

# A Recently Evolved Transcriptional Network Controls Biofilm Development in *Candida albicans*

Clarissa J. Nobile,<sup>1,\*</sup> Emily P. Fox,<sup>1,2</sup> Jeniel E. Nett,<sup>3</sup> Trevor R. Sorrells,<sup>1,2</sup> Quinn M. Mitrovich,<sup>1,5</sup> Aaron D. Hernday,<sup>1,5</sup> Brian B. Tuch,<sup>1,4</sup> David R. Andes,<sup>3</sup> and Alexander D. Johnson<sup>1</sup>

<sup>1</sup>Department of Microbiology and Immunology

<sup>2</sup>Tetrad Program, Department of Biochemistry and Biophysics

University of California, San Francisco, San Francisco, CA 94102, USA

<sup>3</sup>Department of Medicine, University of Wisconsin, Madison, WI 53706, USA

<sup>4</sup>Present address: Genome Analysis Unit, Amgen, South San Francisco, CA 94080, USA

<sup>5</sup>Present address: Amryis, Emeryville, CA 94608, USA

\*Correspondence: clarissa.nobile@ucsf.edu

DOI 10.1016/j.cell.2011.10.048

## SUMMARY

A biofilm is an organized, resilient group of microbes in which individual cells acquire properties, such as drug resistance, that are distinct from those observed in suspension cultures. Here, we describe and analyze the transcriptional network controlling biofilm formation in the pathogenic yeast *Candida albicans*, whose biofilms are a major source of medical device-associated infections. We have combined genetic screens, genome-wide approaches, and two in vivo animal models to describe a master circuit controlling biofilm formation, composed of six transcription regulators that form a tightly woven network with ~1,000 target genes. Evolutionary analysis indicates that the biofilm network has rapidly evolved: genes in the biofilm circuit are significantly weighted toward genes that arose relatively recently with ancient genes being underrepresented. This circuit provides a framework for understanding many aspects of biofilm formation by *C. albicans* in a mammalian host. It also provides insights into how complex cell behaviors can arise from the evolution of transcription circuits.

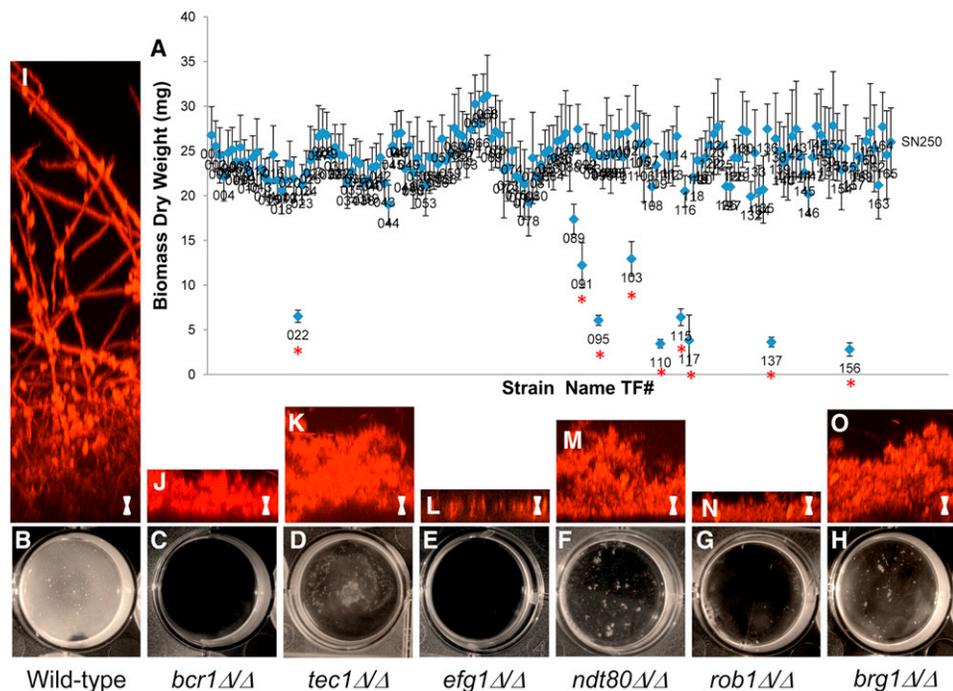
## INTRODUCTION

Biofilms are organized communities of surface-associated micro-organisms embedded in a matrix of extracellular polymers. In this paper, we analyze how *C. albicans*, the predominant fungal pathogen of humans, forms biofilms. Biofilms are a major microbial growth form in natural environments (Kolter and Greenberg, 2006) and a leading cause of persistent human infection (Costerton et al., 1999). These infections are typically seeded from biofilms that form on implanted medical devices,

such as intravascular catheters, and become resistant to drug and mechanical treatments (Donlan and Costerton, 2002). The mechanisms behind biofilm development are thus important to our understanding of microbial ecology (because mixed species biofilms are common) as well as infectious disease.

*C. albicans* biofilm formation can be partitioned into four basic stages, based on studies carried out in vitro (Baillie and Douglas, 1999; Chandra et al., 2001; Douglas, 2003; Hawser and Douglas, 1994; Nobile et al., 2009; Uppuluri et al., 2010a, 2010b). These are: (1) attachment and colonization of yeast-form (nearly spherical) cells to a surface, (2) growth and proliferation of yeast-form cells to allow formation of a basal layer of anchoring microcolonies, (3) growth of pseudohyphae (ellipsoid cells joined end to end) and extensive hyphae (chains of cylindrical cells) concomitant with the production of extracellular matrix material, and (4) dispersal of yeast-form cells from the biofilm to seed new sites. At least some of these features of biofilm formation have also been observed in vivo. For example, *C. albicans* biofilms from denture stomatitis patients confirm the presence of yeast, hyphae, and extracellular matrix (Ramage et al., 2004). Furthermore, biofilm architectures in two animal catheter models and a denture model include numerous yeast cells in the basal region, as well as hyphae and extracellular matrix extending throughout the biofilm (Andes et al., 2004; Nett et al., 2010; Schinabeck et al., 2004).

Here, we combine “classical” genetics, genome-wide approaches, RNA deep sequencing technology, and two in vivo animal models to comprehensively map the transcriptional circuitry controlling biofilm formation in *C. albicans*. The circuit has led to many new predictions about genes involved in biofilm formation, and we have validated a set of these predictions by confirming the roles of several of these genes in biofilm development. The circuit also provides insight into how biofilm formation may have evolved in the *C. albicans* lineage.



**Figure 1. Screening and Characterization of In Vitro Biofilm-Defective Transcription Regulator Mutants**

(A) Biofilm biomass (dry weight) determinations of the entire transcription regulator (TR) mutant library (165 strains). The average total biomass  $\pm$  standard deviation for each TR mutant strain grown under standard biofilm conditions (Experimental Procedures) was calculated from five independent samples of each strain. Statistical significance ( $p$  values) was calculated with a Student's one-tailed paired  $t$  test and is represented by the red asterisk under the nine regulator strains (TF022, TF091, TF095, TF103, TF110, TF115, TF117, TF137, and TF156) with biomasses significantly deviating ( $p < 0.0005$ ) from the reference strain (SN250).

(B–O) Phenotypic characterization of the mutants compared to the wild-type.

(B–H) The visual appearance after 48 hr of growth on polystyrene plates.

(I–O) CSLM side view images of the wild-type and six biofilm-defective mutant strains.

Scale bars represent 20  $\mu$ m. See also Table S1 and Figures S1, S2, and S3.

## RESULTS

### Identification and Phenotypic Characterization of Biofilm-Defective Transcription Regulator Mutants In Vitro

Transcription regulators (defined here as sequence-specific DNA-binding proteins that regulate transcription) play important roles in the control of many developmental pathways; often, they define a group of coregulated target genes that function together to carry out a specific function in the cell. Hence, transcription regulators represent a powerful entry point to understanding a biological process. Using information on transcription regulators taken from a wide variety of species, we constructed a *C. albicans* library of 165 fully vetted transcription regulator (TR) deletion mutants consisting of two independently constructed mutants for each strain (Homann et al., 2009). This library was screened for biofilm formation on the surface of serum-treated polystyrene plates under a standard set of biofilm-inducing conditions (Nobile et al., 2006a, 2006b; Nobile and Mitchell, 2005). The screening was based on biofilm dry weight biomass, visual, and microscopic (confocal) inspection (Figure 1). The screen revealed nine mutants with deficiencies in forming biofilms (Figure 1A and Table S1 and Figure S2A avail-

able online). Three of these mutants were not analyzed further because they exhibited either general growth defects in suspension cultures or a wide variety of other phenotypes in suspension cultures (Extended Experimental Procedures). The remaining six transcription regulator deletion mutants (*bcr1* $\Delta/\Delta$ , *tec1* $\Delta/\Delta$ , *efg1* $\Delta/\Delta$ , *ndt80* $\Delta/\Delta$ , *rob1* $\Delta/\Delta$ , and *brg1* $\Delta/\Delta$ ) have the following characteristics: (1) they were significantly compromised in biofilm formation compared to the wild-type ( $p < 0.0005$ ) (Figures 1B–1H), (2) they did not exhibit general growth defects, and (3) they did not show extensive phenotypes aside from defects in biofilm formation. Of these six transcription regulators, three are newly identified as biofilm regulators (Ndt80/Orf19.2119, Rob1/Orf19.4998 [named for regulator of biofilms], and Brg1/Orf19.4056 [named for biofilm regulator]), and three had been previously implicated in biofilm formation (Bcr1 [Nobile and Mitchell, 2005], Tec1 [Nobile and Mitchell, 2005], and Efg1 [Ramage et al., 2002]). The screen was carried out blindly, and our identification of all previously identified regulators serves as an internal control for both the library construction and the screen.

We further characterized the morphology of the six biofilm-defective mutant strains compared to the wild-type by confocal scanning laser microscopy (CSLM), using silicone squares as the substrate (Figures 1I–1O). By CSLM, the wild-type reference

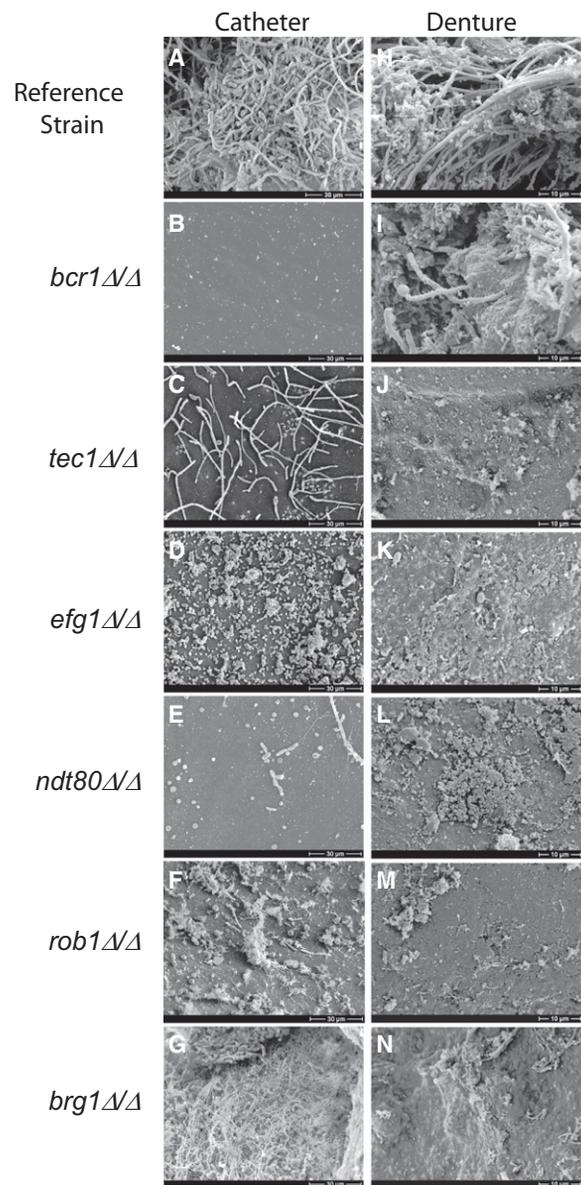
strain formed a biofilm with typical architecture and thickness (Chandra et al., 2001; Douglas, 2003; Nobile and Mitchell, 2005) of  $\sim 250$   $\mu\text{m}$  in depth, containing both round budding yeast-form cells adjacent to the substrate and hyphal cells extending throughout the biofilm (Figure 1I) (see also Figure S1 for CSLM visualization of each regulator mutant over a time course of biofilm development). In all six mutants, only rudimentary biofilms of  $\sim 20$ – $80$   $\mu\text{m}$  in depth were formed, although the detailed phenotypes of the mutants differ (Figures 1J–1O and S1). Reintroduction of an ectopic copy of the wild-type allele back into each mutant reversed the biofilm formation defect of each mutant (Figure S2B). Thus, *BCR1*, *TEC1*, *EFG1*, *NDT80*, *ROB1*, and *BRG1* are required for wild-type biofilm formation in vitro.

Because hyphal development is an important step in normal biofilm development, we assessed the ability of our six biofilm-defective transcription regulator mutants to form normal hyphae when they were not in the context of a biofilm. We found that, with the exception of the *efg1* $\Delta/\Delta$  strain, true hyphae could be detected in the medium surrounding the biofilm (Figure S3A) as well as in suspension cultures using the same medium as that used for biofilm formation (Figure S3B). We also observed hyphal development for all strains except the *efg1* $\Delta/\Delta$  strain in a variety of suspension culture media, although the fraction of hyphal cells was often reduced relative to the parental strain (Figure S3B). Thus, for all of these mutants (with the possible exception of *efg1* $\Delta/\Delta$ ), the defect in biofilm formation was not due to an intrinsic inability to form hyphae.

### Characterization of Biofilm-Defective Transcription Regulator Mutants in Two In Vivo Animal Models

Biofilm formation in vivo is the cause of the majority of new infections in humans, and it is widely appreciated that the conditions for biofilm formation in vivo differ considerably from those in standard in vitro assays (Nett and Andes, 2006). For example, many additional elements are present in vivo, such as liquid flow, host factors, and components of the host immune response. Because biofilm-based catheter infections are a major clinical problem (Kojic and Darouiche, 2004), we used a well-established rat venous catheter model of infection (Andes et al., 2004) to test the six mutants for biofilm formation in vivo. We inoculated the catheters with *C. albicans* cells intraluminally, allowed biofilm formation to proceed for 24 hr, removed the catheters, and visualized the catheter luminal surfaces by scanning electron microscopy (SEM) (Figures 2A–2G and S4A). The wild-type reference strain formed a thick, mature biofilm on the rat catheter, consisting of yeast and hyphal cells and extracellular matrix material (Figure 2A). Of the six transcription regulator mutants, five (*bcr1* $\Delta/\Delta$ , *tec1* $\Delta/\Delta$ , *efg1* $\Delta/\Delta$ , *ndt80* $\Delta/\Delta$ , and *rob1* $\Delta/\Delta$ ) were unable to form biofilms (Figures 2B–2F); *bcr1* $\Delta/\Delta$  had been previously shown to be defective in this model (Nobile et al., 2008). The sixth mutant (*brg1* $\Delta/\Delta$ ) formed a thick biofilm consisting of many adherent cells and a large amount of extracellular matrix material (Figure 2G) but appeared morphologically distinct from the reference strain in that considerably fewer hyphae were observed within the biofilm (compare Figures 2A and 2G).

The most common form of oral candidiasis is denture stomatitis, prevalent largely in the elderly population, and affecting up



**Figure 2. Biofilm Formation in Two In Vivo Rat Models: A Catheter Model and a Denture Model**

(A–G) For the catheter model, the wild-type reference strain SN250 (A) and the six transcription regulator mutant strains (panels B–G) were inoculated into rat intravenous catheters; resulting biofilms were visualized after 24 hr of growth by scanning electron microscopy (SEM). SEM catheter images show the catheter luminal surfaces at magnifications of 1,000 $\times$ .

(H–N) For the denture model, the wild-type reference strain SN425 (H) and the six transcription regulator mutant strains (I–N) were inoculated into rat dentures, and the resulting biofilms were visualized after 24 hr of growth by SEM. SEM denture images show the denture surfaces at magnifications of 2,000 $\times$ .

See also Figure S4.

to 70% of denture wearers (Webb et al., 1998; Wilson, 1998). Denture stomatitis occurs by biofilm colonization and growth over the surface of a denture, leading to inflammation of the palatal mucosa (Ramage et al., 2004). Because biofilm growth

on dentures represents a completely different host environment from that of an intravenous catheter, we also screened our six biofilm-defective regulator mutants in a recently established *in vivo* rat denture model, which was developed to mimic and assess *C. albicans* biofilm formation in denture stomatitis (Nett et al., 2010). In particular, this oral model includes host salivary components, host commensal bacteria, salivary flow dynamics, and direct contact between the denture biofilm and the host mucosal surface (Nett and Andes, 2006). We inoculated the rat dentures with *C. albicans* cells, permitted biofilm formation to proceed for 24 hr, removed the dentures, and visualized the denture surfaces by SEM (Figures 2H–2N). The wild-type reference strain formed a thick, mature biofilm on the surface of the rat denture, consisting predominantly of hyphal *C. albicans* cells interspersed with *C. albicans* yeast-form cells, various host commensal oral bacteria, and extracellular matrix material (Figure 2H). In contrast, the genetically matched mutant strains all showed significant defects in biofilm formation. In particular, *tec1Δ/Δ*, *efg1Δ/Δ*, *ndt80Δ/Δ*, *rob1Δ/Δ*, and *brg1Δ/Δ* were severely defective (Figures 2J–2N), whereas the *bcr1Δ/Δ* mutant, which has previously been shown to be defective in this model (Nett et al., 2010), had less pronounced defects than the other five mutants (Figure 2I). We note that extensive bacterial biofilms consisting of both cocci and rods were seen on the dentures of the six *C. albicans* biofilm-defective mutants (Figure S4B), suggesting a competition between biofilm formation by *C. albicans* and biofilm formation by the native bacteria present in the mouth.

In summary, *BCR1*, *TEC1*, *EFG1*, *NDT80*, *ROB1*, and *BRG1* are each required for normal biofilm formation *in vivo* in both the rat denture and catheter models. The effects of certain deletion mutants (*brg1Δ/Δ* and *bcr1Δ/Δ*) differed to varying degrees between the two models (compare Figures 2G with 2N and 2B with 2I), likely reflecting the influence of the host environment in biofilm formation. The results, taken as a whole, indicate that performing genetic screens and analyzing biofilm formation *in vitro* is a valid approach to understanding clinically relevant *C. albicans* biofilm formation.

