mutants in all isolate backgrounds. Seven other cell wall or cell membrane predicted genes are present in the core set, and may play roles in adhesion or biofilm formation.

Of the three previously uncharacterized genes with newly identified roles in filamentation, *IHD1* and *ERG251* are predicted to localize to the cell membrane. *ERG251* functions in the synthesis of ergosterol, an integral membrane component required for membrane fluidity and integrity (13,14). *IHD1* has been annotated as a putative GPI-linked cell wall protein commonly induced during hyphal development, though its function is unknown (15). For cell periphery genes to affect filamentation, these genes may play roles in signal sensing or signal transduction. Furthermore, given that their mutant phenotypes were only apparent in some strains, complex genetic interactions may exist to mask their mutant phenotypes. One strong candidate for interaction partners would be other cell periphery predicted core targets, and an examination of double mutants may be informative.

LITERATURE CITED, CHAPTER 5

1. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell. 2017 Jun 15;169(7):1177–86.
2. Martin R, Albrecht-Eckardt D, Brunke S, Hube B, Hünniger K, Kurzai O. A core filamentation response network in Candida albicans is restricted to eight genes. PLoS One. 2013;8(3):e58613.
3. Azadmanesh J, Gowen AM, Creger PE, Schafer ND, Blankenship JR. Filamentation Involves Two Overlapping, but Distinct, Programs of Filamentation in the Pathogenic Fungus Candida albicans. G3 (Bethesda). 2017;7(11):3797–808.
4. Nobile CJ, Fox EP, Nett JE, Sorrells TR, Mitrovich QM, Hernday AD, et al. A recently evolved transcriptional network controls biofilm development in Candida albicans. Cell. 2012;148(1–2):126–38.
5. Skrzypek MS, Binkley J, Binkley G, Miyasato SR, Simison M, Sherlock G. The Candida Genome Database (CGD): incorporation of Assembly 22, systematic identifiers and visualization of high throughput sequencing data. Nucleic Acids Res. 2017;45(D1):D592–6.
6. Lane S, Zhou S, Pan T, Dai Q, Liu H. The basic helix-loop-helix transcription factor Cph2 regulates hyphal development in Candida albicans partly via TEC1. Mol Cell Biol. 2001 Oct;21(19):6418–28.
7. Lin CH, Kabrawala S, Fox EP, Nobile CJ, Johnson AD, Bennett RJ. Genetic Control of Conventional and Pheromone-Stimulated Biofilm Formation in Candida albicans. PLoS Pathog. 2013;9(4).
8. Zheng X, Wang Y, Wang Y. Hgc1, a novel hypha-specific G1 cyclin-related protein regulates Candida albicans hyphal morphogenesis. EMBO J. 2004 Apr 21;23(8):1845–56.
9. Min K, Ichikawa Y, Woolford CA, Mitchell AP. Candida albicans Gene Deletion with a Transient CRISPR-Cas9 System. mSphere. 2016;1(3):1–9.
10. Vyas VK, Barrasa MI, Fink GR. A Candida albicans CRISPR system permits genetic engineering of essential genes and gene families. Sci Adv. 2015;1(3):e1500248.
11. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
12. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003 Nov;13(11):2498–504.
13. Liu TT, Lee REB, Barker KS, Lee RE, Wei L, Homayouni R, et al. Genome-wide expression profiling of the response to azole, polyene, echinocandin, and pyrimidine antifungal agents in Candida albicans. Antimicrob Agents Chemother. 2005 Jun;49(6):2226–36.
14. Lv Q-Z, Yan L, Jiang Y-Y. The synthesis, regulation, and functions of sterols in Candida albicans: Well-known but still lots to learn. Virulence. 2016;7(6):649–59.
15. De Groot PWJ, Hellingwerf KJ, Klis FM. Genome-wide identification of fungal GPI proteins. Yeast. 2003 Jul 15;20(9):781–96.

FIGURE LEGENDS, CHAPTER 5

Figure 1. *C. albicans* RNA-seq based biofilm regulator networks

Regulatory relationships between master regulators and all target genes are shown for each clinical isolate. **A.** SC5314 **B.** P76067 **C.** P57055 **D.** P87 **E.** P75010 **F.** Common across all five isolates. Outer rings consist of nodes representing genes differentially expressed in at least one master regulator mutant in at least one background. Node positions are the same in each panel. Center nodes correspond to master regulators, *EFG1, BRG1, UME6,* and *BCR1*. Edges connecting center nodes to outer nodes denote differential expression of the target gene in the connected master regulator mutant. Networks are justified based on the SC5314 regulatory connections. Coloring of outer nodes denotes number of master regulators required for expression: Blue, 1; Purple, 2; Red, 3; Yellow, 4.

Figure 2. Intersection of master regulator targets per isolate background

Venn diagrams depicting intersection of genes depending on each transcription factor per clinical isolate background.

Figure 3. Intersection of select regulated TF correlated targets

Venn diagram depicting intersection of Efg1 target genes who significantly correlated with *ROB1, TEC1,* and *BRG1* respectively.

Figure 4. Distribution of all regulated TF correlated targets

X-axis in each bar graph lists all the differentially expressed transcription factors in at least one clinical isolate background for the indicated master regulator. Bar graphs depict number of differentially expressed target genes that strongly correlate with a given transcription factor.

Figure 5. Common biofilm regulatory network

Regulatory relationships from Figure 1F are arranged to show individual target genes. Target genes are grouped by relationship to master regulators. Target nodes are colored to denote number of master regulators required for expression: Blue, 1; Red, 2; Yellow, 3.

Figure 6. Hyphal induction assays

Filamentation for wild type, *ihd1*, and *hgc1* mutants under planktonic growth conditions in Spider or RPMI + Serum are shown. Strains were grown in each inducing media at 37 degrees Celsius for 4 hours in a roller drum incubator. Scale bars are 20μm in length.

Figure 7. Colony morphology assays

Strains were grown on Spider agar for 7 days at 37 degrees Celsius. Where possible, two isolates were assayed for colony morphology. **A.** Colony morphologies of *ihd2* and *ihd1* mutants in corresponding clinical isolate backgrounds. Only one *ihd1* mutant was obtained in the P57055 background. **B.** Colony morphologies of *erg251* mutants in corresponding clinical isolate backgrounds. Panel A and Panel B depict experiments performed on separate batches of Spider media.

FIGURES, CHAPTER 5

Figure 1.

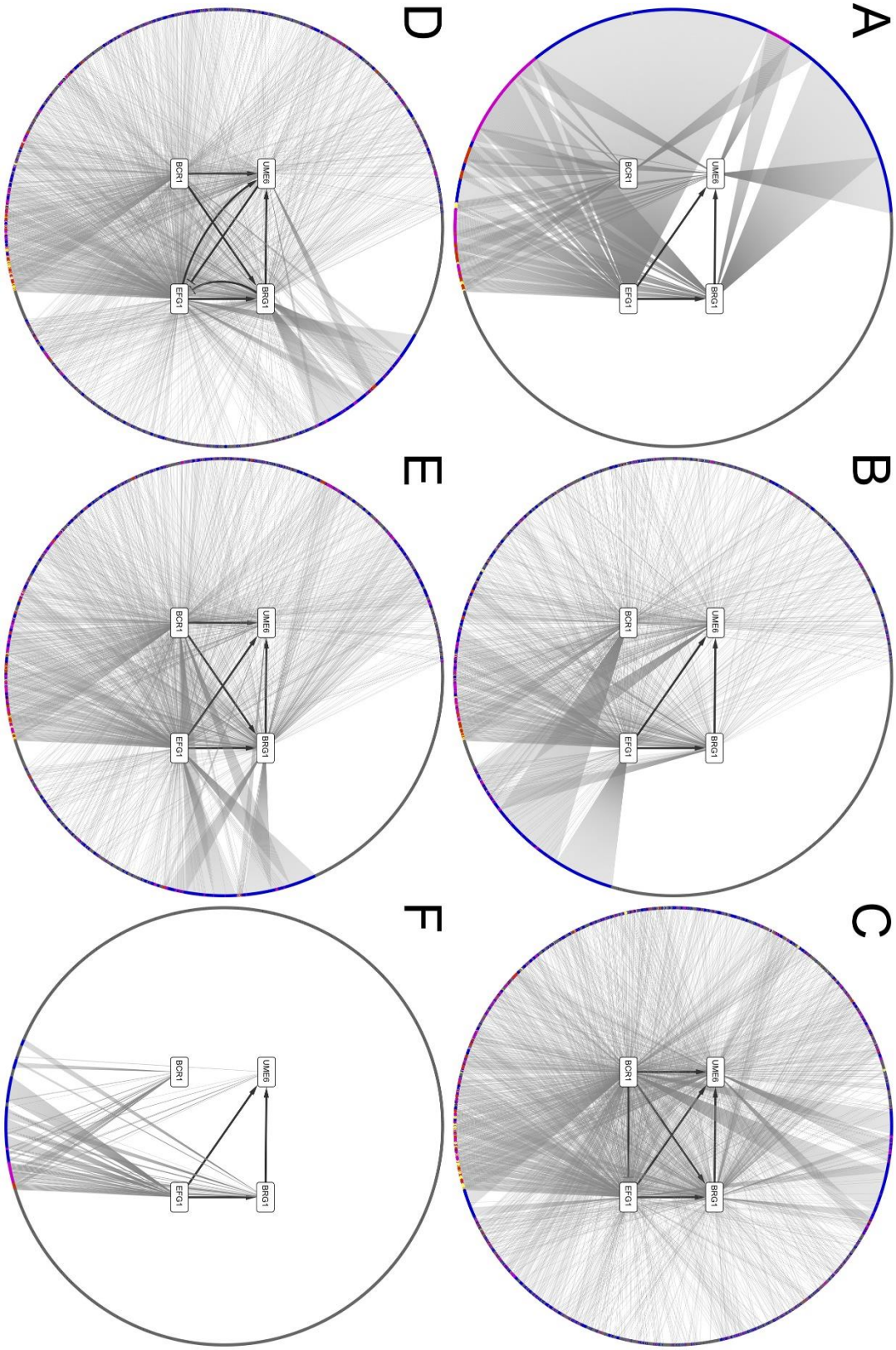


Figure 2.

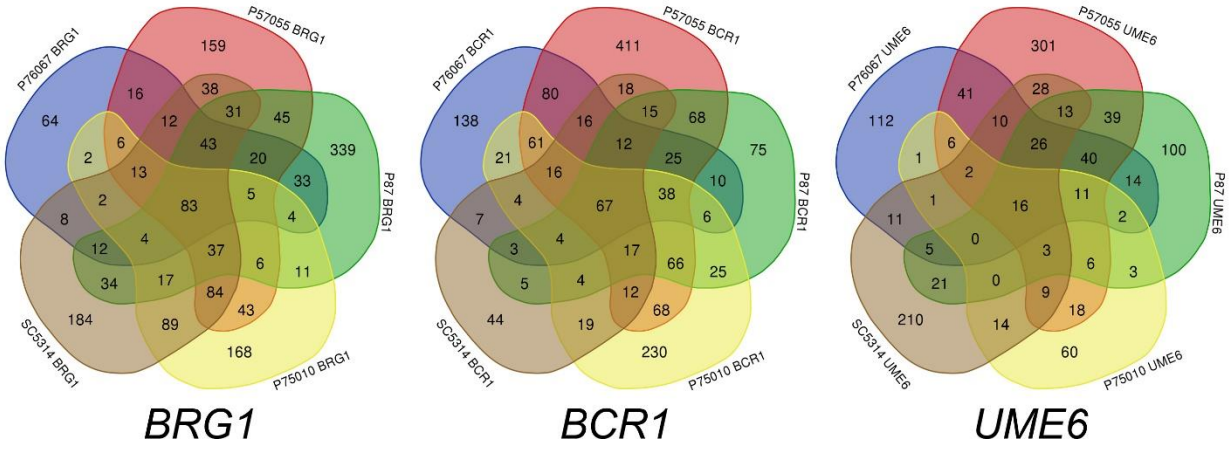


Figure 3.