### Developing Transcriptional Relationships among Biofilm Regulators

To identify genes directly regulated by Bcr1, Tec1, Efg1, Ndt80, Rob1, and Brg1, we performed full-genome chromatin immunoprecipitation microarray (ChIP-chip) to map the position across the genome to which each of the six transcription regulators is bound during biofilm formation. Based on this analysis (see Extended Experimental Procedures for details, Tables S2A–S2F for a complete list of all significantly bound locations for each regulator, and Data S1 for MochiView image plots of every called significant peak for each regulator), we calculate the following number of intergenic regions bound by each regulator: 211 for Bcr1, 76 for Tec1, 328 for Efg1, 558 for Ndt80, 95 for Rob1, and 283 for Brg1 (Table S2G). 831 intergenic regions are bound by one or more regulators, 350 intergenic regions are bound by two or more, 186 intergenic regions are bound by three or more, 111 intergenic regions are bound by four or more, 55 intergenic regions are bound by five or more, and 18 intergenic regions are bound by all six of the biofilm regulators

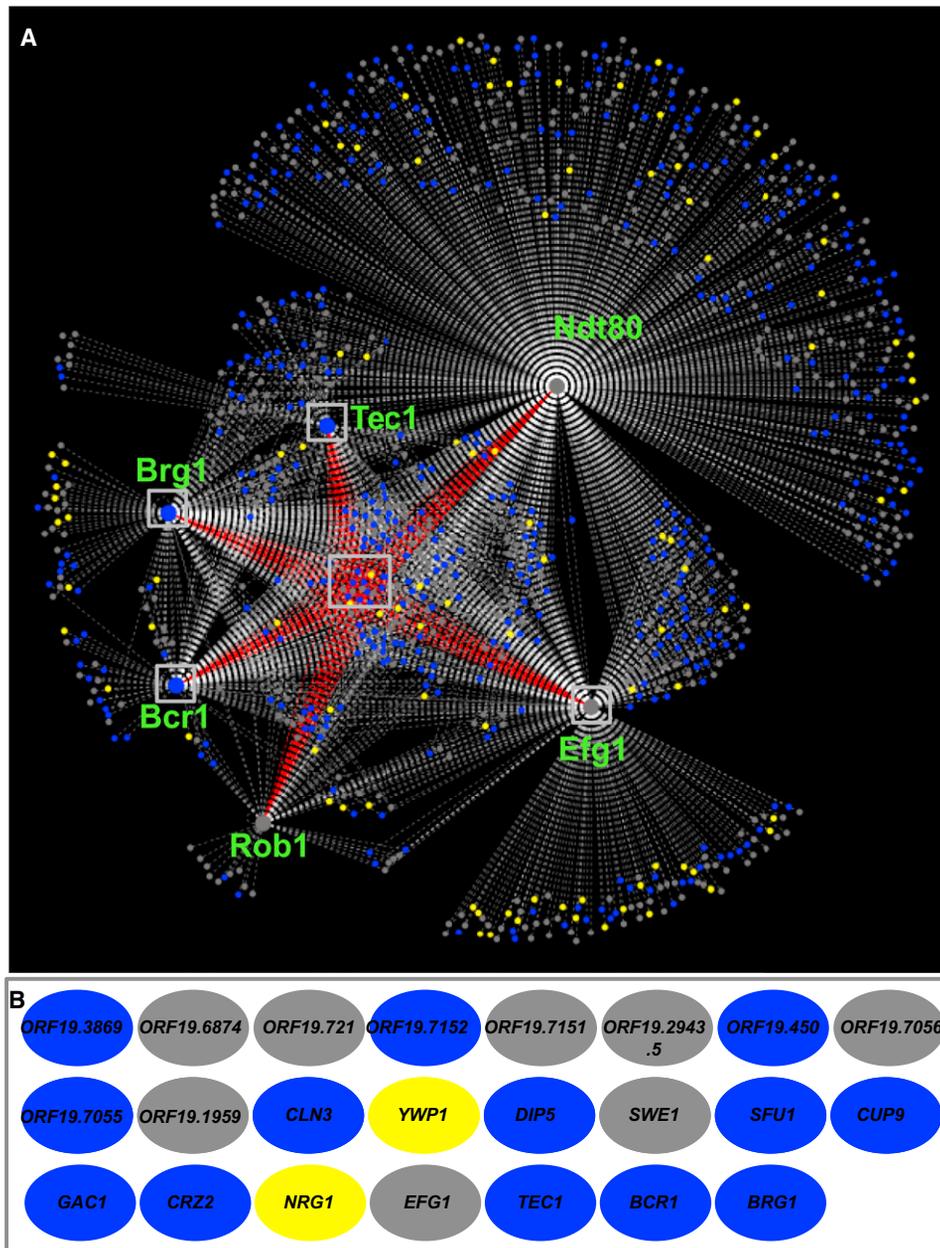
(Table S2G). We noticed two unusual characteristics for the intergenic regions bound by the biofilm regulators. First, the average length of intergenic regions bound by the biofilm regulators is more than twice that of the remainder of the genome (1540 bp compared with 693 bp); this trend is true for all six biofilm regulators (Table S4F). Second, binding peaks are distributed throughout the intergenic regions of the regulator-bound target genes rather than being clustered a fixed distance upstream of the transcription start site (Data S2), as is common for many yeast target genes (Lin et al., 2010).

If we convert bound intergenic regions to genes likely to be controlled (for example, a single bound intergenic region between divergently transcribed genes is counted as two genes), our analysis suggests that the network is composed of 1,061 target genes that are bound in their promoter regions by at least one of the six biofilm regulators (Figure 3 and Table S3A). This regulatory network is shown in Figure 3. Based on the ChIP-chip data, the high degree of overlap between target genes among biofilm regulators suggests that the biofilm regulatory network is considerably interwoven; that is, many of the target genes are controlled by more than one regulator.

The results also indicate that the six regulators originally identified in the genetic screen control each other's expression: all six of the regulators bind to the upstream promoter regions of *BCR1* (Figure 4A), *TEC1* (Figure 4B), *EFG1* (Figure 4C), and *BRG1* (Figure 4F); four of the regulators (Tec1, Efg1, Ndt80, and Rob1) bind to the upstream promoter region of *ROB1* (Figure 4E); and two of the regulators (Efg1 and Ndt80) bind to the upstream promoter region of *NDT80* (Figure 4D).

### De Novo Motif Finding for the Six Master Biofilm Regulators

A test of the self-consistency of ChIP-chip data is the nonrandom occurrence of *cis*-regulatory sequences (motifs). Based on several hundred significant binding events from our ChIP-chip data, we were able to identify statistically significant motifs for all six of the biofilm regulators (Figure 4G, Data S2, and Tables S2H–S2M). This motif generation was based solely on the ChIP-chip data and did not incorporate data from any other experiment or from any other species. We note that the motif generated for Ndt80 (TTACACAAA) is very similar to the reported binding motif for its homolog, Ndt80, in *S. cerevisiae* (GMCACAAA) (Zhu et al., 2009). The motif for Tec1 (RCATTCY) is identical to that determined for its homolog, Tec1, in *S. cerevisiae* (Harbison et al., 2004; Madhani and Fink, 1997). (This Tec1 motif, generated from 107 bound intergenic regions, does not closely resemble the Tec1 motif recently reported in the white-specific pheromone response element [WPRE] [AAAAAAAAAAGAAAG] in *C. albicans*, which was generated from a much smaller set of data [Sahni et al., 2010].) Finally, the Efg1 motif derived from our ChIP-chip data (RTGCATRW) closely resembles the TGCAGNNA consensus sequence of the *S. cerevisiae* ortholog, Sok2 (Harbison et al., 2004). Thus, for three of the biofilm regulators, the motifs developed from our *C. albicans* ChIP-chip data can be independently verified by their similarities to the motifs recognized by their *S. cerevisiae* orthologs. This analysis provides independent support for both the motif analysis and for the validity of the full-genome ChIP data. For the other three regulators, we were able



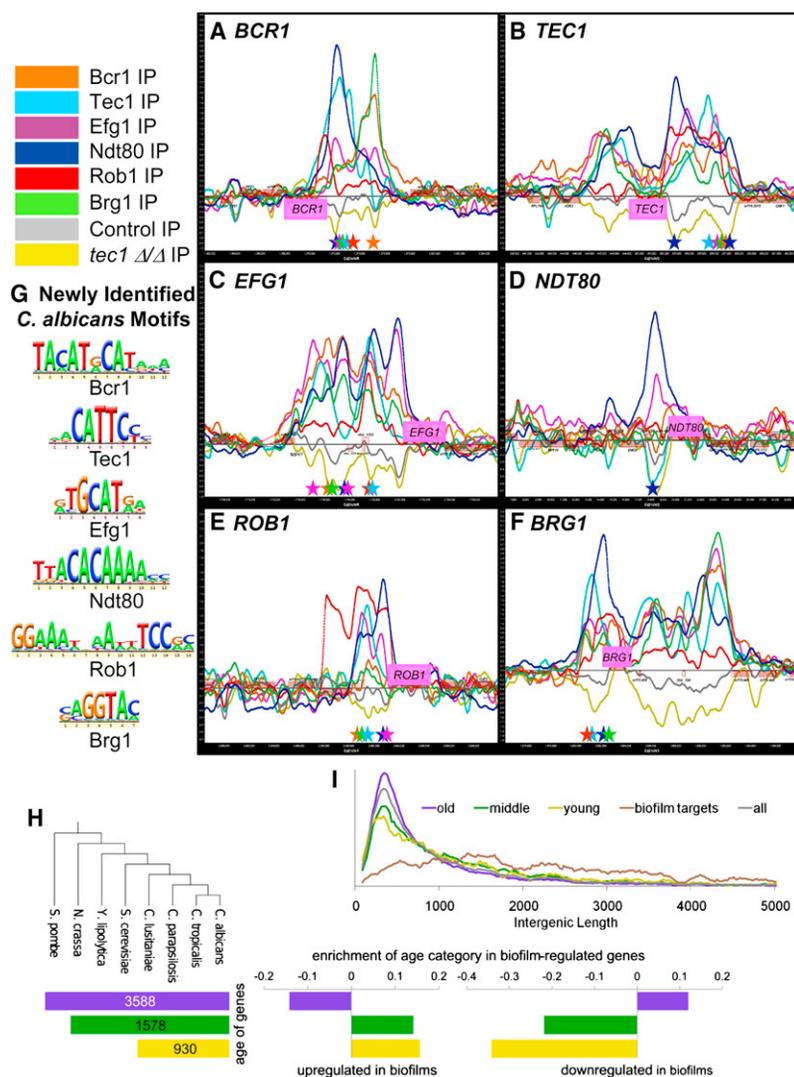
**Figure 3. The Biofilm Regulatory Network**

(A and B) The six master biofilm regulators are represented by the six large circular hubs. Smaller circles represent target genes, which are connected to their respective regulators by dashed lines, indicating a direct interaction as determined by genome-wide ChIP-chip. Genes that are differentially regulated as determined by expression data (using a 2-fold cutoff) in biofilm compared to planktonic cells are shown in blue for those genes upregulated in biofilms, in yellow for those downregulated, and in gray for those with no change. Gray boxes are drawn around the 23 target genes bound by all six regulators and are connected to their respective regulators by red dashed lines (A). The identity of these 23 genes are indicated as the colored ovals in (B) (blue ovals are genes that are upregulated, yellow ovals are genes that are downregulated, and gray ovals are genes with no change in biofilm compared to planktonic cells). Overall, 23 genes are bound by all six, 77 are bound by five or more, 165 are bound by four or more, 265 are bound by three or more, and 458 are bound by two or more of the biofilm regulators. See also Table S3A.

to determine statistically significant motifs, but we were not able to independently verify them by comparison with *S. cerevisiae* because either the orthology relationships are uncertain (Rob1 and Brg1) or the orthologous *S. cerevisiae* regulator has not been characterized (Bcr1).

#### Exploring the Transcriptional Patterns of Biofilms

Although the ChIP-chip experiments reveal the genomic positions where each regulator binds, they do not indicate whether these binding events are associated with differences in gene transcription. We first consider control of the regulators



**Figure 4. Chromatin Immunoprecipitation Mapping and Motif Identification of the Six Master Biofilm Regulators and Evolutionary Analysis of the Biofilm Target Genes**

(A–G) All six regulators bind to each other’s upstream promoter regions (A–F). Immunoprecipitation (IP) binding data for Bcr1-Myc (orange line), Tec1-custom antibody (light blue line), Efg1-Myc (magenta line), Ndt80-Myc (dark blue line), Rob1-Myc (red line), Brg1-Myc (green line), untagged wild-type/control IP (gray line), and *tec1*Δ/Δ (yellow line) strains are shown. The ChIP-chip microarray binding data was mapped and plotted onto the chromosomes containing *BCR1* (A), *TEC1* (B), *EFG1* (C), *NDT80* (D), *ROB1* (E), and *BRG1* (F) using MochiView. The promoters of these genes show significant peak enrichments for the binding of the indicated biofilm regulators. The x axis represents ORF chromosomal locations. The y axis gives the Agilent normalized enrichment value (log<sub>2</sub>) postsmoothing for the binding of each regulator. Genes (pink boxes) plotted above the bold line read in the sense direction; genes plotted below the bold line read in the antisense direction. Using de novo motif finding based on our ChIP-chip data, we identified significantly enriched core binding motifs for all six biofilm regulators (G). Motifs were identified using MochiView and independently verified using MEME, and motif graphics were generated with MochiView. Colored stars corresponding to the colors of the regulators indicate the location of strong instances of the indicated biofilm regulator motifs under the enrichment peaks in A–F.

(H) The evolutionary age of target genes in the biofilm network. Genes were divided into three categories based on when they arose during evolution, with the numbers in each bar giving the number of *C. albicans* genes that fall into that age category (based on the union of RNA-seq and microarray data sets for biofilm versus planktonic cells). The enrichment of each age category in biofilm-regulated genes is log<sub>10</sub> of the observed divided by the expected (for all age categories,  $p < 1.23 \times 10^{-9}$ ).

(I) A histogram of the length of the intergenic regions between tandem and divergent gene pairs targeted by the biofilm regulators. Each category was normalized to the total number of intergenic regions in that category.

See also [Data S1 and S2](#) and [Tables S2, S3, S4, and S5](#).

themselves, as they are all bound by one or more of the other regulators. We deleted each regulator and measured the mRNA levels of the other five (Figure S7A). This analysis revealed that each regulator positively regulates each of the other regulators. We also examined the effect of each regulator on its own synthesis by fusing its upstream region to an mCherry reporter and measuring levels of the reporter in the absence and presence of the regulator (Figure S7B). In all cases, a given regulator activates its own synthesis. Thus, the connections among the six biofilm regulators are primarily, if not exclusively, positive.

To assess the relationship of regulator binding and transcription across the entire circuit, we performed both RNA-seq and gene expression microarray analyses of cells grown in biofilm and planktonic conditions. From our RNA-seq data, we generated 46 million mappable strand-specific sequence reads, expanding our previous gene annotation (Tuch et al., 2010) by identifying 622 “novel transcriptionally active regions” (nTARs) and 161 nTARs that overlap, at least partially, transcribed regions identified in other recent genome-wide experimental annotations

(Bruno et al., 2010; Sellam et al., 2010) (Table S4A). We know from previous work that nTARs identified by RNA-seq include both noncoding RNAs (Mitrovich et al., 2010) and transcripts that encode for proteins too short to have been identified in previous genome annotations (Tuch et al., 2010).

We used our RNA-seq data in addition to our gene expression microarray data to obtain a complete set of genes (coding and noncoding) differentially expressed between planktonic and biofilm conditions (Table S4). Combining the RNA-seq and microarray data, we find 1,599 genes upregulated and 636 genes downregulated at least 2-fold in biofilm compared to planktonic cells (Tables S4B and S4C, respectively). By analyzing the overlap between our ChIP-chip data and our gene expression data (Table S5), we find a strong correlation between transcription regulator binding and differential gene expression. For example, if we consider regions bound by at least four transcription regulators, ~60% of these regions are associated with differentially expressed transcripts. This is significantly greater than that expected by chance ( $p < 0.0001$ ) and suggests, at least broadly,

that binding of the regulators is associated with differential transcription in biofilm versus planktonic cultures. For the correlation between the binding of a given single transcription regulator and differential gene expression, we find a range of 38%–56%, comparable to or greater than the associations documented for other *C. albicans* transcription regulators (Askew et al., 2011; Lavoie et al., 2010; Nobile et al., 2009; Sellam et al., 2009; Tuch et al., 2010).

We examined the evolutionary history of genes that are differentially regulated under biofilm conditions. To do this, we categorized each *C. albicans* gene into an age group based on orthology mappings across the Ascomycota, a large group of yeasts that includes both *C. albicans* and *S. cerevisiae* (Wapinski et al., 2007) (Extended Experimental Procedures). Gene ages were defined using orthology assignments from The Fungal Orthogroups Repository (<http://www.broad.mit.edu/regev/orthogroups/>). The oldest genes are present in distantly related yeast clades, whereas the youngest are found only in *C. albicans*. Young genes can arise in several ways, including relatively rapid mutation that obscures the relation to an ancient gene, horizontal gene transfer, and de novo gene formation (Long et al., 2003). We found that genes upregulated in biofilms are enriched for young and middle-aged genes and are depleted in old genes. The opposite trend was observed for genes that are downregulated in biofilms (Figure 4H). Genes that were not differentially expressed were not strongly enriched for any age category (Table S4E). Young genes typically show longer intergenic regions than old genes (Sugino and Innan, 2011), and this trend may help to explain the unusually long intergenic regions of biofilm circuit genes. However, biofilm genes exhibited significantly longer intergenic regions even when compared to other young genes ( $p < 2.2 \times 10^{-16}$ ) (Figure 4I).

### Identifying Functionally Relevant Target Genes of the Master Biofilm Network

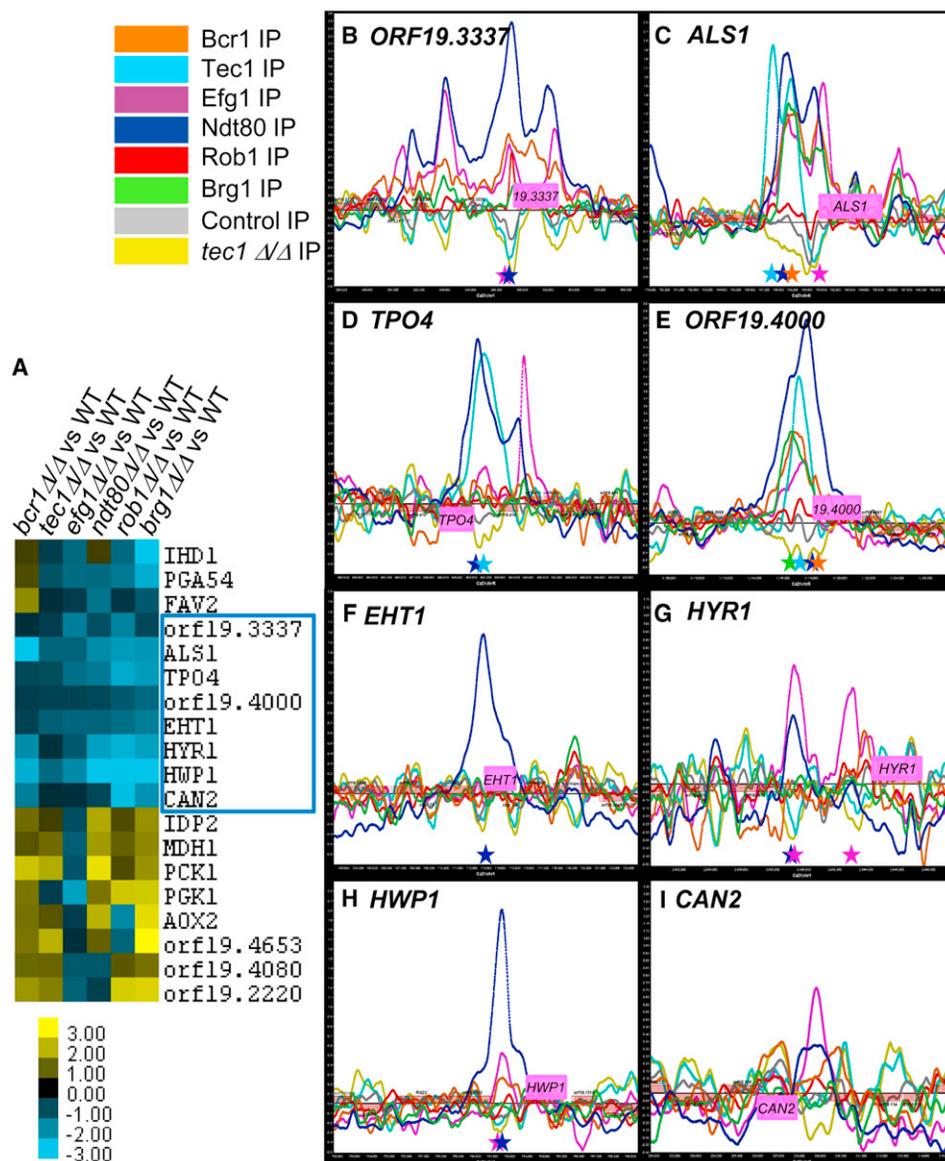
To understand the connections between the six regulators and biofilm development, we performed gene expression microarray experiments of all six regulator mutants compared to a reference strain under biofilm-forming conditions. In interpreting this data, it is important to keep in mind that the mutant strains do not form mature biofilms under these conditions, so many of the transcriptional effects may be indirect consequences of defective biofilms. Consistent with this idea, the transcriptional responses to deletion of each of the biofilm transcription regulators tended to encompass a relatively large set of genes (Table S3A). For example, we found 234 genes that were downregulated and 173 genes that were upregulated in the *bcr1* $\Delta/\Delta$  mutant relative to the isogenic parent (threshold of [ $\log_2 > 0.58$  and  $\log_2 < -0.58$ ]) (Table S3C). Of these genes, Bcr1 binds directly to the promoters of 46 (11%) of them, a number that is significantly higher than that predicted by chance ( $p = 0.0002$ ). Nonetheless, the results indicate that most of the effects of deleting Bcr1 are indirect. Of the genes directly bound by Bcr1, half were downregulated and half were upregulated in the *bcr1* $\Delta/\Delta$  mutant, indicating that Bcr1 can act as both an activator and repressor of its direct target genes. Similar analysis (Table S3C and Extended Experimental Procedures) indicates that Efg1, Ndt80, Rob1, and Brg1 are all both activators and repressors of

their biofilm-relevant direct target genes and that Tec1 is primarily an activator of its biofilm-relevant direct target genes.