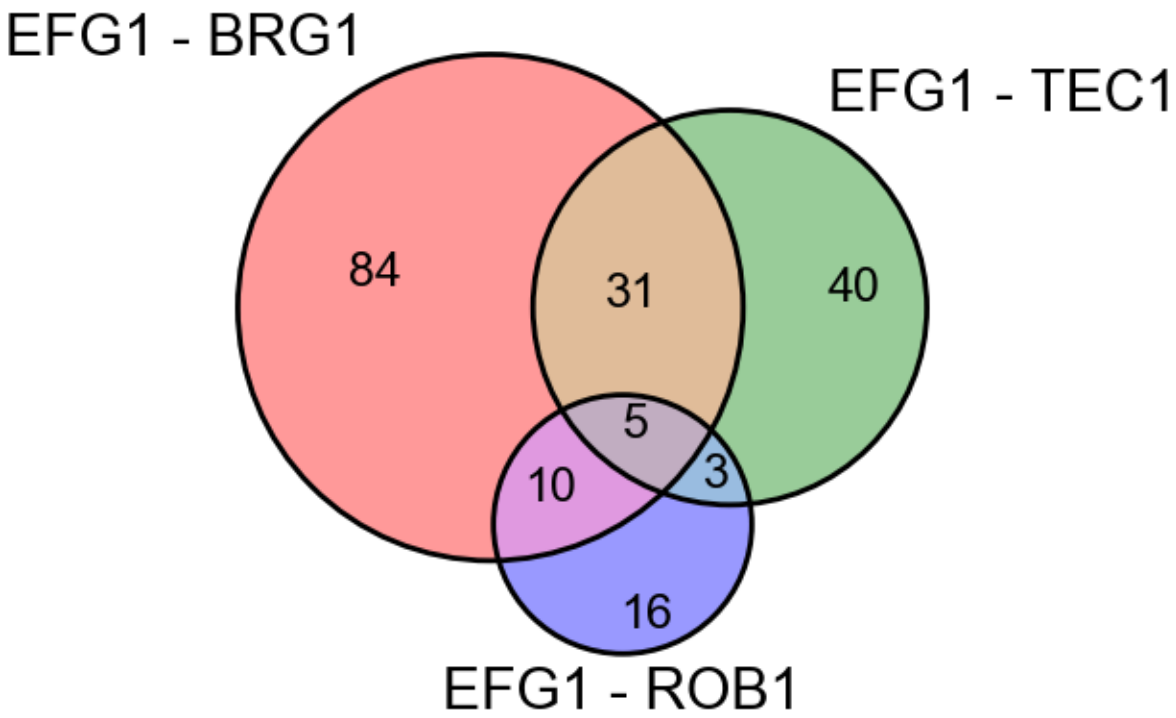


Figure 4.

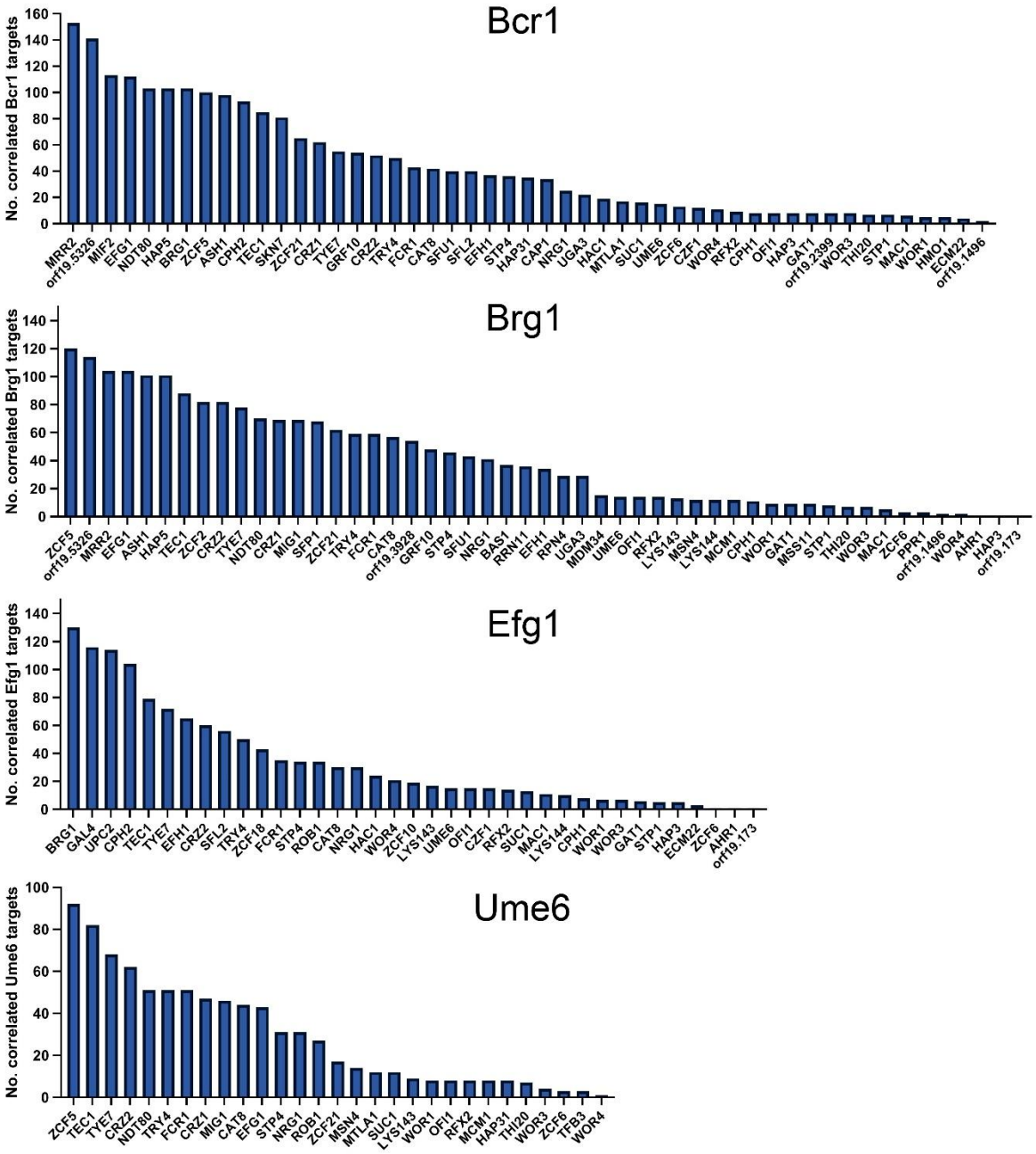

Bcr1
No. correlated Bcr1 targets
MRR2 orf19.5326 MIF2 EFG1 NDT80 HAP5 BRG1 ZCF5 ASH1 CPH2 TEC1 SKN7 ZCF21 CRZ1 TYE7 GRF10 CRZ2 TRY4 FCR1 CAT8 SFU1 SFL2 EFH1 STP4 HAP31 CAP1 NRG1 UGA3 HAC1 MTLA1 SUC1 UME6 ZCF6 CZF1 WOR4 RFX2 CPH1 OFI1 HAP3 GAT1 orf19.2399 WOR3 THI20 STP1 MAC1 WOR1 HMO1 ECM22 orf19.1496
Brg1
No. correlated Brg1 targets
ZCF5 orf19.5326 MRR2 EFG1 ASH1 HAP5 TEC1 ZCF2 CRZ2 TYE7 NDT80 CRZ1 MIG1 SFP1 ZCF21 TRY4 FCR1 CAT8 orf19.3928 GRF10 STP4 SFU1 NRG1 BAS1 RRN11 EFH1 RPN4 UGA3 MDM34 UME6 OFI1 RFX2 LYS143 MSN4 LYS144 MCM1 CPH1 WOR1 GAT1 MSS11 STP1 THI20 WOR3 MAC1 ZCF6 PPR1 orf19.1496 WOR4 AHR1 HAP3 orf19.173
Efg1
No. correlated Efg1 targets
BRG1 GAL4 UPC2 CPH2 TEC1 TYE7 EFH1 CRZ2 SFL2 TRY4 ZCF18 FCR1 STP4 ROB1 CAT8 NRG1 HAC1 WOR4 ZCF10 LYS143 UME6 OFI1 CZF1 RFX2 SUC1 MAC1 LYS144 CPH1 WOR1 WOR3 GAT1 STP1 HAP3 ECM22 ZCF6 AHR1 orf19.173
Ume6
No. correlated Ume6 targets
ZCF5 TEC1 TYE7 CRZ2 NDT80 TRY4 FCR1 CRZ1 MIG1 CAT8 EFG1 STP4 NRG1 ROB1 ZCF21 MSN4 MTLA1 SUC1 LYS143 WOR1 OFI1 RFX2 MCM1 HAP31 THI20 WOR3 ZCF6 TFB3 WOR4

Figure 5.

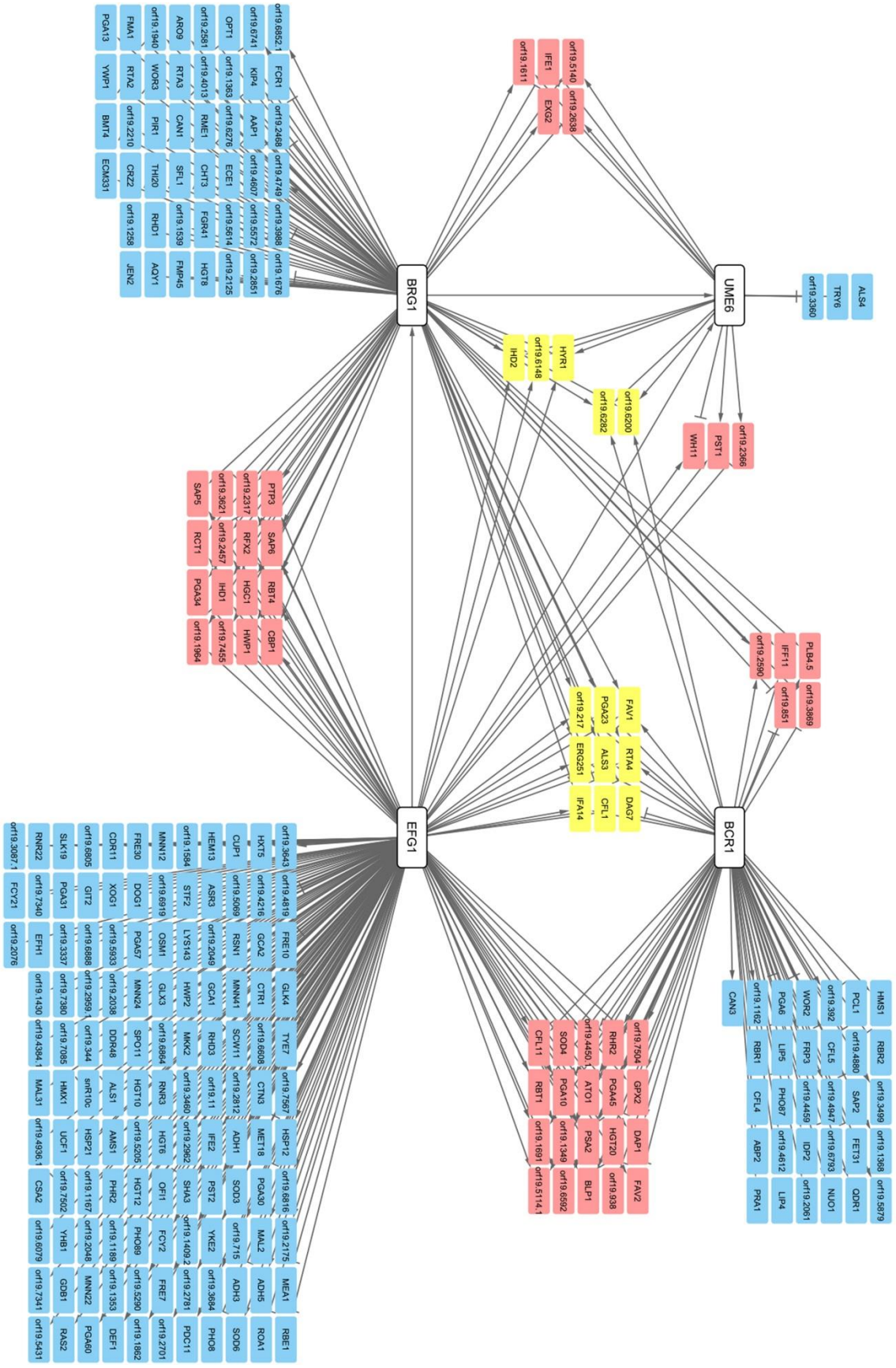

BRG1
UME6
EFG1
BCR1
ALS4
TRY6
orf19.3380
orf19.1611
IFE1
EXG2
orf19.5140
orf19.2638
IHD2
orf19.6148
HYR1
orf19.6282
orf19.6200
WH11
PST1
orf19.2366
PLBA.5
IFF11
orf19.2590
orf19.851
orf19.3869
FAV1
PGA23
orf19.217
ALS3
ERG251
RTA4
CFL1
IFA14
DAG7
PTP3
orf19.2317
SAP6
RBT4
HGC1
HWP1
CBP1
orf19.7455
SAP5
orf19.3821
RFX2
IHD1
orf19.2457
RCT1
PGA34
orf19.1964
orf19.68521
FCR1
orf19.2468
orf19.4749
orf19.3808
orf19.1676
orf19.6741
KIP4
AAP1
orf19.4607
orf19.5572
orf19.2851
OPT1
orf19.1363
orf19.6276
ECE1
orf19.5614
orf19.2125
orf19.2581
orf19.4013
RME1
CHT3
FGR41
HGT8
ARO9
RTA3
CAN1
SFL1
orf19.1539
FMP45
PGA13
YWP1
BMT4
ECM331
FMA1
RTA2
orf19.2210
CRZ2
orf19.1258
RHD1
JEN2
orf19.1940
WOR3
PIR1
THI20
AQY1
HMS1
PCL1
RBR2
orf19.3499
orf19.1368
orf19.5879
orf19.392
orf19.4880
SAP2
FET31
QDR1
CFL5
orf19.4947
orf19.6793
NUO1
WOR2
FRP3
orf19.4459
IDP2
orf19.2061
PGA6
LIP5
PHO87
orf19.4612
LIP4
orf19.1162
RBR1
CFL4
ABP2
PRA1
CAN3
orf19.7504
GPX2
DAP1
FAV2
RHR2
PGA45
HGT20
orf19.938
orf19.4450
ATO1
PSA2
BLP1
SOD4
PGA10
orf19.1349