From these large data sets, we attempted to identify a set of target genes that might be expected to have important roles in biofilm formation. Using hierarchical cluster analysis to characterize genes with similar patterns of expression in each of the six biofilm regulator mutants compared to a reference strain under biofilm conditions, we found 19 target genes that were differentially regulated in all six data sets (threshold of [ $\log_2 > 0.58$ , and  $\log_2 < -0.58$ ]) (Figure 5A and Table S3A). Eight of these target genes (*ORF19.3337*, *ALS1*, *TPO4*, *ORF19.4000*, *EHT1*, *HYR1*, *HWP1*, and *CAN2*) were expressed at lower levels in all six of the biofilm regulator mutants compared to the reference strain (Figure 5A); seven of these genes were also expressed at higher levels in biofilm compared to planktonic wild-type cells (Table S3A). Additionally, all of these eight target genes were bound in their upstream promoter regions by at least one of the six biofilm regulators; most were bound by multiple regulators (Figures 5B–5I).

Further analysis of the regulation of these eight target genes helps to reconcile their expression patterns with the chromatin IP results. As indicated in Figure S5, the transcriptional effects of deleting each one of the six regulators can be accounted for by: (1) direct binding and transcriptional activation by that regulator on the target gene and/or (2) direct binding and activation of a different regulator that, in turn, binds directly to and activates the target gene (Figure S5). This “hierarchical cascade” between the biofilm regulators and target genes, applied more broadly, can explain much of the expression data (Figure S5, Tables S3A and S3C, and Extended Experimental Procedures).

To determine whether the eight target genes identified by this analysis affected biofilm formation, we constructed homozygous deletion strains for each of the eight target genes. We observed significant biofilm defects for *als1* $\Delta/\Delta$  ( $p = 0.01$ ), *hwp1* $\Delta/\Delta$  ( $p = 0.01$ ), and *can2* $\Delta/\Delta$  ( $p = 0.003$ ) mutant strains compared to the reference strain, with the *can2* $\Delta/\Delta$  strain the most defective (Figure 6A). Although all three of these mutants were capable of forming partial biofilms, these biofilms were less stable than those of the wild-type and often detached from the substrate; partial biofilm defects have been previously reported for *als1* $\Delta/\Delta$  and *hwp1* $\Delta/\Delta$  mutant strains (Nobile et al., 2006a, 2006b; Nobile et al., 2008), whereas *can2* $\Delta/\Delta$  is new to this study. The other five knockout strains did not show any obvious biofilm defects under the conditions tested, and we hypothesized that their roles may be masked by genetic redundancy. To explore this idea, we created ectopic expression strains in which each of the eight target genes was ectopically expressed in strains in which each transcription regulator was deleted. In other words, in a grid of  $6 \times 8 = 48$  constructed strains, we determined whether ectopic expression of the target genes could suppress the defect of the original transcription regulator deletion. Overexpression of several of the candidate target genes was able to significantly rescue biofilm formation to varying degrees depending on the target gene mutant background combination ( $p < 0.0005$ ) (Figure 6B; see Figure S6 for CSLM images of the rescued biofilms). For example, overexpression of *ORF19.4000*, *CAN2*, or *EHT1* in the *bcr1* $\Delta/\Delta$  mutant strain background was able to rescue biofilm formation to near wild-type

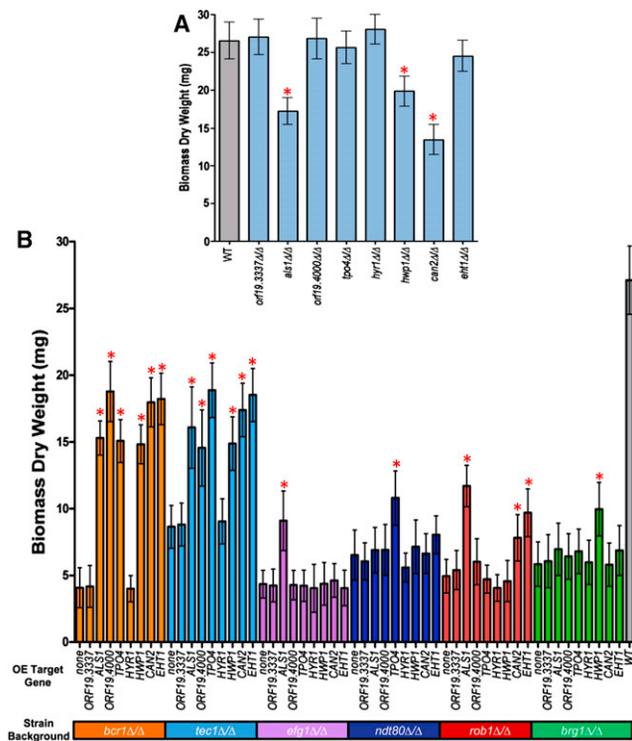


**Figure 5. Core Candidate Biofilm Target Genes**

(A) Using hierarchical cluster analysis of our gene expression microarray data, we identified a set of 19 candidate target genes (*IHD1*, *PGA54*, *FAV2*, *ORF19.3337*, *ALS1*, *TPO4*, *ORF19.4000*, *EHT1*, *HYR1*, *HWP1*, *CAN2/ORF19.111*, *IDP2*, *MDH1*, *PCK1*, *PGK1*, *AOX2*, *ORF19.4653*, *ORF19.4080*, and *ORF19.2220*) that were differentially regulated ( $\log_2 > 0.58$ , and  $\log_2 < -0.58$ ) in all gene expression array experiments that compared each biofilm regulator mutant to a reference strain under biofilm conditions (A). Eight of these targets were differentially regulated in the same direction (all down in the mutants) and were chosen for further functional analyses (A, as indicated by the blue square).

(B–I) ChIP-chip enrichment data for the binding of the six biofilm regulators in the promoters of these eight candidate target genes. IP binding data for Bcr1-Myc (orange line), Tec1 custom antibody (light blue line), Efg1-Myc (magenta line), Ndt80-Myc (dark blue line), Rob1-Myc (red line), Brg1-Myc (green line), untagged wild-type/control IP (gray line), and *tec1*  $\Delta/\Delta$  (yellow line) strains are shown. The ChIP-chip microarray binding data were mapped and plotted onto the chromosomes containing *ORF19.3337* (B), *ALS1* (C), *TPO4* (D), *ORF19.4000* (E), *EHT1* (F), *HYR1* (G), *HWP1* (H), and *CAN2* (I) using MochiView. The promoters of these genes show significant peak enrichments for the binding of the indicated biofilm regulators: *ORF19.3337* by Bcr1, Efg1, Ndt80, and Rob1 (B); *ALS1* by Bcr1, Tec1, Efg1, Ndt80, and Brg1 (C); *TPO4* by Tec1 and Ndt80 (D); *ORF19.4000* by Bcr1, Tec1, Efg1, Ndt80, and Brg1 (E); *EHT1* by Ndt80 (F); *HYR1* by Efg1 (G); *HWP1* by Ndt80 (H); and *CAN2* by Efg1 (I). The x axis represents ORF chromosomal locations. The y axis is the Agilent normalized enrichment value ( $\log_2$ ) postsmoothing for the binding of each regulator. Genes (pink boxes) plotted above the bold line read in the sense direction; genes plotted below the bold line read in the antisense direction. Colored stars corresponding to the colors of the regulators indicate the location of strong instances of the indicated biofilm regulator motifs under the enrichment peaks.

See also Figure S5 and Table S3.



**Figure 6. Functionally Relevant Biofilm Target Genes**

(A and B) Biofilm biomass (dry weight) determinations were measured for the eight core candidate biofilm target gene deletion mutants (A) and the strains in which each of the eight target genes was ectopically expressed in the background of each regulator mutant (B). The average total biomass  $\pm$  standard deviation for each strain grown under standard biofilm conditions was calculated from five independent samples of each strain. Statistical significance ( $p$  values) was calculated with a Student's one-tailed paired  $t$  test and is represented by the red asterisks above the strains with biomasses significantly deviating ( $p < 0.0005$ ) from either the reference strain (WT) for (A) or the corresponding mutant strain for (B). See also Figure S6 and Table S3.

levels of biomass (although the biofilms are fragile) (Figure 6B and Figure S6), implicating these genes in biofilm formation. Taken as a whole, our data suggest that six of the original set of eight candidate target genes have direct roles in biofilm formation. Of course, there are more than 1,000 additional target genes, and their analysis is a future challenge.

## DISCUSSION

### A Master Transcription Network Controlling Biofilm Formation in *C. albicans*

We have described a master circuit of six transcription regulators that controls biofilm formation by *C. albicans* in vitro and in two different animal models. *C. albicans* biofilms are an organized structure of three types of cells (yeast, pseudohyphae, and hyphae) enclosed in an extracellular matrix. The transcription regulators form an elaborate, interconnected transcriptional network: each regulator controls the other five, and most target genes are controlled by more than one master regulator (Figure 3). The circuit appears to be based largely, if not exclusively, on positive regulation (Figures 7, S7A, and S7B). Taking into

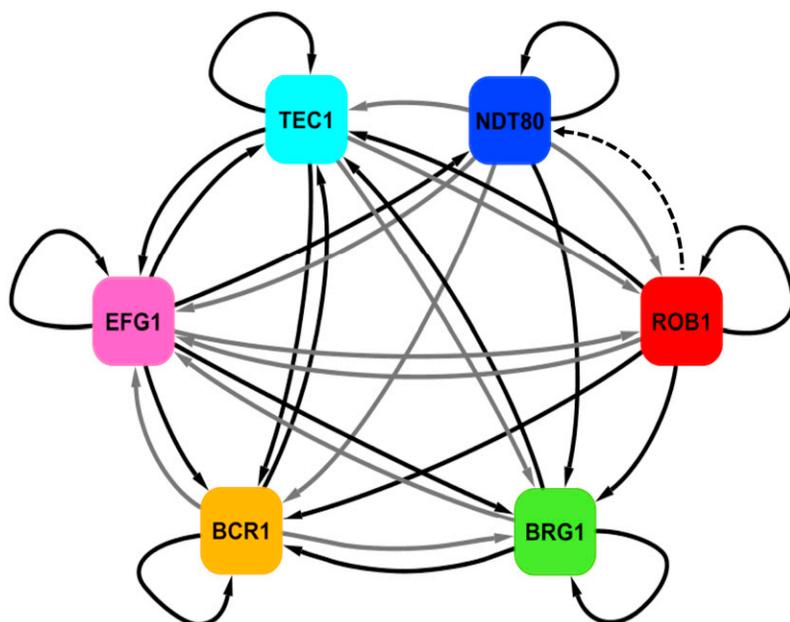
consideration all of the target genes of the six regulators, the biofilm network comprises about 15% of the genes in the genome.

### Circuit Complexity

Although the circuit is large and complex (~1,000 genes and twice that many connections), this level of complexity is not without precedent. For example, circuits that control osmotic stress and pseudohyphal growth pathways of *S. cerevisiae* (Borneman et al., 2006; Ni et al., 2009), competence and spore formation in *Bacillus subtilis* (de Hoon et al., 2010; Hamoen et al., 2003; Losick and Stragier, 1992; Süel et al., 2006), the hematopoietic and embryonic stem cell differentiation pathways of mammals (Wilson et al., 2010; Young, 2011), and the regulation of circadian clock rhythms in *Arabidopsis thaliana* (Alabadi et al., 2001; Locke et al., 2005) show certain similarities: they all consist of a core group of master transcription regulators that control each other and—working together—control a large set of additional target genes.

Several possibilities might account for the complexity of the biofilm network. The regulators we have described can orchestrate biofilm formation in two very different niches of the human host: the bloodstream and the oral cavity. It seems likely that the same circuit also controls biofilm formation in other host niches (for example, in the vagina and gastrointestinal tract). Thus, the biofilm circuit responds to many environmental conditions, such as temperature, nutrient availability, flow rate, surface type, other microbial species, and components of the host immune system. One possibility is that the complex circuit that we have described can integrate a wide range of environmental cues to produce a stereotyped morphological and functional output under many different conditions. Consistent with this idea is the finding that one regulator (Bcr1) plays an important role in biofilm formation in the catheter model but has a less pronounced role in the denture model, whereas another regulator (Brg1) shows the opposite behavior. It is also possible that the complex structure of the network (consisting of many direct and indirect feedback loops, many feed-forward loops, and highly overlapping regulons) is responsible for a form of cell memory that acts over generations to ensure coordinated cooperation among cells in maintaining the biofilm state. A third possibility, as has been suggested for ribosomal protein gene regulation (Müller and Stelling, 2009), is that the more complex the regulatory architecture of a network, the more precisely the dynamics of gene expression can be regulated.

A consideration of the evolution of the biofilm network might also help to explain why it differs structurally from simple regulatory schemes. Incorporation of genes one at a time into a network requires a gain of a binding site upstream of each gene; however, bringing a regulatory protein gene into a network instantly incorporates all of that regulator's targets into the network. Thus, the interconnectedness of the biofilm network may reflect the ease by which many genes can be simultaneously incorporated into an existing circuit. Finally, it is formally possible that the complexity per se of a transcriptional network is not, in itself, adaptive; rather, some aspects of the network complexity could simply be the result of neutral (nonadaptive) evolution (Fernández and Lynch, 2011).



**Figure 7. Regulatory Network Model for Biofilm Formation**

The biofilm network model based on our ChIP-chip and expression data is shown. Solid arrows indicate direct binding interactions determined by our ChIP-chip analysis. Solid black arrows indicate experimentally validated regulatory interactions (as determined by expression profiling data and validated by qPCR) in addition to direct binding interactions (as determined by ChIP-chip data), and solid gray arrows indicate direct binding interactions only. The dashed black arrow indicates an indirect regulatory interaction only. See also Table S3 and Figure S7.

### Evolutionary Conservation of the Biofilm Network

Only a few of the many (probably more than a million) fungal species can proliferate and cause disease in humans. These pathogenic species are widely distributed over the fungal lineage, indicating that survival in a human host probably evolved independently multiple times. Although many fungal species can form aggregates (flocs, mats, biofilms, etc.), it seems likely that *C. albicans* is one of very few fungal species that can efficiently form biofilms in a healthy mammalian host. How then did the biofilm circuit evolve in the *C. albicans* lineage?

Several lines of evidence suggest that the biofilm network in *C. albicans* has undergone extensive evolutionary change relatively recently. First, as described in the Results, “young” genes are enriched in the biofilm circuit and “old” genes are underrepresented (Figure 4H). For example, ~120 *C. albicans* genes appear to have arisen (or at least have changed extensively) after the common ancestor of *C. albicans* and *Candida tropicalis* (a closely related species), and one-third of these are part of the biofilm circuit. Second, if we map (when possible) the *C. albicans* biofilm circuit target genes to other species, we find the motifs of two of the master regulatory proteins (Ndt80 and Efg1) only sporadically enriched in these genes (Figure S7C). Thus, the regulator-target gene connections are not strongly conserved outside of *C. albicans* itself. (This analysis could not be meaningfully performed for the other regulators due to a lack of predictive power of their motifs [see Extended Experimental Procedures]). Third, the intergenic regions targeted by biofilm regulators are much longer than average (Figure 4I), possibly providing a larger mutational target for the gain of binding sites. In combination with short motifs, this may help to explain how new genes have quickly become incorporated into the network. Finally, as we discuss in greater detail below, the functions of the master transcription regulators in *C. albicans*

have diverged significantly from their “assignments” in *S. cerevisiae*. Our data and analyses suggest that the biofilm networks of other CTG clade species (species that translate the CUG codon into serine instead of the conventional leucine, e.g., *C. tropicalis*, *Candida parapsilosis*, *Lodderomyces elongisporus*, *Debaryomyces hansenii*, *Candida guiermondii*, and *Candida lusitanae*) will likely be comprised of different transcription regulators and/or different target genes, or both.

### Evolutionary Reassignment of Transcription Regulators

A direct comparison between *C. albicans* and its nonpathogenic relative *S. cerevisiae* provides additional insight into how the biofilm network evolved. We can ask, for example, whether the six master transcription regulators of biofilm formation in *C. albicans* have clear orthologs in *S. cerevisiae* and, if so, what processes they regulate in *S. cerevisiae*. To explore orthology relationships for the master biofilm regulators, we used SYNERGY and INPARANOID mappings, in addition to hand annotation using constructed gene trees. Details are given in Extended Experimental Procedures.

Overall, this analysis indicates that the biofilm circuit consists of two regulators (Tec1 and Efg1) whose broad function—regulation of cell morphology—is deeply conserved in the fungal lineage. However, the set of target genes controlled by these regulators differs significantly between *S. cerevisiae* and *C. albicans* (Extended Experimental Procedures). A third regulator (Ndt80) is deeply conserved in the fungal lineage, but its function appears completely different between *S. cerevisiae* and *C. albicans*. In the former, it regulates meiosis (Hepworth et al., 1998) and, in the latter, biofilm formation. Two regulators (Rob1 and Brg1) are detectable only in species closely related to *C. albicans*, and the sixth biofilm regulator (Bcr1) has orthologs in *S. cerevisiae*, but they have not been characterized. Given that the DNA binding specificity of Tec1, Efg1, and Ndt80 is strongly conserved, extensive gains and losses of *cis*-regulatory sequence must be responsible, at least in part, for the evolution of the biofilm circuit in the *C. albicans* lineage. The Rob1 and Brg1 proteins appear to have undergone extensive changes in the *C. albicans* lineage such that their direct connection to the ancestor of *C. albicans* and *S. cerevisiae* (if any) has been

obscured. Thus, it seems likely that extensive changes in both regulators and *cis*-regulatory sequences were necessary for the evolution of the modern *C. albicans* biofilm circuit. These considerations, in combination with our analysis of “young” versus “old” genes, indicate that the *C. albicans* biofilm circuit evolved relatively recently, and we suggest that this development had an important role in the ability of *C. albicans* to adapt to its human host.

## EXPERIMENTAL PROCEDURES

### Strain Construction

Primer sequences (Table S7) and *C. albicans* strains (Table S6) are described in the [Extended Experimental Procedures](#); strains were constructed in isogenic backgrounds.

### In Vitro Biofilm Growth, Confocal Microscopy, and Biomass Determination

In vitro biofilm growth assays were carried out in Spider medium as described in detail in the [Extended Experimental Procedures](#). The average total biomass for each strain was calculated from five independent samples. Statistical significance (*p* values) was calculated with a Student's one-tailed paired *t* test.

### In Vivo Rat Catheter Biofilm Model

A rat central-venous catheter infection model (Andes et al., 2004) was used for in vivo biofilm modeling to mimic human catheter infections, as described in detail in the [Extended Experimental Procedures](#). Catheters were removed after 24 hr of *C. albicans* infection to assay biofilm development on the intraluminal surface by scanning electron microscopy (SEM).

### In Vivo Rat Denture Biofilm Model

A rat denture stomatitis infection model (Nett et al., 2010) was used for in vivo biofilm modeling to mimic human denture infections, as described in Nett et al. (2010), with certain modifications described in the [Extended Experimental Procedures](#). Dentures were removed after 24 hr post *C. albicans* infection to assay biofilm development on the denture surface by SEM.