orf19.6592
CFL11
RBT1
orf19.1691
orf19.5114.1
orf19.3643
orf19.4819
FRE10
GLK4
TYE7
orf19.7567
HSP12
orf19.6816
orf19.2175
MEA1
RBE1
HXT5
orf19.4216
GCA2
CTR1
orf19.6808
CTN3
MET18
PGA30
MAL2
ADH5
ROA1
CUP1
orf19.5069
RSN1
MNN41
SCW11
orf19.2812
ADH1
SOD3
orf19.715
ADH3
SOD6
HEM13
ASR3
orf19.2049
GCA1
RHD3
orf19.11
IFE2
PST2
YKE2
orf19.3684
PHO8
orf19.1584
STF2
LYS143
HWP2
MKK2
orf19.3460
orf19.2962
SHA3
orf19.1409.2
orf19.2781
PDC11
MNN12
orf19.6919
OSM1
GLX3
orf19.6884
RNR3
HGT6
OFI1
FCY2
FRE7
orf19.2701
FRE30
DOG1
PGA57
MNN24
SPO11
HGT10
orf19.5205
HGT12
PHO89
orf19.5290
orf19.1862
CDR11
XOG1
orf19.5933
orf19.2038
DDR48
ALS1
AMS1
PHR2
orf19.1189
orf19.1353
DEF1
orf19.6805
GIT2
orf19.6888
orf19.2959.1
orf19.344
snR10c
HSP21
orf19.1167
orf19.2048
MNN22
PGA60
SLK19
PGA31
orf19.3337
orf19.7380
orf19.7085
HMX1
UCF1
orf19.7502
YHB1
GDB1
RAS2
RNR22
orf19.7340
EFH1
orf19.1430
orf19.4384.1
MAL31
orf19.4936.1
CSA2
orf19.6079
orf19.7341
orf19.5431
FCY21
orf19.2076
orf19.3087.1

Figure 6.

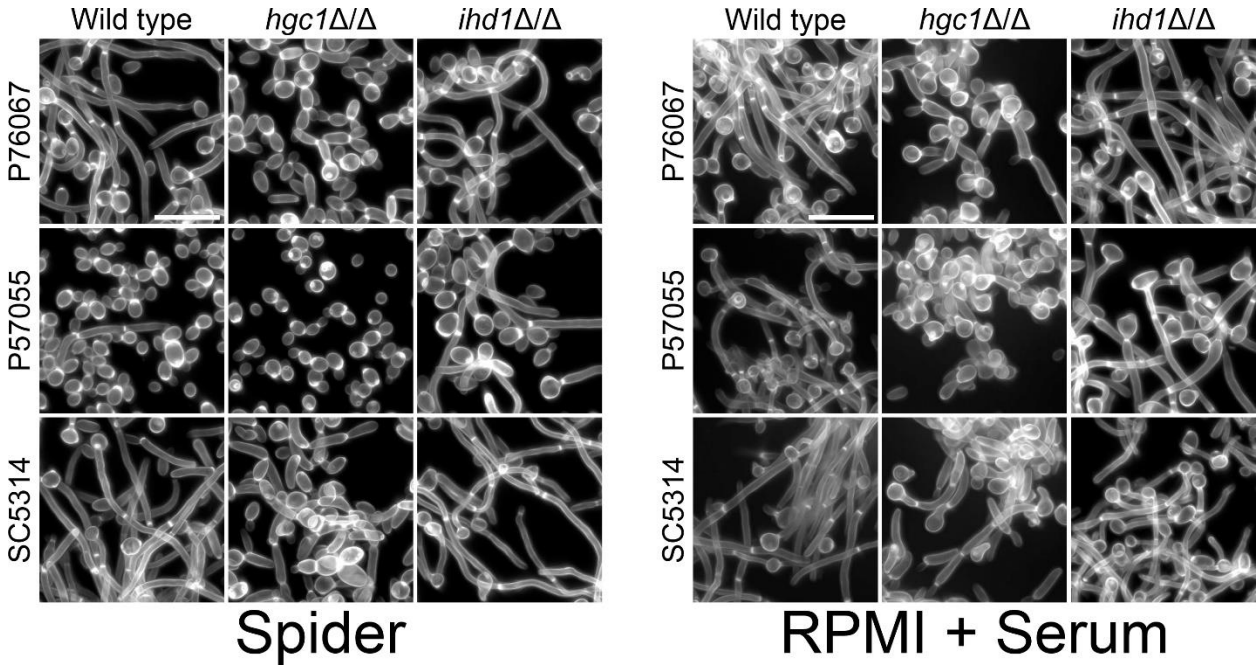
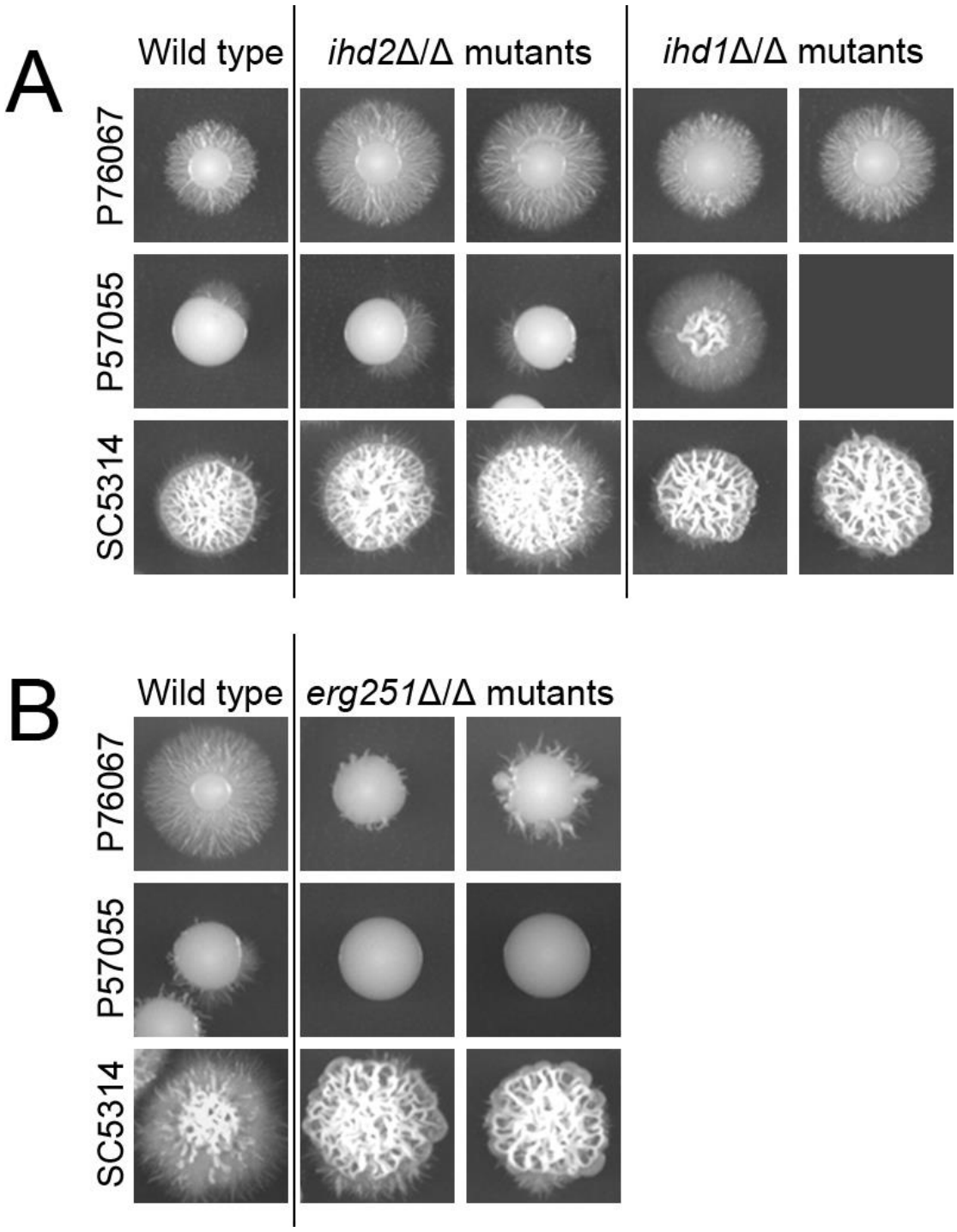