### RNA Sample Preparation and Extraction

Details on growth, cell harvesting, RNA extraction, and treatment of biofilm and planktonic cells used for gene expression microarray and RNA-seq analysis are described in the [Extended Experimental Procedures](#).

### Gene Expression Microarray Design and Analysis

We used custom-designed *C. albicans* oligonucleotide microarrays (AMADID #020166) and a cutoff of 2-fold in both directions ( $\log_2 > 1.0$  and  $\log_2 < -1.0$ ) for the differential expression of biofilm versus planktonic cells and 1.5-fold in both directions ( $\log_2 > 0.58$  and  $\log_2 < -0.58$ ) for the differential expression of mutant over wild-type ([Extended Experimental Procedures](#) and Table S3A).

### Full-Genome Chromatin Immunoprecipitation Tiling Microarray

Each transcription regulator was tagged with a Myc tag at the C- or N-terminal end of the protein in a wild-type reference strain background. (In the case of Tec1, tagging the protein at either the C- or N-terminal end interfered with the protein's activity, and we used a custom-designed polyclonal antibody against an epitope near the C terminus of the Tec1 protein.) The tagged strains were grown under standard biofilm conditions (because the tags do not compromise function, the strains form normal biofilms) and were harvested for chromatin immunoprecipitation. After precipitation using the commercially available Myc antibody or the custom Tec1 antibody, the immunoprecipitated DNA and whole-cell extract were amplified and competitively hybridized to custom whole-genome oligonucleotide tiling microarrays (AMADID #016350) as described in the [Extended Experimental Procedures](#). Display, analysis, and identification of the binding events were determined using MochiView (Homann and Johnson, 2010).

### Motif Analysis

Motif analysis was performed using MochiView. MEME was also used to independently verify motifs found by MochiView. See [Data S2](#), [Tables S2H–S2M](#), and the [Extended Experimental Procedures](#) for details.

### RNA Sequencing of Biofilm and Planktonic Cells

Strand-specific, massively parallel SOLiD System sequencing of RNA from wild-type *C. albicans* biofilm and planktonic cells and mapping of resulting reads were performed as previously described (Tuch et al., 2010). Library amplification and sequencing resulted in 18 million planktonic and 28 million biofilm ~50 nt strand-specific sequence reads mappable to the *C. albicans* genome.

### Identification of Novel Transcriptionally Active Regions in Biofilms

nTARs were identified using MochiView. A previously published transcript annotation (Tuch et al., 2010) was used as a starting scaffold, and additional transcribed regions were identified. This approach identified 783 biofilm nTARs distinct from those in the previous annotation ([Extended Experimental Procedures](#) and Table S4A).

### Differential Expression Analysis of RNA-Seq Data

For every transcribed region in our expanded biofilm genome annotation, mean per nucleotide sequence coverage was extracted from both biofilm and planktonic data sets and transformed into pseudo-RPKM values, and transcripts differentially expressed between the two data sets were determined as described in [Extended Experimental Procedures](#). The union of the RNA-seq and microarray data sets was used to determine the final set of differentially expressed genes (Tables S4B and S4C). Statistical significance (*p* values) for the association of binding and differential transcription was calculated using a two-tailed Fisher's exact test.

### Association of Transcription Regulator Binding Sites with Adjacent Transcripts

To determine the association between transcription regulator binding and differential gene expression, the binding peaks identified by ChIP-chip were mapped to immediately adjacent, divergently transcribed genes. A transcription regulator binding site was considered to be associated with differential expression if at least one divergent flanking transcript was differentially expressed in either the microarray or the RNA-seq comparison.

### Exploring Orthology Relationships and Defining Gene Age Categories

Orthologs of the *C. albicans* and *S. cerevisiae* biofilm regulators and their direct targets were identified using freely available orthology mapping programs and by hand annotation using gene trees (See [Extended Experimental Procedures](#)). *C. albicans* gene age categories were defined as follows: “old” are members of gene families found in all Ascomycetes, “middle-aged” are members of gene families that arose after the divergence of *Schizosaccharomyces pombe* and *Schizosaccharomyces japonicus* but before the divergence of the CTG clade, and “young” are found only in CTG clade species. Overlap of age categories with biofilm-induced genes is described by the hypergeometric distribution ([Extended Experimental Procedures](#)).

### ACCESSION NUMBERS

All data have been deposited into the NCBI Gene Expression Omnibus (GEO) portal under the accession numbers GSE21291 (RNA-seq), GSE29785 (ChIP-chip), and GSE30474 (GE Array).

### SUPPLEMENTAL INFORMATION

Supplemental Information includes [Extended Experimental Procedures](#), seven figures, seven tables, and two data files and can be found with this article online at [doi:10.1016/j.cell.2011.10.048](https://doi.org/10.1016/j.cell.2011.10.048).

## ACKNOWLEDGMENTS

We thank Oliver Homann for developing MochiView, Christopher Baker and Isabel Necedal for help with evolutionary analysis, Francisco De La Vega for making possible the RNAseq analysis, Chiraj Dalal for computational advice, Lauren Booth for comments on the manuscript, and Sudarsi Desta, Jeanselle Dea, and Jorge Mendoza for technical assistance. We are grateful for the advice of Kurt Thorn in the acquisition of the CSLM images at the Nikon Imaging Center at UCSF. This study was supported by NIH grants R01AI073289 (D.R.A.) and R01AI049187 and R01AI083311 (A.D.J.). C.J.N. was supported by NIH fellowships T32AI060537 and F32AI088822. The content is the responsibility of the authors and does not necessarily represent the views of the NIH.

Received: May 3, 2011

Revised: August 9, 2011

Accepted: October 18, 2011

Published: January 19, 2012

## REFERENCES

- Alabadi, D., Oyama, T., Yanovsky, M.J., Harmon, F.G., Más, P., and Kay, S.A. (2001). Reciprocal regulation between TOC1 and LHY/CCA1 within the Arabidopsis circadian clock. *Science* *293*, 880–883.
- Andes, D., Nett, J., Oschel, P., Albrecht, R., Marchillo, K., and Pitula, A. (2004). Development and characterization of an in vivo central venous catheter *Candida albicans* biofilm model. *Infect. Immun.* *72*, 6023–6031.
- Askew, C., Sellam, A., Epp, E., Mallick, J., Hogues, H., Mullick, A., Nantel, A., and Whiteway, M. (2011). The zinc cluster transcription factor Ahr1p directs Mcm1p regulation of *Candida albicans* adhesion. *Mol. Microbiol.* *79*, 940–953.
- Baillie, G.S., and Douglas, L.J. (1999). Role of dimorphism in the development of *Candida albicans* biofilms. *J. Med. Microbiol.* *48*, 671–679.
- Borneman, A.R., Leigh-Bell, J.A., Yu, H., Bertone, P., Gerstein, M., and Snyder, M. (2006). Target hub proteins serve as master regulators of development in yeast. *Genes Dev.* *20*, 435–448.
- Bruno, V.M., Wang, Z., Marjani, S.L., Euskirchen, G.M., Martin, J., Sherlock, G., and Snyder, M. (2010). Comprehensive annotation of the transcriptome of the human fungal pathogen *Candida albicans* using RNA-seq. *Genome Res.* *20*, 1451–1458.
- Chandra, J., Kuhn, D.M., Mukherjee, P.K., Hoyer, L.L., McCormick, T., and Ghannoum, M.A. (2001). Biofilm formation by the fungal pathogen *Candida albicans*: development, architecture, and drug resistance. *J. Bacteriol.* *183*, 5385–5394.
- Costerton, J.W., Stewart, P.S., and Greenberg, E.P. (1999). Bacterial biofilms: a common cause of persistent infections. *Science* *284*, 1318–1322.
- de Hoon, M.J., Eichenberger, P., and Vitkup, D. (2010). Hierarchical evolution of the bacterial sporulation network. *Curr. Biol.* *20*, R735–R745.
- Donlan, R.M., and Costerton, J.W. (2002). Biofilms: survival mechanisms of clinically relevant microorganisms. *Clin. Microbiol. Rev.* *15*, 167–193.
- Douglas, L.J. (2003). *Candida* biofilms and their role in infection. *Trends Microbiol.* *11*, 30–36.
- Fernández, A., and Lynch, M. (2011). Non-adaptive origins of interactome complexity. *Nature* *474*, 502–505.
- Hamoen, L.W., Venema, G., and Kuipers, O.P. (2003). Controlling competence in *Bacillus subtilis*: shared use of regulators. *Microbiology* *149*, 9–17.
- Harbison, C.T., Gordon, D.B., Lee, T.I., Rinaldi, N.J., Macisaac, K.D., Danford, T.W., Hannett, N.M., Tagne, J.B., Reynolds, D.B., Yoo, J., et al. (2004). Transcriptional regulatory code of a eukaryotic genome. *Nature* *431*, 99–104.
- Hawser, S.P., and Douglas, L.J. (1994). Biofilm formation by *Candida* species on the surface of catheter materials in vitro. *Infect. Immun.* *62*, 915–921.
- Hepworth, S.R., Friesen, H., and Segall, J. (1998). NDT80 and the meiotic recombination checkpoint regulate expression of middle sporulation-specific genes in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* *18*, 5750–5761.
- Homann, O.R., and Johnson, A.D. (2010). MochiView: versatile software for genome browsing and DNA motif analysis. *BMC Biol.* *8*, 49.
- Homann, O.R., Dea, J., Noble, S.M., and Johnson, A.D. (2009). A phenotypic profile of the *Candida albicans* regulatory network. *PLoS Genet.* *5*, e1000783.
- Kojic, E.M., and Darouiche, R.O. (2004). *Candida* infections of medical devices. *Clin. Microbiol. Rev.* *17*, 255–267.
- Kolter, R., and Greenberg, E.P. (2006). Microbial sciences: the superficial life of microbes. *Nature* *441*, 300–302.
- Lavoie, H., Hogues, H., Mallick, J., Sellam, A., Nantel, A., and Whiteway, M. (2010). Evolutionary tinkering with conserved components of a transcriptional regulatory network. *PLoS Biol.* *8*, e1000329.
- Lin, Z., Wu, W.S., Liang, H., Woo, Y., and Li, W.H. (2010). The spatial distribution of cis regulatory elements in yeast promoters and its implications for transcriptional regulation. *BMC Genomics* *11*, 581.
- Locke, J.C., Southern, M.M., Kozma-Bognár, L., Hibberd, V., Brown, P.E., Turner, M.S., and Millar, A.J. (2005). Extension of a genetic network model by iterative experimentation and mathematical analysis. *Mol. Syst. Biol.* *1*, 2005.0013.
- Long, M., Betrán, E., Thornton, K., and Wang, W. (2003). The origin of new genes: glimpses from the young and old. *Nat. Rev. Genet.* *4*, 865–875.
- Losick, R., and Stragier, P. (1992). Crisscross regulation of cell-type-specific gene expression during development in *B. subtilis*. *Nature* *355*, 601–604.
- Madhani, H.D., and Fink, G.R. (1997). Combinatorial control required for the specificity of yeast MAPK signaling. *Science* *275*, 1314–1317.
- Mitrovich, Q.M., Tuch, B.B., De La Vega, F.M., Guthrie, C., and Johnson, A.D. (2010). Evolution of yeast noncoding RNAs reveals an alternative mechanism for widespread intron loss. *Science* *330*, 838–841.
- Müller, D., and Stelling, J. (2009). Precise regulation of gene expression dynamics favors complex promoter architectures. *PLoS Comput. Biol.* *5*, e1000279.
- Nett, J., and Andes, D. (2006). *Candida albicans* biofilm development, modeling a host-pathogen interaction. *Curr. Opin. Microbiol.* *9*, 340–345.
- Nett, J.E., Marchillo, K., Spiegel, C.A., and Andes, D.R. (2010). Development and validation of an in vivo *Candida albicans* biofilm denture model. *Infect. Immun.* *78*, 3650–3659.
- Ni, L., Bruce, C., Hart, C., Leigh-Bell, J., Gelperin, D., Umansky, L., Gerstein, M.B., and Snyder, M. (2009). Dynamic and complex transcription factor binding during an inducible response in yeast. *Genes Dev.* *23*, 1351–1363.
- Nobile, C.J., and Mitchell, A.P. (2005). Regulation of cell-surface genes and biofilm formation by the *C. albicans* transcription factor Bcr1p. *Curr. Biol.* *15*, 1150–1155.
- Nobile, C.J., Andes, D.R., Nett, J.E., Smith, F.J., Yue, F., Phan, Q.T., Edwards, J.E., Filler, S.G., and Mitchell, A.P. (2006a). Critical role of Bcr1-dependent adhesins in *C. albicans* biofilm formation in vitro and in vivo. *PLoS Pathog.* *2*, e63.
- Nobile, C.J., Nett, J.E., Andes, D.R., and Mitchell, A.P. (2006b). Function of *Candida albicans* adhesin Hwp1 in biofilm formation. *Eukaryot. Cell* *5*, 1604–1610.
- Nobile, C.J., Schneider, H.A., Nett, J.E., Sheppard, D.C., Filler, S.G., Andes, D.R., and Mitchell, A.P. (2008). Complementary adhesin function in *C. albicans* biofilm formation. *Curr. Biol.* *18*, 1017–1024.
- Nobile, C.J., Nett, J.E., Hernday, A.D., Homann, O.R., Deneault, J.S., Nantel, A., Andes, D.R., Johnson, A.D., and Mitchell, A.P. (2009). Biofilm matrix regulation by *Candida albicans* Zap1. *PLoS Biol.* *7*, e1000133.
- Ramage, G., VandeWalle, K., López-Ribot, J.L., and Wickes, B.L. (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.
- Ramage, G., Tomsett, K., Wickes, B.L., López-Ribot, J.L., and Redding, S.W. (2004). Denture stomatitis: a role for *Candida* biofilms. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endod.* *98*, 53–59.
- Sahni, N., Yi, S., Daniels, K.J., Huang, G., Srikantha, T., and Soll, D.R. (2010). Tec1 mediates the pheromone response of the white phenotype of *Candida*

- albicans: insights into the evolution of new signal transduction pathways. *PLoS Biol.* 8, e1000363.
- Schinabeck, M.K., Long, L.A., Hossain, M.A., Chandra, J., Mukherjee, P.K., Mohamed, S., and Ghannoum, M.A. (2004). Rabbit model of *Candida albicans* biofilm infection: liposomal amphotericin B antifungal lock therapy. *Antimicrob. Agents Chemother.* 48, 1727–1732.
- Sellam, A., Tebbji, F., and Nantel, A. (2009). Role of Ndt80p in sterol metabolism regulation and azole resistance in *Candida albicans*. *Eukaryot. Cell* 8, 1174–1183.
- Sellam, A., Hogues, H., Askew, C., Tebbji, F., van Het Hoog, M., Lavoie, H., Kumamoto, C.A., Whiteway, M., and Nantel, A. (2010). Experimental annotation of the human pathogen *Candida albicans* coding and noncoding transcribed regions using high-resolution tiling arrays. *Genome Biol.* 11, R71.
- Süel, G.M., Garcia-Ojalvo, J., Liberman, L.M., and Elowitz, M.B. (2006). An excitable gene regulatory circuit induces transient cellular differentiation. *Nature* 440, 545–550.
- Sugino, R.P., and Innan, H. (2011). Natural selection on gene order in the genome re-organization process after whole genome duplication of yeast. *Mol. Biol. Evol.* Published online May 5, 2011. 10.1093/molbev/msr118.
- Tuch, B.B., Mitrovich, Q.M., Homann, O.R., Hernday, A.D., Monighetti, C.K., De La Vega, F.M., and Johnson, A.D. (2010). The transcriptomes of two heritable cell types illuminate the circuit governing their differentiation. *PLoS Genet.* 6, e1001070.
- Uppuluri, P., Chaturvedi, A.K., Srinivasan, A., Banerjee, M., Ramasubramanian, A.K., Köhler, J.R., Kadosh, D., and Lopez-Ribot, J.L. (2010a). Dispersion as an important step in the *Candida albicans* biofilm developmental cycle. *PLoS Pathog.* 6, e1000828.
- Uppuluri, P., Pierce, C.G., Thomas, D.P., Bubeck, S.S., Saville, S.P., and Lopez-Ribot, J.L. (2010b). The transcriptional regulator Nrg1p controls *Candida albicans* biofilm formation and dispersion. *Eukaryot. Cell* 9, 1531–1537.
- Wapinski, I., Pfeffer, A., Friedman, N., and Regev, A. (2007). Natural history and evolutionary principles of gene duplication in fungi. *Nature* 449, 54–61.
- Webb, B.C., Thomas, C.J., Willcox, M.D., Harty, D.W., and Knox, K.W. (1998). *Candida*-associated denture stomatitis. Aetiology and management: a review. Part 2. Oral diseases caused by *Candida* species. *Aust. Dent. J.* 43, 160–166.
- Wilson, J. (1998). The aetiology, diagnosis and management of denture stomatitis. *Br. Dent. J.* 185, 380–384.
- Wilson, N.K., Foster, S.D., Wang, X., Knezevic, K., Schütte, J., Kaimakis, P., Chilarska, P.M., Kinston, S., Ouwehand, W.H., Dzierzak, E., et al. (2010). Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. *Cell Stem Cell* 7, 532–544.
- Young, R.A. (2011). Control of the embryonic stem cell state. *Cell* 144, 940–954.
- Zhu, C., Byers, K.J., McCord, R.P., Shi, Z., Berger, M.F., Newburger, D.E., Saulrieta, K., Smith, Z., Shah, M.V., Radhakrishnan, M., et al. (2009). High-resolution DNA-binding specificity analysis of yeast transcription factors. *Genome Res.* 19, 556–566.

## EXTENDED EXPERIMENTAL PROCEDURES

### Media

*C. albicans* strains were grown at 30°C in YPD medium (2% Bacto Peptone, 2% dextrose, 1% yeast extract) unless otherwise indicated. Transformants were selected for on SC medium (2% dextrose, 6.7% YNB with ammonium sulfate, and auxotrophic supplements) or on YPD+clonNAT400 (2% Bacto Peptone, 2% dextrose, 1% yeast extract, and 400 µg/ml nourseothricin (clonNAT, WERNER BioAgents)) for nourseothricin-resistant isolates. To obtain nourseothricin-sensitive isolates having flipped out the *SAT1* marker (Reuss et al., 2004), nourseothricin-resistant transformants were grown for 10 hr in YPM liquid medium, plated at a cell density of 200 cells/plate on YPD+clonNat25 (2% Bacto Peptone, 2% dextrose, 1% yeast extract, and 25 µg/ml nourseothricin (clonNAT, WERNER BioAgents)), and allowed to grow for 24 hr at 30°C as previously described (Reuss et al., 2004). Biofilms were grown in Spider medium (Liu et al., 1994) at 37°C. Planktonic cells were grown in Spider medium at 37°C or in SC+Uri medium (SC supplemented with 100 mg/l uridine) at 23°C.

### Plasmid and Strain Construction

All *C. albicans* strains used in this study are listed in Table S6 with the exception of the transcription regulator (TR) mutant deletion library. This TF library, consisting of 165 strains, has been previously described (Homann et al., 2009), and is publicly available at the Fungal Genetics Stock Center (<http://www.fgsc.net/>). Strains were constructed in isogenic strain backgrounds, and were derived from strain SN87 (His-Leu-) or SN152 (His-Leu-Arg-) (Noble and Johnson, 2005). The transcription regulator mutant deletion strain library (His+Leu+Arg- background), and reference strains SN250 (His+Leu+Arg-), SN425 (His+Leu+Arg+), and QMY23 (His+Leu+Arg+) have been previously described (Homann et al., 2009; Mitrovich et al., 2007; Noble et al., 2010). All primer sequences used in this study are listed in Table S7.