Wild type
hgc1Δ/Δ
ihd1Δ/Δ
P76067
P57055
SC5314
Spider
Wild type
hgc1Δ/Δ
ihd1Δ/Δ
P76067
P57055
SC5314
RPMI + Serum

Figure 7.


A
Wild type
ihd2Δ/Δ mutants
ihd1Δ/Δ mutants
P76067
P57055
SC5314
B
Wild type
erg251Δ/Δ mutants
P76067
P57055
SC5314

TABLES, CHAPTER 5

Table 1. Biofilm Regulator Target Sizes

	Numbers of TF responsive genes					
	P76067	P57055	P87	P75010	SC5314	Common
BCR1	508	990	440	658	263	67
UME6	298	569	299	152	369	16
BRG1	327	641	724	574	691	83
	Proportion of targets common across isolates					
	P76067	P57055	P87	P75010	SC5314	Max
BCR1	0.132	0.068	0.152	0.102	0.255	0.255
UME6	0.054	0.028	0.054	0.105	0.043	0.105
BRG1	0.254	0.129	0.115	0.145	0.120	0.254

Table 2. Circuit Diversification Network Modeling

	Total Unique Regulated Targets	Total TF Correlated Elements
BCR1	1538	810
UME6	1097	398
BRG1	1578	752
EFG1	1546	607

Table 3. Regulated TF Per Background

	Regulated TF per background					
	P76067	P57055	P87	P75010	SC5314	Unique TF
BCR1	11	33	17	28	6	52
UME6	8	16	16	3	11	31
BRG1	16	29	31	18	28	55
EFG1	20	20	24	23	27	41

Table 4. Summary of Core Targets

Gene:	Connection to Hyphae/Biofilm Formation	Summary:
FAV1	No	
RBT4	No	
RFX2	Yes	Transcriptional repressor that regulates filamentation and adhesion. Mutant reported to be hyperfilamentous on solid media.
ALS3	Yes	Cell wall adhesin required for biofilm formation.
IHD2	Yes*	Protein of unknown function whose mutant appears hyperfilamentous on solid media (in P76067).
HGC1	Yes	G1 cyclin-related protein required for filamentation and biofilm formation through interaction with Cdc28.
IHD1	Yes*	Cell wall protein whose mutant appears appears hyperfilamentous in solid and liquid media (in P57055).
RCT1	No	
HYR1	No	
RTA4	Yes	Fatty acid transporter that is related to filamentous growth; a transposon insertion mutant proximal to the *RTA4* ORF shows decreased filamentation on solid Spider media.
CFL1	Yes	Ferric reductase involved in cell wall integrity, deletion leads to an iron-dependent defect in filamentation and decreased adhesion to polystyrene.
DAG7	No	
PGA23	No	
ERG251	Yes*	C-4 sterol methyl oxidase involved in ergosterol synthesis. Deletion leads to a filamentation defect on solid Spider media.
orf19.6148	No	
orf19.1964	No	
orf19.7455	No	
orf19.2317	No	
CBP1	No	
PGA34	No	
IFA14	No	
orf19.2457	No	
HWP1	Yes	Cell wall adhesin required for biofilm formation.
orf19.3621	No	
orf19.217	No	
SAP5	Yes	Secreted aspartyl protease, mutant shows decreased biofilm formation.
SAP6	Yes	Secreted aspartyl protease, mutant shows decreased biofilm formation.
PTP3	No	