All genotypes were verified for correct integration by colony PCR using corresponding flanking detection primers. Plasmids for complementation (pCJN598 (*EFG1*), pCJN600 (*TEC1*), pCJN602 (*BCR1*), pCJN604 (*BRG1*), pCJN606 (*ROB1*), pCJN608 (*NDT80*)) were constructed using PCR and homologous recombination in *S. cerevisiae*, and were designed to contain a wild-type copy for each gene of interest with upstream promoter and downstream UTR sequences, followed by the *C. dubliniensis* *ARG4* gene between upstream and downstream flanking *C. albicans* *LEU2* gene sequences. The complemented strains CJN2318 (*EFG1*), CJN2320 (*TEC1*), CJN2322 (*BCR1*), CJN2324 (*BRG1*), CJN2326 (*ROB1*), and CJN2328 (*NDT80*) were made by transforming the transcription regulator mutant strains TF156 (*efg1Δ/Δ*), TF115 (*tec1Δ/Δ*), TF137 (*bcr1Δ/Δ*), TF022 (*brg1Δ/Δ*), TF110 (*rob1Δ/Δ*), and TF095 (*ndt80Δ/Δ*), respectively, with their corresponding PmeI-digested complementation plasmid described above. The transcription regulator mutant strains were made Arg+ by transforming with PmeI-digested pSN105 (Noble et al., 2010) to yield strains CJN2302 (*efg1Δ/Δ*), CJN2330 (*tec1Δ/Δ*), CJN2334 (*bcr1Δ/Δ*), CJN2338 (*brg1Δ/Δ*), CJN2408 (*rob1Δ/Δ*), and CJN2412 (*ndt80Δ/Δ*). SN425 (His+Leu+Arg+) (Noble et al., 2010) was used as the marker-matched prototrophic reference strain for all Arg+ add-back strains. Wild-type reference strain QMY23 (Mitrovich et al., 2007) was used for the RNA sequencing of biofilm and planktonic cells (see below).

The N-terminal Myc-tagged nourseothricin-resistant Myc-Rob1 (CJN2144) strain was constructed by transforming the reference strain SN250 using PCR products from template plasmid pADH70 (containing a 7XMyC epitope tag immediately preceding the *SAT1*-flipper cassette (see description below)) and primers CJNO1137 and CJNO1140. These primers amplify the entire 7XMyC epitope tag and complete *SAT1* flipper cassette with 66 bp of hanging homology to *ROB1* up to the start codon for the forward primer and 70 bp of hanging homology to *ROB1* precisely after the start codon for the reverse primer. The homology in these primers allows recombination of the entire 7XMyC epitope tag and complete *SAT1* flipper cassette directly upstream of *ROB1*, so that *ROB1* contains an in frame N-terminal 7XMyC epitope tag translational fusion after the marker has been flipped out. Correct integration of the N-terminal 7XMyC epitope tag and *SAT1* flipper for Rob1 was verified by colony PCR using detection primers CJNO1141 and AHO624 to check the upstream integration, and CJNO1142 and AHO613 to check the downstream integration. The N-terminal tagged nourseothricin-sensitive Myc-Rob1 (CJN2208) was constructed by flipping out the *SAT1* cassette from strain CJN2144, as described previously (Reuss et al., 2004). The following primer pairs were used in colony PCR to confirm the clean “flipping out” of the *SAT1*-flipper cassette for *ROB1*: CJNO1141 with AHO628 and CJNO1142 with AHO276. The 7XMyC epitope tag and the region of homology to the 5' end of the ORF of interest used for integration of the *SAT1*-flipper cassette was confirmed by sequencing the colony PCR product generated using primers CJNO1141 and AHO628.

The C-terminal Myc-tagged nourseothricin-resistant Efg1-Myc (CJN1775), Bcr1-Myc (CJN1785), Brg1-Myc (CJN1700), and Ndt80-Myc (CJN1707) strains were constructed by transforming the reference strain SN250 using PCR products from template plasmid pADH34 (containing a 13XMyC epitope tag immediately preceding the *SAT1*-flipper cassette, described in (Nobile et al., 2009)) and primers AHO385 and AHO386 (for Efg1), CJNO1046 and CJNO1047 (for Bcr1), CJNO1012 and CJNO1013 (for Brg1), and CJNO1024 and CJNO1025 (for Ndt80), respectively. These primers amplify the entire 13XMyC epitope tag and complete *SAT1* flipper cassette with 65 bp of hanging homology to the ORF of interest (minus stop codon) for the forward primer and 65 bp of hanging homology to the 3' UTR of the ORF of interest (precisely downstream of the stop codon) for the reverse primer. The homology in these primers allows recombination of the entire 13XMyC epitope tag and complete *SAT1* flipper cassette directly downstream of the ORF of interest, lacking its natural stop codon, so that the ORF contains a C-terminal 13XMyC epitope tag translational fusion. Correct integration of the C-terminal 13XMyC epitope tag and *SAT1* flipper was verified by colony PCR using detection primers AHO279 and AHO300 (for Efg1), CJNO1048 and AHO300 (for Bcr1), CJNO1014 and AHO300 (for Brg1), and CJNO1026

and AHO300 (for Ndt80) to check the upstream integration, and AHO281 and AHO301 (for Efg1), CJNO1049 and AHO301 (for Bcr1), CJNO1015 and AHO301 (for Brg1), and CJNO1027 and AHO301 (for Ndt80) to check the downstream integration. The C-terminal tagged nourseothricin-sensitive Efg1-Myc (CJN1781), Bcr1-Myc (CJN1787), Brg1-Myc (CJN1734), and Ndt80-Myc (CJN1748) strains were constructed by flipping out the *SAT1* cassette from strains CJN1775, CJN1785, CJN1700, and CJN1707, respectively, as described previously (Reuss et al., 2004). The following primer pairs were used in colony PCR to confirm the clean “flipping out” of the *SAT1*-flipper cassette for the ORF of interest: AHO279 with AHO300 and AHO281 with AHO302 (for Efg1), CJNO1048 with AHO300 and CJNO1049 with AHO302 (for Bcr1), CJNO1014 with AHO300 and CJNO1015 with AHO302 (for Brg1), and CJNO1026 with AHO300 and CJNO1027 with AHO302 (for Ndt80). The 13XMyC epitope tag and the region of homology to the 3' end of the ORF of interest used for integration of the *SAT1*-flipper cassette was confirmed by sequencing the colony PCR product generated using primers AHO279 with AHO283 (for Efg1), CJNO1048 with AHO283 (for Bcr1), CJNO1014 with AHO283 (for Brg1), and CJNO1026 with AHO283 (for Ndt80).

The *NAT1-TDH3* promoter plasmid pCJN542 (Nobile et al., 2008) was used for gene overexpression. All overexpression strains (CJN2555, CJN2527, CJN2600, CJN2530, CJN2531, CJN2533, CJN2567, CJN2549, CJN2604, CJN2552, CJN2569, CJN2570, CJN2473, CJN2476, CJN2479, CJN2480, CJN2483, CJN2486, CJN2395, CJN2609, CJN2397, CJN2536, CJN2611, CJN2499, CJN2537, CJN2539, CJN2557, CJN2558, CJN2561, CJN2564, CJN2573, CJN2576, CJN2607, CJN2578, CJN2580, CJN2583, CJN2541, CJN2351, CJN2601, CJN2544, CJN2546, CJN2354, CJN2585, CJN2587, CJN2590, CJN2591, CJN2595, CJN2597, CJN2704, CJN2684, CJN2700, CJN2691, CJN2690, CJN2687) were constructed by transforming the various transcription regulator mutant strains, using PCR products from template plasmid pCJN542 and specific primers (listed in Table S7) that amplify the entire *Ashbya gossypii* *TEF1* promoter, the *C. albicans* *NAT1* open reading frame, the *A. gossypii* *TEF1* terminator, and the *C. albicans* *TDH3* promoter with 70-100 bp of hanging homology to 500 bp upstream into the promoter of the gene being overexpressed for the forward primer and 70-100 bp of hanging homology from exactly the start codon of the gene being overexpressed. The homology in these primers allows for homologous recombination of the entire cassette directly upstream of the natural locus of the gene being overexpressed so that its expression is driven by the *TDH3* promoter instead of its natural promoter. Transformation into *C. albicans* strains and selection on YPD+clonNAT400 plates has been described for this construct (Nobile et al., 2008). Integration of the constructs was verified by colony PCR with a gene-specific forward detection primer annealing to a sequence within the promoter of each gene and the reverse primer CJNO875 annealing to a sequence found in the *NAT* gene (listed in Table S7). The overexpression strains were assayed by quantitative real time PCR (qPCR) to ensure that the chosen genes of interest were overexpressed in their corresponding strain backgrounds (data not shown).

The target gene deletion mutant strains TFT66a (*orf19.3337Δ/Δ*), TFT64b (*als1 Δ/Δ*), TFT68b (*tpo4Δ/Δ*), TFT70a (*eht1Δ/Δ*), TFT54a (*hyr1Δ/Δ*), TFT60d (*hwp1 Δ/Δ*), and TFT72b (*can2Δ/Δ*) were constructed with the method previously described (Homann et al., 2009; Noble and Johnson, 2005) by fusion PCR using the primers listed in Table S7. Target gene deletion mutant strain TF021 (*orf19.4000Δ/Δ*) was previously constructed (Homann et al., 2009). All deletions were verified by diagnostic PCR of the flanks surrounding the introduced markers and by attempting to amplify a small internal fragment of the ORF (for a successful deletion, this intra-ORF PCR yielded no product while a wild-type control yielded a clear product).

Transcriptional reporter strains of the biofilm transcriptional regulators (where one copy of each regulator's promoter was fused to an mCherry transcriptional reporter in both a homozygous regulator mutant as well as a heterozygous regulator mutant) were constructed as follows. The mCherry heterozygous reporter strains CJN2614, CJN2616, CJN1619, CJN2672, CJN2629, and CJN2621 were constructed by transforming SN425, using PCR products from template plasmid pADH77 (see description below) and gene-specific primers (listed in Table S7); transformants were selected on YPD+clonNAT400 plates, as described above. Integration of the constructs was verified by colony PCR with a gene-specific forward detection primer and the reverse primer AHO702 (see Table S7). The mCherry homozygous reporter strains CJN2708, CJN2710, CJN2712, CJN2736, CJN2715, and CJN2718 were constructed by transforming CJN2724, CJN2725, CJN2726, CJN2727, CJN2728, and CJN2729, respectively, using PCR products from template plasmid pADH77 and gene-specific primers (listed in Table S7); transformants were selected on SC-His+clonNAT200 plates, as described above. Integration of the constructs was verified by colony PCR with a gene-specific forward detection primer and the reverse primer AHO702 (see Table S7).

#### Description of N-Terminal Myc-Tagging Plasmid pADH70

The sequence of plasmid pADH70 (containing a 7XMyC epitope tag immediately preceding the *SAT1*-flipper cassette) used for N-terminal myc-tagging is available at GenBank (<http://www.ncbi.nlm.nih.gov/genbank>), accession # JN795133).

#### Description of mCherry-Tagging Plasmid pADH77

The sequence of plasmid pADH77 (containing an mCherry fluorescent tag immediately preceding the *SAT1*-flipper cassette) used for creating transcriptional reporters of the biofilm regulators is available at GenBank (<http://www.ncbi.nlm.nih.gov/genbank>), accession # JN795134).

#### In Vitro Biofilm Growth, Confocal Microscopy, and Biomass Determination

In vitro biofilm growth assays were carried out in Spider medium by growing the biofilm on either the silicone square substrate (Cardiovascular Instruments Corp, PR72034-060N) as described in (Nobile and Mitchell, 2005), or directly on the bottom of

12-well polystyrene plates (BD Falcon), as follows. Strains were grown overnight in YPD at 30°C, diluted to an optical density at 600 nm ( $OD_{600}$ ) of 0.5 in 2 ml Spider medium. The 12-well plate alone or 12-well plate with silicone squares had been pretreated overnight with bovine serum (Sigma, B-9433) and washed with 2 ml phosphate-buffered saline (PBS) to prepare it for the biofilm assay. The inoculated plate was incubated at 37°C for 90 min at 200 rpm agitation for initial adhesion of cells in an ELMI digital thermostatic shaker. The plates were washed with 2 ml PBS, and 2 ml of fresh Spider medium was added. The plate was incubated at 37°C for an additional 48 hr at 200 rpm agitation to allow biofilm formation. For visualization of strains over time, biofilms were allowed to form for 0 hr (imaging was done directly after the 90 min incubation step), 8 hr, 24 hr, or 48 hr. Biofilms grown on the silicone squares were used for confocal scanning laser microscopy (CSLM) visualization. For CSLM, biofilms were stained with 50  $\mu$ g/ml of concanavalin A Alexa Fluor 594 conjugate (conA-594) (Molecular Probes, C-11253) in the dark for 1 hr with 200 rpm agitation at 37°C. CSLM was performed in the Nikon Imaging Center at UCSF with a Nikon Eclipse C1si upright spectral imaging confocal microscope using a 40x/0.80W Nikon objective. For conA-594 visualization, a 561 nm laser line was used. Images were acquired using Nikon EZ-C1 Version 3.80 software, and assembled into maximum intensity Z-stack projections using Nikon NIS Elements Version 3.00 software. Biofilms grown on the bottom of the 12-well plates were used for biomass determination. For dry mass measurements, five replicate wells containing biofilms were used. The medium was removed, 2 ml of PBS was added to each well, the biofilms were disrupted and resuspended by pipetting, and the contents of each well were vacuum filtered over a pre-weighed 0.8  $\mu$ m nitrocellulose filter (Millipore, AAWG02500). A control well with no cells added was also vacuum filtered. The biofilm-containing filters were dried overnight, and weighed the following day. The average total biomass for each strain was calculated from five independent samples after subtracting the mass of the filter with no cells added. Statistical significance (*P* values) was calculated with a Student's one-tailed paired *t* test.

### In Vivo Rat Catheter Biofilm Model

A well established rat central-venous catheter infection model (Andes et al., 2004) was used for in vivo biofilm modeling to mimic human catheter infections, as described previously (Andes et al., 2004; Nobile et al., 2006a). For this model, specific-pathogen-free female Sprague-Dawley rats weighing 400 g (Harlan Sprague-Dawley) were used. A heparinized (100 U/ml) polyethylene catheter with 0.76 mm inner and 1.52 mm outer diameters was inserted into the external jugular vein and advanced to a site above the right atrium. The catheter was secured to the vein with the proximal end tunneled subcutaneously to the midscapular space and externalized through the skin. The catheters were inserted 24 hr prior to infection to permit a conditioning period for deposition of host protein on the catheter surface. Infection was achieved by intraluminal instillation of 500  $\mu$ l *C. albicans* cells ( $10^6$  cells/ml). After a 4 hr dwelling period, the catheter volume was withdrawn and the catheter flushed with heparinized 0.15 M NaCl. Catheters were removed after 24 hr of *C. albicans* infection to assay biofilm development on the intraluminal surface by scanning electron microscopy (SEM). Catheter segments were washed with 0.1 M phosphate buffer, pH 7.2, fixed in 1% glutaraldehyde/4% formaldehyde, washed again with phosphate buffer for 5 min, and placed in 1% osmium tetroxide for 30 min. The samples were dehydrated in a series of 10 min ethanol washes (30%, 50%, 70%, 85%, 95%, and 100%), followed by critical point drying. Specimens were mounted on aluminum stubs, sputter coated with gold, and imaged using a Hitachi S-5700 or JEOL JSM-6100 scanning electron microscope in the high-vacuum mode at 10 kV. Images were assembled using Adobe Photoshop Version 7.0.1 software.

### In Vivo Rat Denture Biofilm Model

A recently developed rat denture stomatitis infection model (Nett et al., 2010) was used for in vivo biofilm modeling to mimic human denture infections, as described in Nett et al., with certain modifications described here. For this model, specific-pathogen-free male Sprague-Dawley rats weighing 350 g (Harlan Sprague-Dawley) were used. Rats were immunosuppressed with a single dose of 200 mg/kg subcutaneous cortisone at the time of denture placement. For denture placement, a stainless steel wire was threaded between the cheek teeth, a metal spatula was placed over the hard palate to create a space for *C. albicans* inoculation, and acrylic denture material (Maxitemp HP) was applied over the cheek teeth and wire to create the denture surface. The denture material was then molded into place, allowed to solidify for 5 min, and the spatula was removed. For denture infection, the hard palate beneath the acrylic denture was inoculated with 100  $\mu$ l *C. albicans* cells ( $10^8$  cells/ml). In order to reduce enteric bacterial colonization, ampicillin sodium/sulbactam sodium at 100 mg/kg was subcutaneously administered twice daily while the dentures were in place. Dentures were removed after 24 hr post *C. albicans* infection to assay biofilm development on the denture surface by SEM. Dentures were processed for SEM as described previously (Andes et al., 2004) and briefly above, for the catheters.

### RNA Sample Preparation and Extraction

Biofilms for gene expression microarray and RNA-seq analysis were grown in Spider medium at 37°C directly on the bottom of 6-well polystyrene plates, as described above, but with all volumes at 4 ml. One 6-well plate containing biofilms for one strain yields sufficient RNA for gene expression microarray and RNA-seq analysis. Biofilms were harvested by gently pipetting up and down along the bottoms of the 6-well plates, and combining the biofilm slurry of the same strain from each well of one 6-well plate in a 50 ml conical tube. Biofilm slurries were then centrifuged at 3,000  $\times$  g for 5 min, and total RNA was extracted using the RiboPure-Yeast RNA kit (Ambion, AM1926) or by the hot-phenol method (Hernday et al., 2010). To isolate total RNA for gene expression microarrays from planktonic cells, Spider medium was inoculated with organisms from a YPD 30°C overnight culture to obtain a starting  $OD_{600}$  of

0.05, incubated at 37°C until harvested by vacuum filtration when the OD<sub>600</sub> was 1.0, snap-frozen in liquid nitrogen, and total RNA was extracted using the RiboPure-Yeast RNA kit (Ambion, AM1926). To isolate RNA for RNA-seq from planktonic cells, cells were grown in SC+Uri medium to an OD<sub>600</sub> of 1.0 at 30°C, harvested by centrifugation (5 min, 2,000 x g), and snap-frozen in liquid nitrogen; total RNA was extracted by the hot phenol method (Hernday et al., 2010). For all total RNA samples used for RNA-seq, poly(A) RNA was isolated from 50 µg of total RNA by two rounds of purification using the Poly(A)Purist MAG kit (Ambion, AM1922).

### Gene Expression Microarray Design and Analysis

Synthesis of cDNA and dye coupling were performed as previously described (Nobile et al., 2009). The gene expression microarrays were custom-designed oligonucleotide microarrays, containing at least two independent probes for each ORF from the *C. albicans* Assembly 21 genome (<http://www.candidagenome.org/>), and printed by Agilent Technologies (AMADID #020166). The gene expression microarray experiments were performed and analyzed as previously described using LOWESS normalization (Lohse and Johnson, 2010). Microarray data was clustered using Cluster Version 3.0 (de Hoon et al., 2004), and visualized using Java TreeView Version 1.13 (Saldanha, 2004). Expression microarray data are reported in Table S3A as the median of three independent experiments. We used a cutoff of twofold in both directions ( $\log_2 > 1.0$ , and  $\log_2 < -1.0$ ) for the differential expression of biofilm versus planktonic cells, and 1.5-fold in both directions ( $\log_2 > 0.58$ , and  $\log_2 < -0.58$ ) for the differential expression of mutant over wild-type. Raw gene expression array data are available at the Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo>, accession # GSE30474).

### Full-Genome Chromatin Immunoprecipitation Tiling Microarray

Each transcription regulator was tagged with a Myc tag at the C- or N-terminal end of the protein in a wild-type reference strain background. (In the case of Tec1, tagging the protein at either the C- or N-terminal end interfered with the protein's activity, and we used a custom-designed polyclonal antibody against an epitope near the C terminus of the Tec1 protein.) The tagged strains were grown under standard biofilm conditions (because the tags do not compromise function, the strains form normal biofilms), and the biofilm cells harvested for chromatin immunoprecipitation. After precipitation using the commercially available Myc antibody or the custom Tec1 antibody, the immunoprecipitated DNA and whole-cell extract were amplified and competitively hybridized to custom whole-genome oligonucleotide tiling microarrays. The ChIP-chip tiling microarrays were designed by tiling 181,900 probes of 60 bp length across 14.3 Mb included in the *C. albicans* Assembly 21 genome (<http://www.candidagenome.org/>), as previously described (Tuch et al., 2008), and printed by Agilent Technologies (AMADID #016350). All Myc-tagged regulator strains, untagged strains, and delete strains were grown under the same biofilm-inducing conditions as the strains grown for gene expression microarray analysis. The ChIP-chip experiments were performed as previously described (Nobile et al., 2009) with two independent biological replicates for each strain. Normalized enrichment values were determined for every probe on the microarray by LOWESS normalization using Agilent Chip Analytics Version 1.2 software (see Agilent Chip Analytics manual for details). Display, analysis and identification of the binding events were determined using MochiView Version 1.45 software (<http://johnsonlab.ucsf.edu/sj/mochiview-start/>) (Homann and Johnson, 2010), where peaks for the Myc-tagged strain (plus Myc antibody) or the wild-type strain (plus custom antibody), are compared to peaks from an untagged reference strain (plus Myc antibody) or the deletion strain (plus custom antibody). All ChIP-chip data analysis was performed with MochiView by mapping to the *Candida albicans* Assembly 21 genome (<http://www.candidagenome.org/>) downloaded on 4/1/2010. Identification of binding events for both independent biological replicates was determined by smoothing the two data sets together using the "Extract Peaks from Data Set(s)" utility described in detail in the MochiView manual. Briefly, a smoothing function is first applied to the Chip Analytics  $\log_2$  enrichment values, followed by the application of a peak detection algorithm, where all binding peaks are assigned a *P* value using permutation testing. Peak-finding significance thresholds were kept at the default settings,  $p \leq 0.001$  for the experimental IPs (i.e., Myc-tagged regulator strains and wild-type strains with custom antibodies), and  $p \leq 0.05$  for the control IPs (i.e., untagged strains and deletion strains). For greater confidence, the amount of sampling was increased tenfold from the default setting to 100,000 (number of random samples to compare against each peak), and 100 (maximum number of random samples passing for inclusion of peak). The user-defined cutoffs for the minimum value for peak inclusion postsmoothing (values ranging from 0.40-0.58) were determined using the distribution of  $\log_2$ -ratios for each regulator, and were set at two standard deviations from the mean of  $\log_2$ -transformed fold enrichments. User-defined cutoffs for the minimum value for peak inclusion postsmoothing ranging from 0.27-0.36 (1.5 standard deviations from the mean of  $\log_2$ -transformed fold enrichments) were used for the untagged and delete IP control data sets. We note that adjustments to these peak inclusion cutoffs do not alter the primary conclusions that we make from our ChIP-chip data. We have chosen highly stringent significance thresholds ( $p \leq 0.001$ ) for the analysis of our ChIP-chip data in order to maximize our confidence in indicating a direct binding event through a ChIP signal, and note that, as with all ChIP-chip data, the lack of a called peak does not necessarily indicate the absence of a direct binding event in that region. Raw ChIP-chip data are available at the Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo>, accession # GSE29785).

### Assessment of Myc-Tagged Proteins for Functionality

We did not observe a biofilm phenotype for our six heterozygous transcriptional regulator mutant strains. Thus, to ensure that Myc-tagging the regulators in the wild-type background did not interfere with the function of the regulator of interest, we additionally

created a Myc-tagged allele over a deletion to confirm that the Myc-tagged allele is functional in promoting biofilm production. In addition, when Myc-tagging our regulators of interest, the constructs are sequenced to ensure that there are no obvious coding errors, and a Western blot is performed before attempting a ChIP experiment.

### Custom Antibody for Tec1

Since C-terminal and N-terminal Myc-tagging of Tec1 interfered with the function of the protein, we used a custom-designed polyclonal antibody to epitope ELPSSAKPVRLQK from amino acids 612–625 of the Tec1 protein. The peptide was synthesized, conjugated, polyclonal antibodies were produced in rabbits, and the antibody was affinity purified by Promab Biotechnologies. This Tec1 antibody (Promab, 29307) was used at a concentration of 10  $\mu$ g/ml instead of the Myc antibody during the Tec1 ChIP-chip experiments.

### Motif Analysis

Motif finding and assessment was performed with MochiView's "motif finder" function using 250 bp centered on the midpoint of half of the extracted peaks for each regulator, and analyzed for significance of enrichment in the remaining half of extracted peaks for that regulator using MochiView's "enrichment" function (Data S2). This utility determines the likelihood of finding the identified motif by chance in random intergenic regions of the same length. MEME Version 3.4.7 (<http://meme.nbcr.net>) software using 250 bp centered on the midpoint of all of the significantly called peaks was also used to independently verify the motifs found by MochiView. Locations of motif instances for each regulator, identified within their corresponding binding peaks, are shown in Table S2H–M. Overall, the fraction of bound locations containing instances of the motifs depends on the regulator, and ranges from 30%–100% (30% for Bcr1, 100% for Tec1, 100% for Efg1, 93% for Ndt80, 43% for Rob1, 44% for Brg1).

Motif distribution summary plots (Data S2) were constructed with MochiView's "compact motif/data/location plot" function using the minimum LOD score cutoffs established from the enrichment plots for each of the regulator motifs.

Motif distributions relative to binding peaks and to start codons (Data S2) were determined using MochiView's "motif  $\rightarrow$  distribution  $\rightarrow$  relative to locations" function. *P* values are calculated for the motif distributions at 50%, 60%, 70%, 80%, 90%, and 100% of the maximum motif score using a bootstrapped chi-square goodness of fit test (see MochiView manual). Motif occurrences for Tec1, Efg1, and Ndt80 were significantly centered at the area of ChIP-chip binding peak enrichment (Data S2). Motifs for all regulators were uniformly distributed across promoters, and there was no significant bias toward the start codon (Data S2).

### RNA Sequencing of Biofilm and Planktonic Cells

Strand-specific, massively-parallel SOLiD System sequencing of RNA from wild-type *C. albicans* biofilm and planktonic cells and mapping of resulting reads were performed as previously described (Tuch et al., 2010) using poly(A) RNA. Library amplification was performed using barcoded SOLiD PCR Primer Sets 5 and 10 for planktonic and biofilm samples, respectively. Sequencing was performed on a full slide with eight other samples (not presented here), and resulted in 18 million planktonic and 28 million biofilm  $\sim$ 50 nt strand-specific sequence reads mappable to the *C. albicans* genome. Sequence data are available at the Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo>, accession # GSE21291).

### Identification of Novel Transcriptionally Active Regions in Biofilms

Novel transcriptionally active regions (nTARs) were identified using the biofilm RNA-seq dataset and the "Create data set by extracting enriched regions from tiled set" feature of MochiView (Homann and Johnson, 2010). A previously published transcript annotation (Tuch et al., 2010) was used as a starting scaffold, and additional transcribed regions identified using a sliding window size of 125 nt, a trim multiplier of 0.01, a minimum location size of 50 nt, a location threshold cutoff of 20 and a location merge interval of 50 nt. This approach identified 783 biofilm nTARs distinct from those in the previous annotation (Table S4A).

### Differential Expression Analysis of RNA-Seq Data

For every transcribed region in our expanded biofilm genome annotation, mean per-nucleotide sequence coverage was extracted from both biofilm and planktonic WIG files (generated using Life Technologies' SOLiD Whole Transcriptome Pipeline, available at <http://solidsoftwaretools.com/gf/project/transcriptome>), and transformed into pseudo-RPKM values (reads per kilobase per million mapped reads). The DEGseq package (Wang et al., 2010) was then used to determine which transcripts were differentially expressed between the two datasets, using a likelihood ratio test with a false discovery rate (q-value) cutoff of 0.005 (Storey and Tibshirani, 2003); an expression change of at least twofold was also required. The union of the RNA-seq and microarray datasets was used to determine the final set of differentially expressed genes (Table S4B and Table S4C). The union was chosen (rather than the intersection) of differentially expressed genes because we wanted to increase the sensitivity and be more inclusive of the differentially expressed targets in the biofilm network. We noticed that genes called by only one set were often just below the significance threshold in the other set, but still appeared to be differentially expressed. We found that some of these genes proved to be important for biofilm formation, and thus we wanted to be sure to include them. There are very few cases (24 out of 2,235 genes) where the directions between the microarrays and RNA-seq data were conflicting, and these genes are listed as "unchanged" in Table S4D. Statistical significance (*P* values) for the association of binding and differential transcription was calculated using a two-tailed Fisher's exact test.

### Association of Transcription Regulator Binding Sites with Adjacent Transcripts

To determine the association between transcription regulator binding and differential gene expression, the binding peaks identified by ChIP-chip were mapped to immediately adjacent, divergently transcribed (i.e., downstream) genes (with no threshold on distance). In our experience, certain chromosomal loci frequently exhibit non-specific (artifactual) chromatin immunoprecipitation; such loci were removed from our sets of predicted binding sites by screening out those that overlapped centromeres (0.7% of peaks) or heavily transcribed regions (mean biofilm RNA-seq coverage of > 60 within 100 bp surrounding peak on either strand; 4% of peaks). The Assembly 21 annotation of the *C. albicans* genome was used (van het Hoog et al., 2007), but with transcription start sites defined by RNA-seq (Tuch et al., 2010), and supplemented with previously annotated nTARs (Tuch et al., 2010) and with biofilm nTARs annotated herein. Likely spurious genes (identified as those with mappable sequences having a maximal pseudo-RPKM value of < 1 in all of our RNA-seq datasets) were removed from the annotation prior to peak association. A transcription regulator binding site was considered to be associated with differential expression if at least one divergent flanking transcript was differentially expressed in either the microarray or the RNA-seq comparison, using the criteria described above. The expected background association between transcription regulator binding and differential expression (0.32) was calculated as the average likelihood of a chance association across all binding sites; the likelihood for individual binding events was calculated as 0 for regions flanked by convergent transcripts, 0.26 (i.e., the overall proportion of differentially expressed genes) for regions flanked by one divergent transcript, and the probability that at least one of two randomly selected genes is differentially expressed ( $1 - (1 - 0.26)^2$ ) for regions flanked by two divergent transcripts.

### Exploring Orthology Relationships and Defining Gene Age Categories

The hemiascomycete species include the model yeast, *S. cerevisiae*, and the predominant fungal pathogen of humans, *C. albicans* (Pappas et al., 2004). These species are estimated to have diverged from a common ancestor between 300-700 million years ago (Hedges et al., 2004); in terms of protein conservation, the two yeasts differ from each other to about the same extent as humans and fish (Dujon et al., 2004).

Orthologs of the *C. albicans* and *S. cerevisiae* biofilm regulators were identified by orthology mappings using SYNERGY (Remm et al., 2001) from The Fungal Orthogroups Repository (<http://www.broad.mit.edu/regev/orthogroups/>) and InParanoid (Wapinski et al., 2007) from the *Candida* Genome Database (CGD) (Costanzo et al., 2006) (<http://www.candidagenome.org>). In the cases of ambiguous orthology relationships from these two sources, further analysis was performed by hand annotation using gene trees generated from alignments of similar genes from 32 fungal genomes as identified by PSI-BLAST. Orthology comparisons for the direct targets of the regulators were based on orthology tables from The Fungal Orthogroups Repository (<http://www.broad.mit.edu/regev/orthogroups/>) using our *C. albicans* ChIP-chip direct targets, and *S. cerevisiae* direct targets from the YEASTRACT database (Teixeira et al., 2006) (<http://www.yeasttract.com/>).

Tec1, a TEA/ATTS protein family member, represents the simplest case. The *C. albicans* Tec1, which regulates hyphal growth in biofilms (Nobile and Mitchell, 2005; Schweizer et al., 2000), has a one-to-one orthologous relationship to Tec1 in *S. cerevisiae*, where it controls pseudohyphal development (Gavrias et al., 1996; Liu et al., 1993). (*S. cerevisiae* does not make true hyphae.) Thus, Tec1 promotes aspects of cell morphology in both species. Despite the fact that the Tec1 protein and its recognition motif are conserved between *S. cerevisiae* and *C. albicans*, the set of genes controlled by Tec1 in the two species (as determined by ChIP-chip) has diverged considerably; for example, only 3% of the *C. albicans* Tec1-controlled genes are also controlled by Tec1 in *S. cerevisiae* (data presented here compared with that of Borneman et al., 2007). This situation is not unusual: even though a transcription regulator might control the same general process in *S. cerevisiae* and *C. albicans* (in this case, production of an elongated cell morphology), the genes controlled can vary significantly (Bennett et al., 2003; Booth et al., 2010; Borneman et al., 2007; Martchenko et al., 2007).

The situation with Efg1, a basic-helix-loop-helix protein, is similar in principle. *C. albicans* Efg1 and the closely related *C. albicans* protein Efh1 have a two-to-two orthology relationship with *S. cerevisiae* Phd1 and Sok2, which, like Tec1, are regulators of pseudohyphal growth (Gimeno and Fink, 1994; Ward et al., 1995). (We note that deletion of Efh1 had no apparent effect on *C. albicans* biofilm formation in our screen (Figure 1)). Like Tec1, Efg1 likely orchestrated changes in cell morphology (from spherical yeast form cells to more elongated forms) in the common ancestor of *S. cerevisiae* and *C. albicans*.

Ndt80, a member of an unusual class of DNA-binding proteins, seems to have very different roles in *C. albicans* and *S. cerevisiae* despite an almost identical recognition motif. *C. albicans* has two Ndt80 paralogs (the other is Orf19.513, whose deletion has no effect on biofilm formation (Figure 1)), while *S. cerevisiae* has a single ortholog, which is produced only in meiosis and is devoted to regulating the middle meiosis genes (Hepworth et al., 1998; Unal et al., 2011). *C. albicans* does not appear to undergo meiosis (having a parasexual cycle (Bennett and Johnson, 2005)), and, as we show here, Ndt80 in this species plays a prominent role in biofilm production. A universal feature of microbial biofilms is their resistance to drugs and other antimicrobial agents, and, although *NDT80* is required for formation of biofilms in *C. albicans*, it also plays a role in regulating drug resistance by controlling the expression of *CDR1*, which encodes a major drug efflux pump (Chen et al., 2004). Indeed, our data also indicate that Ndt80 directly and indirectly controls the expression of many drug efflux pumps in biofilms (for example, *CDR3*, and *CDR4* directly; *CDR11* indirectly). At this point, we do not know whether Ndt80 controlled meiosis or biofilm production in the ancestor of *S. cerevisiae* and *C. albicans*, but it is clear that the regulator plays very different roles in the two modern species.

Finally, *C. albicans* Bcr1, a C<sub>2</sub>H<sub>2</sub> zinc finger protein, has a one-to-two orthology relationship (as a result of the whole genome duplication) with *S. cerevisiae* Usv1 and Rgm1. However, neither of these genes has been extensively studied, and it is therefore not

possible to meaningfully compare their function between the two species. *C. albicans* Rob1, a zinc cluster protein, and Brg1, a GATA protein, have no identifiable orthologs outside of the *Candida* (CTG) clade (with the possible exception of a Brg1 ortholog in *Yarrowia lipolytica*, a species recently reported to form biofilms (Dusane et al., 2008).

Genes of different age groups have been shown to have different functional properties (Capra et al., 2010). Gene ages were defined using orthology assignments from The Fungal Orthogroups Repository (<http://www.broad.mit.edu/regev/orthogroups/>). “Old” *C. albicans* genes are members of gene families found in all Ascomycetes. “Middle-aged” *C. albicans* genes are members of gene families that arose after the divergence of *S. pombe* and *S. japonicus* but before the divergence of the CTG clade. “Young” genes are found only in CTG clade species. Overlap of age categories with biofilm-induced genes is described by the hypergeometric distribution, which was approximated by the Pearson’s chi-square test without Yates’ continuity correction to obtain *P* values. For all age categories (Table S4E),  $p < 1.23E-9$ . Similar results to those in Figure 4H were obtained using the expression, binding, and inter-section of those datasets.

### Details on Identification and Phenotypic Characterization of Biofilm-Defective Transcription Regulator Mutants In Vitro

Of the nine mutants that were revealed from our visual and biomass biofilm screens, we did not follow up on TF091 (*orf19.1685Δ/Δ*) because its biofilm defect is due to a growth defect specific to Spider medium (the mutant is unable to utilize mannitol as a carbon source (Homann et al., 2009)), TF103 (*orf19.3063Δ/Δ*) because its biofilm defect is the result of an overall growth defect (Homann et al., 2009), and TF117 (*tup1Δ/Δ*) because this mutant is highly pleiotropic (Braun and Johnson, 1997; Homann et al., 2009; Zhao et al., 2002) and its phenotype interfered with the biofilm assay.

A comprehensive phenotypic characterization of the homozygous transcription regulator mutant library that we describe here has been previously reported (Homann et al., 2009). Homann et al. assessed phenotypes of this mutant library under 55 different growth conditions, including several conditions for assaying colony morphologies and drug susceptibilities (Homann et al., 2009). Biofilm formation assays, however, were not previously performed on this library. Our six biofilm defective transcription regulator mutants that we identify here, do not appear to have any one phenotype in common based on the phenotypic profiles reported in Homann et al. We note that *tec1Δ/Δ* strains exhibited reduced colony wrinkling and reduced peripheral filamentation in several media, *bcr1Δ/Δ* and *ndt80Δ/Δ* strains exhibited enhanced invasion on several media, and *efg1Δ/Δ*, *brg1Δ/Δ*, and *rob1Δ/Δ* strains exhibited reduced invasion on several media (Homann et al., 2009). Based on this information, it is possible that reductions in peripheral filamentation, or the presence of an invasive growth phenotype in either direction (hyper- or hypo- invasion), may potentially be an indicator of a biofilm defect. However, there are many mutants in the library that display reductions in peripheral filamentation (e.g., *rim101Δ/Δ*, *isw2Δ/Δ*, *orf19.7381Δ/Δ*, and *orf19.1168Δ/Δ*), enhanced invasive growth (e.g., *sfl1Δ/Δ*, *stp4Δ/Δ*, *orf19.6102Δ/Δ*, *orf19.6874Δ/Δ*, and *orf19.6798Δ/Δ*), and reduced invasive growth (e.g., *cph2Δ/Δ* and *orf19.921Δ/Δ*) (Homann et al., 2009), that we did not find to be defective in biofilm formation (Table S1 and Figure 1). Thus, the characteristics that contribute to defining a biofilm-defective mutant cannot be predicted based on other morphological attributes of the mutant.

### Assessing Hyphal Formation under Biofilm and Planktonic Conditions In Vitro

For assessing hyphal formation under biofilm conditions, strains were grown under standard biofilm-inducing conditions in Spider medium at 37°C for 48 hr. Two hundred cells free-floating in the medium were counted and analyzed for hyphal formation by light microscopy. From this, the percentage of true hyphae was determined for each transcription regulator mutant strain under biofilm conditions (Figure S3). For assessing hyphal formation under planktonic conditions, strains were grown planktonically at 37°C under three types of filament-inducing conditions: 1) RPMI medium for 90 min, 2) Spider medium for 3 hr, and 3) YPD+10% Serum for 2 hr. Strains were inoculated from a saturated overnight YPD culture into the corresponding filament-inducing medium at an  $OD_{600} = 0.2$ . Two hundred cells from each medium were counted and analyzed for hyphal formation by light microscopy. From this, the percentage of true hyphae was determined for each transcription regulator mutant strain under planktonic conditions (Figure S3).