* Denotes characterized in this work

Table 5. Mutant Phenotypes of Brg1-Efg1 Core Targets

	P76067			P57055			SC5314		
Gene:	Spider-L	RPMI + Serum L	Spider Solid	Spider-L	RPMI + Serum L	Spider Solid	Spider-L	RPMI + Serum L	Spider Solid
FAV1	normal	normal	n/a	normal	normal	n/a	normal	normal	n/a
RBT4	normal	normal	n/a	normal	normal	n/a	normal	normal	n/a
RFX2	normal	hyperfil	n/a	normal	normal	n/a	normal	normal	n/a
ALS3	normal	normal	n/a	normal	normal	n/a	normal	normal	n/a
IHD2	normal	normal	hyperfil	normal	normal	normal	normal	normal	hyperfil
HGC1	defective	defective	n/a	defective	defective	n/a	defective	defective	n/a
IHD1	hyperfil	normal	normal	hyperfil	normal	hyperfil	normal	normal	normal
RCT1	normal	normal	abnormal	normal	normal	normal	normal	normal	normal
HYR1	normal	normal	normal	normal	normal	normal	normal	normal	normal
RTA4	normal	normal	n/a	normal	normal	n/a	normal	normal	n/a
CFL1	normal	normal	abnormal	normal	normal	defective	normal	normal	abnormal
DAG7	normal	normal	normal	normal	normal	normal	normal	normal	abnormal
PGA23	normal	normal	normal	normal	normal	normal	normal	normal	abnormal
ERG251	normal	normal	defective	normal	normal	defective	normal	normal	abnormal
orf19.6148	normal	normal	normal	normal	normal	normal	normal	normal	normal
orf19.1964	normal	normal	n/a	normal	normal	n/a	normal	normal	n/a
orf19.7455	normal	normal	n/a	normal	normal	n/a	normal	normal	n/a
orf19.2317	normal	normal	n/a	normal	normal	n/a	normal	normal	n/a
CBP1	normal	normal	n/a	normal	normal	n/a	normal	normal	n/a
PGA34	normal	normal	n/a	normal	normal	n/a	normal	normal	n/a
IFA14	normal	normal	n/a	normal	normal	n/a	normal	normal	n/a
orf19.2457	normal	normal	n/a	normal	normal	n/a	normal	normal	n/a
HWP1	normal	normal	n/a	normal	normal	n/a	normal	normal	n/a
orf19.3621	normal	normal	n/a	normal	normal	n/a	normal	normal	n/a
orf19.217	normal	normal	n/a	normal	normal	n/a	normal	normal	n/a
PTP3	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
SAP5	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
SAP6	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Chapter 6: Concluding Discussion

While laboratory strains have been vital to the study of *C. albicans* pathogenesis and biology, their use has proceeded with the tacit understanding that laboratory adapted strains do not represent the full breadth of diversity within the species. Until recently (ca. 2015), manipulation of clinical isolates proved a daunting task, constrained by slow methodology and a lack of reliable markers. The development of CRISPR-Cas9 techniques, two of which have been described here, greatly simplified the task of genetic manipulation and allowed us to thoroughly analyze natural variation of virulence in this clinically relevant pathogen. Here we have shown that regulatory networks are highly variable between members of the same species, and that both regulatory differences and similarities may have functional impact on phenotypic outcomes.

Regulatory Relationships

Core targets conserved across strains were found to be highly enriched for annotated functional targets. Furthermore, we examined uncharacterized core targets requiring both Brg1 and Efg1 and identified several genes with novel functional roles in filamentation. As the phenotypic impact on *C. albicans* filamentation and biofilm formation of an *efg1* mutation is greater than a *brg1* mutation, it may also be likely that other core targets required for biofilm formation exist among genes dependent solely on Efg1. In either case, the evolutionary conservation of interactions between master regulators and functional core targets raises several interesting questions. The evolutionary history of a particular strain reflects the environmental pressures encountered throughout its genealogy, and *C. albicans* strains may inhabit a number of

niches in the human body, each specifying different stressors and nutrient conditions. Passage through antibiotic-treated mouse GI tract selects against strains competent in virulence and filamentation, allowing *flo8, efg1,* or *ume6* mutants to outcompete (1–3). However in a healthy gut, normal gut flora may exert selective pressures to maintain capacity for virulence and filamentation (1). What selective pressures act on, and what mechanisms maintain regulatory relationships between master regulators and core targets across dynamic, shifting environments?

One hypothesis is that master regulator binding sites upstream of core targets differ qualitatively from master regulator binding sites upstream of subset specific targets. Under this model, a point mutation in the former may not significantly affect binding affinity, whereas a point mutation in the latter may. Binding sites upstream of core targets may contain optimal motifs allowing the strongest possible transcription factor binding. A SNP in an optimal motif may still allow strong transcription factor binding, but a SNP in a sub-optimal motif may abolish transcription factor binding. ChIP-seq analysis of master regulator binding will reveal if binding motifs calculated from peaks upstream of core targets differ from motifs calculated from peaks upstream of specific/subset targets. If distinct motif populations exist, future directions may include an analysis of each population of binding motifs by comparison of ChIP-seq peak heights or assays of binding affinity using convenient reporter constructs. A non-exclusive, second possibility may be that recruitment by cofactors or other transcription factors binding at adjacent sites may overcome decreased affinity for a mutated binding site. Although direct protein-protein interactions between master biofilm regulators has not been examined in detail, promoters of hyphal associated genes are often bound by multiple biofilm

regulators (4). Analysis of motifs adjacent to ChIP-seq peaks upstream of core targets can also reveal possible binding partners.

Functional targets were also identified among strain/subset specific targets for both regulators with uniform and variable mutant phenotypes (e.g. *ALS1* dependence on Brg1, *BRG1* dependence on Bcr1). Similarly, while hyphal gene expression in different environments converged upon core targets, some functional targets were still identified in temporally and environmentally distinct portions of filamentation responses (5,6). The presence of these functional targets in the "optional" regulatory regimes of regulators with uniform phenotypic impact raises the question: what are the minimum criteria both sufficient for a *C. albicans* strain to form a biofilm? One simplified model is that some combination of core network genes may be both necessary and sufficient for biofilm formation while "optional" relationships may modulate overall phenotypic outcome and are insufficient for biofilm formation. This model would further suggest that circuit diversification affects phenotypic variation through interactions with "optional" targets.

Alternatively, it may be that "optional" targets may also be sufficient for biofilm formation with some combination of core targets. This conceptually more complex model is supported by findings that protein abundance is under greater selective pressure than mRNA abundance (7). Selective advantages for biofilm formation may only exert pressure on expression of any basic combination of biofilm effector proteins, and not a specific "core" set.

To test these models, future experiments may examine if heterologous expression of core network genes on an artificial chromosome in *S. cerevisiae* drives filamentation or

biofilm formation. We may also determine if engineered constitutive expression of all or some subset of core network genes drives filamentation and biofilm formation under non-inducing conditions. Alternatively, if CRISPR/dCas9 epigenome editing is adapted for *C. albicans*, targeted activation of subsets of genes may be achieved by the use of multiplex sgRNAs (8,9). The minimum set of necessary and sufficient targets may then be determined by a library approach using all possible combinations of sgRNAs. This approach may additionally be employed to examine if any combination of core and "optional" targets is sufficient for biofilm formation. If a minimum sufficient set of targets may be identified, we may ask if base substitution rates across isolates are lower for these specific targets compared to core targets or optional targets and identify how selective pressures influence networks.