We found that, with the exception of the *efg1Δ/Δ* strain, we were able to detect true hyphae suspended in the medium surrounding the biofilm (but not necessarily in the biofilm itself) for the mutants under biofilm conditions (Figure S3A). Note that cells dispersed from biofilms are predominantly in the yeast form; 21% of cells suspended in the medium from a wild-type biofilm are hyphae (Figure S3A). The *efg1Δ/Δ* strain was the only strain completely defective for hyphal formation under biofilm conditions; 0% of cells suspended in the medium from an *efg1Δ/Δ* strain are true hyphae (Figure S3A). The *efg1Δ/Δ* strain was also defective for hyphal formation under every planktonic hyphal-inducing condition that we tested (Figure S3B). Thus, consistent with previous findings (Lo et al., 1997; Ramage et al., 2002; Stoldt et al., 1997), *efg1Δ/Δ* has a strong hyphal defect under various in vitro conditions, including biofilm conditions, and has the strongest hyphal defect of our six biofilm-defective regulator mutants. We also observed that the absolute number of true hyphae suspended in the medium for the *tec1Δ/Δ*, *ndt80Δ/Δ*, and *rob1Δ/Δ* strains was slightly reduced (6%, 7%, and 10% hyphae, respectively), compared to wild-type (21% hyphae), under our biofilm conditions (Figure S3A). In addition, the *tec1Δ/Δ*, *ndt80Δ/Δ*, and *rob1Δ/Δ* strains were defective for hyphal formation, compared to wild-type, under some planktonic hyphal-inducing conditions that we tested (Figure S3B). Hyphal defects have been previously reported for *tec1Δ/Δ* and *ndt80Δ/Δ* *in vitro* (Schweizer et al., 2000; Sellam et al., 2010). We did not detect any hyphal defects for the *bcr1Δ/Δ* or *brg1Δ/Δ* strains under our biofilm conditions (Figure S3A) or under the planktonic hyphal-inducing conditions that we tested (Figure S3B). This finding is consistent with the previous finding that Bcr1 is not required for hyphal formation per se, but is important for overall adherence of hyphae and yeast cells

(Nobile and Mitchell, 2005). We note that our newly identified biofilm regulator mutant, *brg1*  $\Delta/\Delta$ , at least in terms of its hyphal formation phenotype, appears most similar to *bcr1*  $\Delta/\Delta$ .

#### Additional Information on the Characterization of Biofilm-Defective Transcription Regulator Mutants In Vitro

Based solely on the in vitro biofilm phenotypes of the six core transcription regulator mutants that we describe, below we attempt to categorize the mutants according to their appearances in our assay. We visualized each regulator mutant over time during the development of the biofilm (Figure S1). At the 0 hr time-point (90 min postadherence), the wild-type reference strain and the *bcr1*  $\Delta/\Delta$  strain both had a uniform distribution of yeast-form cells, many of which began to form hyphae. In contrast, the *rob1*  $\Delta/\Delta$ , *ndt80*  $\Delta/\Delta$ , *brg1*  $\Delta/\Delta$ , and *tec1*  $\Delta/\Delta$  strains formed a dense mat of primarily yeast-form cells. The *efg1*  $\Delta/\Delta$  strain only sparsely colonized the substrate and occasional pseudohyphae were observed. The *bcr1*  $\Delta/\Delta$  strain did not appear deficient at the 0 hr time-point, however it began to show a defect at the 8 hr time-point. At 8 hr, the wild-type strain contained many hyphal cells and was 4 times thicker than at the 0 hr time-point, while the *bcr1*  $\Delta/\Delta$  strain had not changed substantially since the 0 hr time-point. The other mutant strains also made minimal progress and were thicker and/or denser at 8 hr than at 0 hr, but were largely lacking hyphae within their biofilms. The *tec1*  $\Delta/\Delta$  strain appeared to contain some of the same long and vertical hyphae as seen in wild-type, although they were sparse compared to wild-type and disappeared by the 24 hr time-point. At 24 hr, wild-type formed a robust biofilm, while most of the mutants were still rudimentary mats composed largely of yeast-form and pseudohyphal cells. Interestingly, the *bcr1*  $\Delta/\Delta$  was capable of forming a morphologically intact biofilm at this time-point, however, it was extremely delicate and sloughed off the substrate, leaving behind a few scattered yeast and hyphal cells. It appeared that by 48 hr, the *bcr1*  $\Delta/\Delta$  began to fill in the sloughed-off regions with yeast and hyphal cells, and somewhat resembled the other mutant regulators at this stage. From this phenotypic description over time, we can categorize the *bcr1*  $\Delta/\Delta$  as being defective at a later time-point than the other mutants. The other mutants were deficient at either making or retaining hyphal cells immediately after adhesion as well as throughout biofilm development. The *efg1*  $\Delta/\Delta$  strain had the most severe phenotype overall as it appeared defective in adhesion, hyphal formation, and thickness at every stage of biofilm formation. The *tec1*  $\Delta/\Delta$  and *bcr1*  $\Delta/\Delta$  strains were the most successful at incorporating hyphae into the biofilm, at least transiently. While none of the mutants appeared retarded at any particular developmental stage throughout biofilm development, per se, *rob1*  $\Delta/\Delta$ , *brg1*  $\Delta/\Delta$ , *ndt80*  $\Delta/\Delta$ , and *efg1*  $\Delta/\Delta$  strains most closely resembled wild-type at the 0 hr time-point, while *tec1*  $\Delta/\Delta$  and *bcr1*  $\Delta/\Delta$  strains, in terms of biofilm architecture, made some attempts at forming the components of a mature biofilm, but ultimately failed by the 48 hr time-point likely due to adherence deficiencies.

#### Additional Information on the Characterization of Biofilm-Defective Transcription Regulator Mutants in a Rat Denture In Vivo Model

In the initial characterization of this in vivo denture model, the *bcr1*  $\Delta/\Delta$  strain was the strain that was used to validate this model, and our findings are consistent with what was observed by Nett et al. (Nett et al., 2010). Nett et al. reported that the *bcr1*  $\Delta/\Delta$  biofilm formed on the rat denture contained 4-fold fewer adherent *C. albicans* cells, and a 50% higher bacterial burden compared to the reference strain (Nett et al., 2010). One simple hypothesis for this finding is that, during oral colonization of the denture, the presence of fewer colonizing *C. albicans* cells will result in an increase in bacterial colonization of the denture. This inference is consistent with what we observe for the other five mutant strains, which had little to no colonizing *C. albicans* cells on their dentures (Figure 2); we were, however, able to detect extensive bacterial biofilms consisting of both cocci and rod bacteria on their dentures (Figure S4B).

#### Additional Information on the Analysis of the ChIP-Chip Data

To identify genes directly regulated by Bcr1, Tec1, Efg1, Ndt80, Rob1, and Brg1, we performed full-genome chromatin immunoprecipitation microarray (ChIP-chip) to map the position across the genome to which each of the six transcription regulators are bound. We tagged each transcription regulator with a Myc tag at the C- or N-terminal end of the protein (choosing the scheme that did not interfere with the protein's activity) in a wild-type reference strain background. (In the case of Tec1, tagging the protein at either the C- or N-terminal end interfered with the protein's activity, and we used a custom-designed polyclonal antibody against an epitope near the C terminus of the Tec1 protein.) We then grew the tagged strains under standard biofilm conditions (because the tags do not compromise function, the strains form normal biofilms), and harvested the biofilm cells for chromatin immunoprecipitation. After precipitation using the commercially available Myc antibody or the custom Tec1 antibody, the immunoprecipitated DNA and whole-cell extract were amplified and competitively hybridized to custom whole-genome oligonucleotide tiling microarrays.

For analysis of the ChIP-chip data, peaks for the Myc-tagged strain (plus Myc antibody) or the wild-type strain (plus custom antibody), are compared to peaks from an untagged reference strain (plus Myc antibody) or the deletion strain (plus custom antibody), respectively, by mapping the data onto each chromosome. We chose stringent significant thresholds  $p < 0.001$  for the peak-finding analysis of our ChIP-chip data in order to maximize our confidence in indicating a direct binding event through a ChIP signal, with the caveat in mind that the lack of a called significant peak does not necessarily indicate the absence of a direct binding event in that region. Our analysis revealed the following number of significant binding peaks (binding events): 314 for Bcr1, 90 for Tec1, 393 for Efg1, 660 for Ndt80, 100 for Rob1, and 309 for Brg1 (see Table S2A–F for a complete list of every bound location for each regulator and Data S1 for MochiView image plots of every called significant peak for each regulator). The numbers of peaks that we report here include counting multiple peaks that may bind the same intergenic region. We then mapped these called significant peaks to intergenic regions of the *C. albicans* genome by taking the maximum peak enrichment value ( $\log_2$  normalized enrichment) of the peak/s

restricted to an intergenic region, thereby only counting a single maximum peak value per bound intergenic region. Our analysis based on intergenic regions containing a peak from the ChIP-chip signals that meets our significance thresholds indicates the following number of bound regions for each regulator: 211 for Bcr1, 76 for Tec1, 328 for Efg1, 558 for Ndt80, 95 for Rob1, and 283 for Brg1 (see Table S2G for a list of every intergenic region containing a significant called peak). Our analysis indicates that the biofilm regulatory network consists of 831 intergenic regions bound by one or more regulator, 350 intergenic regions bound by two or more, 186 intergenic regions bound by three or more, 111 intergenic regions bound by four or more, 55 intergenic regions bound by five or more, and 18 intergenic regions bound by all six of the biofilm regulators (Table S2G). Indeed, 42% of the 831 bound intergenic regions are bound by at least two of the six regulators.

Finally, to determine which target genes contain a significant called peak in their upstream promoter regions, we assigned peak enrichment values to promoters (defined as in (Tuch et al., 2010)), and listed bound target genes based on the presence of a binding event in their upstream promoter region (see Table S3A for a list of every gene significantly bound in their upstream promoter by each regulator). In the case of the latter dataset, binding events are listed twice if the upstream promoter region is flanked by two divergent ORFs, thus listing each candidate target gene's promoter; and binding events occurring outside of promoter regions (e.g., between two convergent ORFs) are not listed. From our analysis, we found the following number of bound promoters containing a significant called peak from the ChIP-chip signals of each regulator: 252 for Bcr1, 107 for Tec1, 447 for Efg1, 836 for Ndt80, 96 for Rob1, and 311 for Brg1 (see Table S3A for a list of every promoter containing a significant called peak). Of these 1,061 target genes - 23 are bound by all six, 77 are bound by five or more, 165 are bound by four or more, 265 are bound by three or more, and 458 (43%) are bound by two or more of the biofilm regulators.

### Additional Information on Identifying Functionally Relevant Target Genes of the Core Biofilm Network

We observed the following transcriptional responses from largest to smallest when *EFG1*, *ROB1*, *BRG1*, *NDT80*, *TEC1*, and *BCR1* were deleted (48%, 35%, 29%, 16%, 7%, and 7% of the genome changing, respectively, based on our significance threshold of  $\log_2 > 0.58$ , and  $\log_2 < -0.58$ ) (Table S3A). For Bcr1, we found 234 genes that were downregulated and 173 genes that were upregulated in the *bcr1*  $\Delta/\Delta$  mutant (Table S3C). Of these 407 differentially regulated genes, 46 of them (11%) were also directly bound in their promoters by Bcr1 (Table S3C). Precisely half of these 46 direct target genes were downregulated and half of them were upregulated in the *bcr1*  $\Delta/\Delta$  mutant, arguing that although Bcr1 is typically thought of as an activator of its target genes (Nobile and Mitchell, 2005), we show that Bcr1 can act as both an activator and repressor of its biofilm-relevant direct targets. For Tec1, we found 233 genes that were downregulated and 226 genes that were upregulated in the *tec1*  $\Delta/\Delta$  mutant (Table S3C). Of these 459 differentially regulated genes, 40 of them (9%) were also directly bound in their promoters by Tec1 (Table S3C). 33 of these directly bound target genes (90%) were downregulated in the *tec1*  $\Delta/\Delta$  mutant, while only 4 direct target genes were upregulated in the *tec1*  $\Delta/\Delta$  mutant, suggesting that Tec1 is predominantly an activator of its biofilm-relevant direct targets. For Efg1, we found 1,410 genes that were downregulated and 1,537 genes that were upregulated in the *efg1*  $\Delta/\Delta$  mutant (Table S3C). Of these 2,947 differentially regulated genes, 276 (9%) were also directly bound in their promoters by Efg1 (Table S3C). 179 of these directly bound target genes (65%) were downregulated in the *efg1*  $\Delta/\Delta$  mutant, and 97 (35%) were upregulated in the *efg1*  $\Delta/\Delta$  mutant, suggesting a dual role for Efg1 as both an activator and repressor of its biofilm-relevant direct targets. For Ndt80, we found 593 genes that were downregulated and 406 genes that were upregulated in the *ndt80*  $\Delta/\Delta$  mutant (Table S3C). Of these 999 differentially regulated genes, 273 (27%) were also directly bound in their promoters by Ndt80 (Table S3C). 143 of these directly bound target genes (52%) were downregulated in the *ndt80*  $\Delta/\Delta$  mutant, and 130 (48%) were upregulated in the *ndt80*  $\Delta/\Delta$  mutant, suggesting that Ndt80 is both an activator and repressor of its biofilm-relevant direct targets. For Rob1, we found 1,121 genes that were downregulated and 1,029 genes that were upregulated in the *rob1*  $\Delta/\Delta$  mutant (Table S3C). Of these 2,150 differentially regulated genes, 46 (2%) were also directly bound in their promoters by Rob1 (Table S3C). 33 of these directly bound target genes (72%) were downregulated in the *rob1*  $\Delta/\Delta$  mutant, and 13 (28%) were upregulated in the *rob1*  $\Delta/\Delta$  mutant, suggesting that Rob1 is both an activator and repressor of its biofilm-relevant direct targets. For Brg1, we found 822 genes that were downregulated and 931 genes that were upregulated in the *brg1*  $\Delta/\Delta$  mutant (Table S3C). Of these 1,753 differentially regulated genes, 130 (7%) were also directly bound in their promoters by Brg1 (Table S3C). 101 of these directly bound target genes (78%) were downregulated in the *brg1*  $\Delta/\Delta$  mutant, and 29 (22%) were upregulated in the *brg1*  $\Delta/\Delta$  mutant, suggesting that Brg1 is both an activator and repressor of its biofilm-relevant direct targets.

By looking at the correlation between our ChIP-chip binding data and our gene expression array data when that regulator is deleted, we note that we do not find an improvement in this correlation by considering regions bound by only the relevant transcriptional regulator (or a smaller subset of the regulators). Thus, the lack of a complete correlation between binding and regulation is unlikely to be a result of redundancy of binding (see Table S3B).

In Figure 5, we identified eight target genes that are positively regulated by all six transcriptional regulators; however they are not bound by all six regulators. We find that a hierarchical cascade between the biofilm regulators can explain this indirect regulation. These eight genes are not direct targets of all six regulators, however they are direct targets of at least one regulator, and indirect targets can be explained by regulators that are direct targets (Figure S5). For example, Ndt80 binds and directly regulates six out of the eight target genes. Ndt80 also directly regulates *BRG1*, which in turn directly regulates *TEC1*, which then directly regulates *EFG1*, which then directly regulates the other two target genes (Table S3C; Figure S5). The hierarchical cascade can be followed in this way, from each transcriptional regulator to each of the eight target genes, explaining the differential regulation of these genes, even in the absence of a direct binding event by the initial regulator.

In addition to identifying the direct targets of each regulator, the ChIP-chip and expression profiling data indicate that differential expression of indirect target genes of any one regulator may be explained via regulation of the other five transcriptional regulators (Table S3C). For example, Rob1 directly regulates only 2% of its targets, however it also directly regulates Brg1, Bcr1, and Tec1; at least one of each binds an additional 12% of Rob1's indirect targets. If the other biofilm regulators (Ndt80 and Efg1) are included, 9% more of Rob1's indirect targets are bound, making a total of 23% of Rob1's indirect targets bound by at least one regulator in the biofilm circuit. If this analysis is done for each biofilm regulator, 6% of Ndt80's indirect targets (and 33% of all of its targets) are explained; 18% of Brg1's indirect targets (and 25% of all of its targets) are explained; 11% of Efg1's indirect targets (and 21% of all of its targets) are explained; 31% of Bcr1's indirect targets (and 42% of all of its targets) are explained; and 27% of Tec1's indirect targets (and 36% of all of its targets) are explained. Additionally, Ndt80, Brg1, Efg1, Bcr1, Tec1, and Rob1 bind and regulate, respectively, 29, 26, 29, 4, 4, and 2, transcriptional regulators other than the 6 regulators in the biofilm circuit. Thus, it is possible that indirect targets of the biofilm regulators may be regulated by these other regulators. Interestingly, the direct targets of the biofilm regulators are significantly enriched for transcriptional regulators ( $p = 4.573E-12$ ; Pearson's chi-square test): 11% of Ndt80 targets, 21% of Brg1 targets, 12% of Efg1 targets, 11% of Bcr1 targets, 15% of Tec1 targets, and 11% of Rob1 targets; only 5% of genes in the genome encode transcriptional regulators (Table S3D). We note that of the 52 transcriptional regulators directly regulated by a member of the biofilm circuit, we screened deletion mutants for 34 of these (as they were in the library), and found no biofilm defects by dry weight. It is possible that deletion of one of the remaining 15 transcriptional regulators may have a biofilm defect, but 8 of these may be essential (as homozygous deletion mutants were not obtained) and the other 7 are yet to be attempted.

Similar to the biofilm network, the regulators in the pseudohyphal growth circuit of *S. cerevisiae* bind many additional regulators outside of the main circuit (Borneman et al., 2006). In contrast, transcriptional regulators that are not major players in complex regulatory networks, but rather control their targets in more simple genetic pathways, such as Zap1 in *C. albicans* and Gal4 in *S. cerevisiae*, bind only a few, if any transcriptional regulators (Nobile et al., 2009; Ren et al., 2000). Thus, we suggest that transcriptional regulators that are main players in complex networks may be more likely to regulate other transcriptional regulators.

### GO Term Functional Analysis of Biofilm Target Genes

Genes specifically regulated by only one biofilm transcriptional regulator or genes that are regulated by multiple biofilm transcriptional regulators may share common functions. However, 75% of the *C. albicans* genome remains uncharacterized, which makes it difficult to assign function to groups of genes; we attempted to do this analysis using Gene Ontology. While most of the targets of the regulators were not significantly enriched for any particular GO category, there were some exceptions. Targets that were regulated by all 6 biofilm transcriptional regulators were enriched for the following GO terms: adhesion, biofilms, and hexose biosynthesis ( $p = 1E-4$  to 0.05). Targets that were regulated by 5 biofilm transcriptional regulators were enriched for multi-organism processes including biofilm formation and host interaction ( $p = 1E-8$  to  $1E-6$ ). Targets that were regulated by 4 biofilm transcriptional regulators were enriched for organic substrate transport and membrane transport ( $p = 1E-6$  to  $1E-4$ ). For genes that were regulated by only one biofilm transcriptional regulator, only the targets of Ndt80 and Rob1 had any sort of functional enrichment. Ndt80 targets were enriched for genes involved in reproduction ( $p = 1E-4$  to 0.05), including *KIC1*, *HHF1*, *HHF22*, and *VPS11*. Rob1 targets were enriched for intracellular transport ( $p = 1E-4$  to 0.05), including *PEP7*, *VPS21*, *REI1*, *ORF19.3128*, *ORF19.479.2*, and *ORF19.7202*. While these GO term categories provide a hint to the function of various target gene sets, they explain only a small fraction of regulated genes. Note that the majority of genes in the *C. albicans* genome do not have a GO term, and those that do are based on not only hand-curated entries but also high throughput data and computational prediction.

### Additional Information on Target Gene Ectopic Expression

We note that there were no genes among similar functional classes as the eight target genes that we chose for overexpression that were not differentially regulated in all of the regulator mutants compared to the reference strain. However, among target genes within similar functional classes, there was one gene that was differentially regulated in just one of the regulator mutants. Therefore, we overexpressed this gene, *TPO5* (*ORF19.151*), which is predicted to be within the same functional class as *TPO4* (*ORF19.473*), one of the target genes we implicate in playing a role in biofilm formation. While *TPO4* is differentially regulated in all six regulator mutants compared to reference strain, *TPO5* is differentially regulated (just under 2-fold) only in the *brg1Δ/Δ* mutant strain background. When *TPO5* is overexpressed in the backgrounds of all of the regulator mutants, it is not able to rescue biofilm formation (data not shown). We believe that this is consistent with our hypothesis that some of our target genes are important in biofilm formation, while other genes (even from similar functional classes) are not.