Defining Core Networks

Should P75010 be included in the filter sets defining a core network? The wild type P75010 strain is the most defective in biofilm formation, and will only form intermittent hyphae in our acute hyphal induction assays. We present two arguments to support the inclusion of P75010. First, the regulatory targets of the P75010 biofilm regulatory network are not an outlier in comparison to the other isolates. The P75010 network is roughly commensurate in size with the other regulatory networks, and exclusion of P75010 does not greatly change the proportion of any master regulator's core targets. Efg1 regulates 697 targets in P75010, and Brg1 regulates 579 targets in P75010. The P75010 Brg1 and Efg1 regulons intersect with other backgrounds fairly evenly (see Chapter 5, Figure 2; Chapter 4, Figure 6B).

Second, the use of P75010 affords an opportunity to filter out targets whose differential expression stems from expression changes that correlate solely with filamentous growth. As the P75010 wild type forms predominantly yeast cells, the shift between wild type and mutant shows the least change in yeast to hypha ratio, providing dynamic range to identify what target genes solely correlate with morphology. Furthermore, on a conceptual level, while P75010 has low filamentation capacity, it is still a bona fide bloodstream isolate, and is in fact more virulent in a mouse model of disseminated candidiasis than P87, an oral candidiasis isolate (10). Lastly, one may argue that any strain that contains functional alleles of a transcription factor should be considered in any core network analyses, as by definition, core regulatory targets should be those present in all strains. Each regulator clearly was still functional in the P75010 background. While the P75010 wild type is the weakest filamenter, intermittent hyphae are still observed. When each master regulator mutant was examined, the mutants always entirely failed to form hyphae.

One argument against the inclusion of P75010 in the core network is that P75010 may carry variant alleles of *BRG1* and *EFG1*. In Chapter 4, we validated our *brg1* and *efg1* mutants in the P75010 background by reintroducing a copy of the corresponding SC5314 *BRG1* or *EFG1* allele. The P75010 validation strains carrying SC5314 alleles were more strongly filamentous as compared to the P75010 wild type, suggesting that the SC5314 alleles we introduced contain cis-elements that may affect *BRG1* or *EFG1* expression or function (See Chapter 4, S4 Figure). Interestingly, we could not identify any SNPs specific to the P75010 *BRG1* or *EFG1* alleles. $BRG1^{P75010}$ and $EFG1^{P75010}$ alleles shared SNPs with at least one allele in each other examined clinical isolate

background. While we did not examine the relative expression levels of specific *BRG1* alleles and the differences in allele strength still require validation, this finding may suggest that the relevant binding sites upstream of functional targets of *BRG1* or *EFG1* are clearly intact in P75010. Differentially expressed genes between strains carrying alleles from different sources may lend insight into how regulatory variation affects phenotypic diversity.

Circuit Diversification

Our primary criteria for circuit diversification was if a target gene was differentially expressed in one background, but not (or oppositely differentially expressed) in another. One aspect of regulatory variation we did not consider was relative differences in the magnitude of differential expression, i.e. if a target gene's expression increases by 3 fold in background A, but by 9 fold in background B. These differences may reflect additional differences in interactions with other regulatory mechanisms, even for genes we identified as shared or "core" targets.

Herein we have examined circuit diversification on the level of transcriptional response, but circuit diversification may affect regulatory networks at post-transcriptional levels. Differences in mRNA abundance may often agree with differences in protein abundance. While a consensus in the literature does not exist, a landmark study by Albert and colleagues showed that more than half of loci which affected mRNA expression levels similarly affected protein abundance, highlighting general correspondence between expression QTL and protein QTLs (11). Furthermore, while transcriptional variation appears to be primarily governed by *trans* acting QTL, a study

by McManus and colleagues identified in a comparison between *S. cerevisiae* and *Saccharomyces paradoxus*, divergence in translational efficiency was primarily governed by *cis*-regulatory differnces (12). Herein we have identified both *cis*-regulatory (differences in allele strength, such as between P75010 and SC5314 *BRG1* and *EFG1* alleles), as well as *trans*-regulatory elements (variable dependence of *BCR1* on Brg1) that affect filamentation or biofilm outcomes. Future studies to address these additional mechanisms of circuit diversification will provide greater resolution on the scope and nature of regulatory variation in *C. albicans*.

LITERATURE CITED, CHAPTER 6

1. Tso GHW, Reales-Calderon JA, Tan ASM, Sem X, Le GTT, Tan TG, et al. Experimental evolution of a fungal pathogen into a gut symbiont. Science. 2018;362(6414):589–95.
2. Desai J V, Lionakis MS. Setting Up Home: Fungal Rules of Commensalism in the Mammalian Gut. Cell Host Microbe. 2019;25(3):347–9.
3. Witchley JN, Penumetcha P, Abon N V, Woolford CA, Mitchell AP, Noble SM. Candida albicans Morphogenesis Programs Control the Balance between Gut Commensalism and Invasive Infection. Cell Host Microbe. 2019 Mar 13;25(3):432-443.e6.
4. Nobile CJ, Fox EP, Nett JE, Sorrells TR, Mitrovich QM, Hernday AD, et al. A recently evolved transcriptional network controls biofilm development in Candida albicans. Cell. 2012;148(1–2):126–38.
5. Martin R, Albrecht-Eckardt D, Brunke S, Hube B, Hünniger K, Kurzai O. A core filamentation response network in Candida albicans is restricted to eight genes. PLoS One. 2013;8(3):e58613.
6. Azadmanesh J, Gowen AM, Creger PE, Schafer ND, Blankenship JR. Filamentation Involves Two Overlapping, but Distinct, Programs of Filamentation in the Pathogenic Fungus Candida albicans. G3 (Bethesda). 2017;7(11):3797–808.
7. Laurent JM, Vogel C, Kwon T, Craig SA, Boutz DR, Huse HK, et al. Protein abundances are more conserved than mRNA abundances across diverse taxa. Proteomics. 2010 Dec;10(23):4209–12.
8. Braun SMG, Kirkland JG, Chory EJ, Husmann D, Calarco JP, Crabtree GR. Rapid and reversible epigenome editing by endogenous chromatin regulators. Nat Commun. 2017;8(1):560.
9. Pulecio J, Verma N, Mejía-Ramírez E, Huangfu D, Raya A. CRISPR/Cas9-Based Engineering of the Epigenome. Cell Stem Cell. 2017;21(4):431–47.
10. Wu W, Lockhart SR, Pujol C, Srikantha T, Soll DR. Heterozygosity of genes on the sex chromosome regulates Candida albicans virulence. Mol Microbiol. 2007 Jun;64(6):1587–604.
11. Albert FW, Treusch S, Shockley AH, Bloom JS, Kruglyak L. Genetics of single-cell protein abundance variation in large yeast populations. Nature. 2014 Feb 27;506(7489):494–7.
12. McManus CJ, May GE, Spealman P, Shteyman A. Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. Genome Res. 2014 Mar;24(3):422–30.