### Additional Information on the Functionally Relevant Target Genes Controlled by the Biofilm Circuit

In this study, we prioritized our focus on the eight target genes that were downregulated in all six transcription regulator mutant strains when compared to WT (as it turned out, they were also bound by at least one regulator). We made homozygous deletion mutant strains for each of these eight genes and also created strains where each target gene was overexpressed in the background of each regulator mutant. We analyzed the resulting 56 strains by dry weight assays and CSLM. From the deletion strains, three were deficient in biofilm formation: *als1Δ/Δ* and *hwp1Δ/Δ* (both of which were previously known to have biofilm defects (Nobile et al., 2006a; Nobile et al., 2006b; Nobile et al., 2008)), and also *can2Δ/Δ*, which has not been identified previously as important for biofilm formation. All three of these proteins are predicted to be cell-surface localized. Hwp1 is a hyphal-specific protein and

Als1 is expressed in both yeast and hyphal cells (Coleman et al., 2010; Green et al., 2005; Staab et al., 1996). Als1 and Hwp1 both play roles in cell-cell adhesion as well as surface adhesion, and it is thought that decreased adhesion contributes to poor biofilm formation in the null mutants.

CAN2 is predicted to encode an amino acid permease, and the protein sequence has 81% identity to Can1, a confirmed lysine/arginine/histidine permease. Amino acid permeases not only provide amino acid substrates for metabolic pathways, but also transport drugs and toxic substances and may be involved in nutrient sensing and signaling pathways (Hundal and Taylor, 2009; Sophianopoulou and Diallinas, 1995). In *C. albicans*, arginine activates the filamentation pathway, and arginine biosynthesis is required for *C. albicans* to escape macrophages via filamentation (Ghosh et al., 2009). It is possible that Can2 contributes to filamentation pathways via arginine uptake. Alternatively, Can2 may uptake other small molecules or participate in nutrient signaling pathways important for biofilm formation. Interestingly, CAN2 expression was previously found to be upregulated in wild-type biofilms formed in the rat catheter model (Nett et al., 2009), whereas the other known amino acid permease, CAN1 was not differentially regulated. This data indicates that Can2 plays an important role in the host environment.

Three of the eight top target gene candidates were required for normal biofilm formation, however, all but two of them (ORF19.3337 and HYR1) were able to partially rescue biofilm formation when overexpressed in at least one transcription regulator mutant strain background. This is not surprising, given that many proteins have functional redundancy. Overexpression of six of the eight target genes was able to partially rescue both the *bcr1*  $\Delta/\Delta$  and *tec1*  $\Delta/\Delta$  mutants. Previous work has shown that BCR1 expression is dependent on Tec1, so this may explain why these mutant backgrounds are similarly rescued (Nobile and Mitchell, 2005).

Further study will be needed to fully understand the roles of each of these target genes in biofilm formation. However, we can form some hypotheses based on previous work on some of these proteins in *C. albicans* and also homology to *S. cerevisiae* proteins. As described above, Als1 and Hwp1 are both adhesin proteins, thus enhancing adhesion between biofilm cells; Als1 has no clear orthologs in *S. cerevisiae*. Little is known about the other four genes that resulted in partial rescue. ORF19.4000 is a predicted homeodomain transcription regulator, potentially regulating its own set of target genes that may compensate for the gene sets no longer regulated by the lack of Bcr1 or Tec1. Like CAN2, TPO4 encodes a putative transporter, but with similarity to both polyamine and MFS drug transporters. Polyamines are essential for normal cell growth and polyamine levels are carefully regulated in *E. coli*, *S. cerevisiae*, and higher eukaryotes (Igarashi and Kashiwagi, 2010a, 2010b). It is unclear what the roles of Tpo4 and Can2 are in *C. albicans*, but both may be transporting small molecules that affect biofilm formation. The final rescue was mediated by overexpression of EHT1. EHT1 encodes a putative alcohol acyl transferase, and contains a predicted hydrolase catalytic domain. The *S. cerevisiae* orthologs, Eht1 and Eeb1, play roles in lipid metabolism (Athenstaedt et al., 1999). In *C. albicans*, EHT1 is induced in response to alpha pheromone, thus it may play a role in polarized growth, which is also important in biofilm formation (Bennett and Johnson, 2006).

Our genome wide approach provided us with an unbiased list of candidate target genes potentially involved in biofilm formation. We chose to focus first on the eight genes that are positively regulated by all six of the biofilm regulators. Of the 8 target genes, 6 had a biofilm phenotype, which validates our approach for selecting a high confidence set of target genes. Future work will examine the roles of these 6 genes in biofilm formation and will also screen additional target genes identified here for biofilm phenotypes.

### Additional Information on the Evolutionary Conservation of the Biofilm Network

Our evolutionary analysis indicated that biofilm genes are enriched for young and middle-aged genes, while old genes are underrepresented in the biofilm network. In addition, we noticed that the intergenic regions bound by biofilm regulators were longer than the genome-wide average. Because young genes tend to have longer intergenic regions (Sugino and Innan, 2011), we asked whether the enrichment of young genes can explain the increased length of biofilm regulator targets. However, biofilm targets had much longer intergenic regions than most young genes ( $p < 2.2E-16$ ) (Figure 4I).

We note that the enrichment of young and middle-aged genes and underrepresentation of old genes is also true for regulator-bound genes. By incorporating the array data, we find that old genes are indirectly regulated, whereas young genes are directly regulated by the biofilm regulators (Table S4E).

The biofilm network has additional features that may point to its rapid evolution. For example, the mean length of intergenic regions bound by each individual biofilm regulator is longer than those that are not bound by the biofilm regulators (see Figure 4I and Table S4F). In addition, and in contrast to other transcription regulators whose position is confined to a narrow region upstream of the transcription start site (e.g., Mat $\alpha$ 2 in *S. cerevisiae*; Johnson and Herskowitz, 1985), the regulatory sites for the biofilm regulators were found interspersed throughout intergenic regions (Data S2).

### Information on the Enrichment of the Biofilm Regulator Motifs across Other Yeast Species

Genes that are known targets of biofilm regulators in *C. albicans* were mapped to gene families using The Fungal Orthogroups Repository (<http://www.broad.mit.edu/reggev/orthogroups/>). Upstream intergenic regions from thirteen species were scored using the motifs of each biofilm regulator and two control regulators (Lavoie et al., 2010) using MAST (Bailey and Gribskov, 1998) (Figure S7C). Ndt80 and Efg1 were the only biofilm regulators with motifs containing sufficient predictive power in *C. albicans* to be scored across the other yeast species (Figure S7C). Enrichment of the motif upstream of biofilm regulator targets was quantified using a log-ratio. A cutoff for the motif score from MAST was selected to maximize the log-ratio of each regulator in *C. albicans* and was then used for the other species. Based on the control regulator, we considered a log ratio greater than 2.8 as significant, below 2.0 as insignificant, and in between 2.0 and 2.8 as borderline significant. Overall we found that the *C. albicans* biofilm targets were not well conserved to other

fungal species. Nonetheless, *C. albicans* Ndt80 biofilm targets were the most conserved, where we observed significant conservation in *C. tropicalis*, *C. parapsilosis*, and *L. elongisporus*. Based on Ndt80 and Efg1 biofilm target gene conservation, there does not appear to be any trend in targeting between pathogenic and non-pathogenic fungal species (Figure S7C).

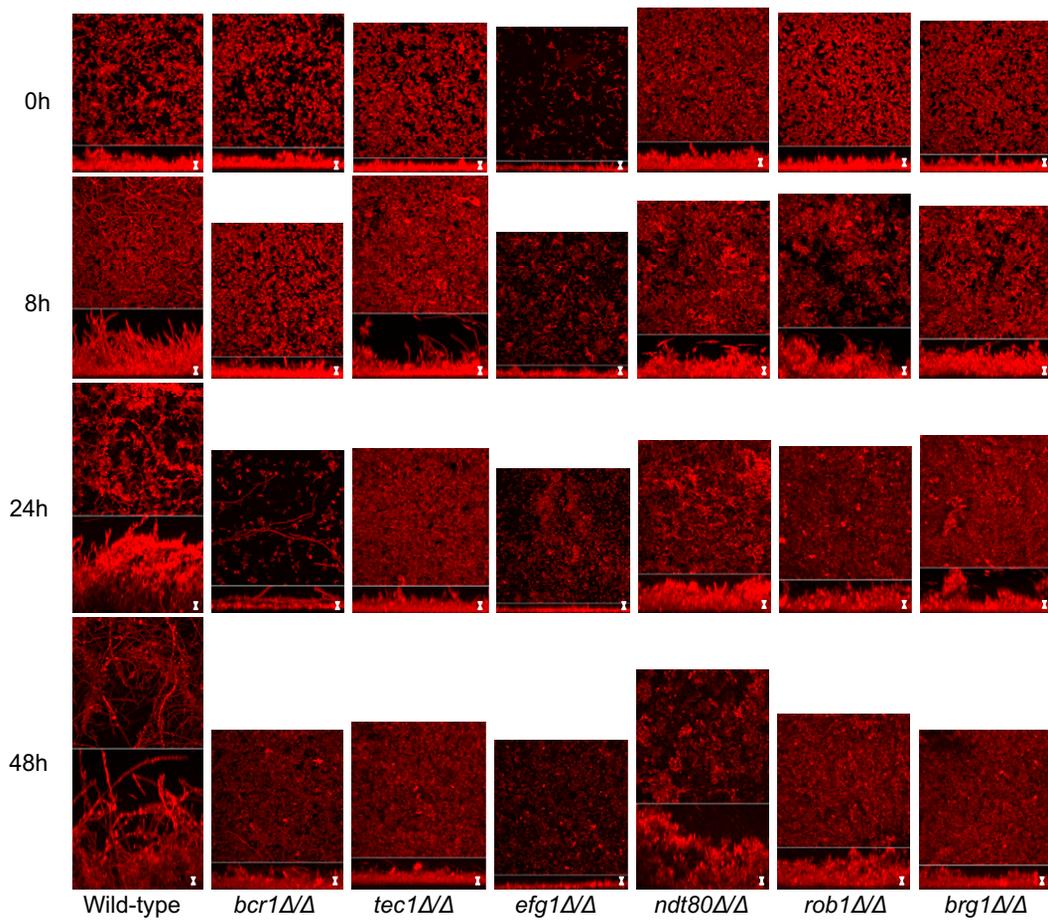
### Evolutionary Analysis of *C. albicans* Genes without Orthologs in *S. cerevisiae*

According to the current CGD orthology mappings, there are 1,801 ORFs with no orthologs in *S. cerevisiae*. We find that these 1,801 *C. albicans* ORFs with no orthologs in *S. cerevisiae* map to both pathogenic and non-pathogenic species of the *Candida* clade. Overall, 1199 map to *C. tropicalis* (pathogenic), 1047 map to *D. hansenii* (non-pathogenic), 1040 map to *C. parapsilosis* (pathogenic), 984 map to *L. elongisporus* (non-pathogenic), 977 map to *C. guillermoidii* (pathogenic), and 925 map to *C. lusitanae* (pathogenic). A total of 468 of these ORFs are specific to *C. albicans*. Of these genes that do not map to *S. cerevisiae*, 35 map to *Candida glabrata*, an independently-evolved pathogen, which diverged after the whole genome duplication. We find that 649 of these ORFs map to all CTG clade species. Genes found only in pathogenic species are potentially interesting because they may identify a pathogenesis module (Butler et al., 2009). We find that 722 of these ORFs (half of which are not yet characterized) map to all pathogenic CTG clade species (this excludes *D. hansenii* and *L. elongisporus*). If we include *C. glabrata* in addition to all of the pathogenic CTG clade species, 19 of these ORFs (half of which are not yet characterized) map to these species. Some of these ORFs that may form a pathogenesis module contain some genes that are in the biofilm network (e.g., gene families encoding HYR/Iff-like, Als-like adhesins, Pga-like, and major facilitator transporters proteins), however they are not statistically enriched for biofilm genes ( $p > 0.05$ ).

### SUPPLEMENTAL REFERENCES

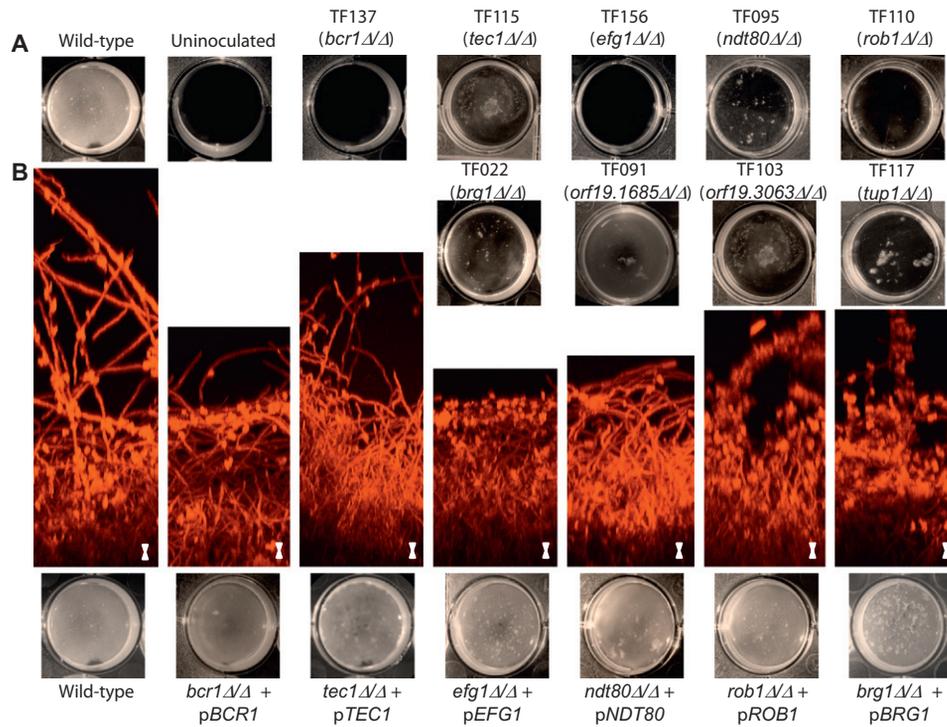
- Athenstaedt, K., Zweyck, D., Jandrositz, A., Kohlwein, S.D., and Daum, G. (1999). Identification and characterization of major lipid particle proteins of the yeast *Saccharomyces cerevisiae*. *J. Bacteriol.* *181*, 6441–6448.
- Bailey, T.L., and Gribskov, M. (1998). Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* *14*, 48–54.
- Bennett, R.J., and Johnson, A.D. (2005). Mating in *Candida albicans* and the search for a sexual cycle. *Annu. Rev. Microbiol.* *59*, 233–255.
- Bennett, R.J., and Johnson, A.D. (2006). The role of nutrient regulation and the Gpa2 protein in the mating pheromone response of *C. albicans*. *Mol. Microbiol.* *62*, 100–119.
- Bennett, R.J., Uhl, M.A., Miller, M.G., and Johnson, A.D. (2003). Identification and characterization of a *Candida albicans* mating pheromone. *Mol. Cell. Biol.* *23*, 8189–8201.
- Booth, L.N., Tuch, B.B., and Johnson, A.D. (2010). Intercalation of a new tier of transcription regulation into an ancient circuit. *Nature* *468*, 959–963.
- Borneman, A.R., Gianoulis, T.A., Zhang, Z.D., Yu, H., Rozowsky, J., Seringhaus, M.R., Wang, L.Y., Gerstein, M., and Snyder, M. (2007). Divergence of transcription factor binding sites across related yeast species. *Science* *317*, 815–819.
- Braun, B.R., and Johnson, A.D. (1997). Control of filament formation in *Candida albicans* by the transcriptional repressor TUP1. *Science* *277*, 105–109.
- Butler, G., Rasmussen, M.D., Lin, M.F., Santos, M.A., Sakthikumar, S., Munro, C.A., Rheinbay, E., Grabherr, M., Forche, A., Reedy, J.L., et al. (2009). Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. *Nature* *459*, 657–662.
- Capra, J.A., Pollard, K.S., and Singh, M. (2010). Novel genes exhibit distinct patterns of function acquisition and network integration. *Genome Biol.* *11*, R127.
- Chen, C.G., Yang, Y.L., Shih, H.I., Su, C.L., and Lo, H.J. (2004). CaNdt80 is involved in drug resistance in *Candida albicans* by regulating CDR1. *Antimicrob. Agents Chemother.* *48*, 4505–4512.
- Coleman, D.A., Oh, S.H., Zhao, X., and Hoyer, L.L. (2010). Heterogeneous distribution of *Candida albicans* cell-surface antigens demonstrated with an Als1-specific monoclonal antibody. *Microbiology* *156*, 3645–3659.
- Costanzo, M.C., Arnaud, M.B., Skrzypek, M.S., Binkley, G., Lane, C., Miyasato, S.R., and Sherlock, G. (2006). The *Candida* Genome Database: facilitating research on *Candida albicans* molecular biology. *FEM. Yeast Res.* *6*, 671–684.
- de Hoon, M.J., Imoto, S., Nolan, J., and Miyano, S. (2004). Open source clustering software. *Bioinformatics* *20*, 1453–1454.
- Dujon, B., Sherman, D., Fischer, G., Durrens, P., Casaregola, S., Lafontaine, I., De Montigny, J., Marck, C., Neuvéglise, C., Talla, E., et al. (2004). Genome evolution in yeasts. *Nature* *430*, 35–44.
- Dusane, D.H., Nanchaiah, Y.V., Venugopalan, V.P., Kumar, A.R., and Zinjarde, S.S. (2008). Biofilm formation by a biotechnologically important tropical marine yeast isolate, *Yarrowia lipolytica* NCIM 3589. *Water Sci. Technol.* *58*, 2467–2475.
- Gavrias, V., Andrianopoulos, A., Gimeno, C.J., and Timberlake, W.E. (1996). *Saccharomyces cerevisiae* TEC1 is required for pseudohyphal growth. *Mol. Microbiol.* *19*, 1255–1263.
- Ghosh, S., Navarathna, D.H., Roberts, D.D., Cooper, J.T., Atkin, A.L., Petro, T.M., and Nickerson, K.W. (2009). Arginine-induced germ tube formation in *Candida albicans* is essential for escape from murine macrophage line RAW 264.7. *Infect. Immun.* *77*, 1596–1605.
- Gimeno, C.J., and Fink, G.R. (1994). Induction of pseudohyphal growth by overexpression of PHD1, a *Saccharomyces cerevisiae* gene related to transcriptional regulators of fungal development. *Mol. Cell. Biol.* *14*, 2100–2112.
- Green, C.B., Zhao, X., Yeater, K.M., and Hoyer, L.L. (2005). Construction and real-time RT-PCR validation of *Candida albicans* PALS-GFP reporter strains and their use in flow cytometry analysis of ALS gene expression in budding and filamenting cells. *Microbiology* *151*, 1051–1060.
- Hedges, S.B., Blair, J.E., Venturi, M.L., and Shoe, J.L. (2004). A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC Evol. Biol.* *4*, 2.
- Hernday, A.D., Noble, S.M., Mitrovich, Q.M., and Johnson, A.D. (2010). Genetics and Molecular Biology in *Candida albicans*. *Methods Enzymol.* *470*, 737–758.
- Hundal, H.S., and Taylor, P.M. (2009). Amino acid transceptors: gate keepers of nutrient exchange and regulators of nutrient signaling. *Am. J. Physiol. Endocrinol. Metab.* *296*, E603–E613.

- Igarashi, K., and Kashiwagi, K. (2010a). Characteristics of cellular polyamine transport in prokaryotes and eukaryotes. *Plant Physiol. Biochem.* **48**, 506–512.
- Igarashi, K., and Kashiwagi, K. (2010b). Modulation of cellular function by polyamines. *Int. J. Biochem. Cell Biol.* **42**, 39–51.
- Johnson, A.D., and Herskowitz, I. (1985). A repressor (MAT alpha 2 Product) and its operator control expression of a set of cell type specific genes in yeast. *Cell* **42**, 237–247.
- Liu, H., Styles, C.A., and Fink, G.R. (1993). Elements of the yeast pheromone response pathway required for filamentous growth of diploids. *Science* **262**, 1741–1744.
- Liu, H., Köhler, J., and Fink, G.R. (1994). Suppression of hyphal formation in *Candida albicans* by mutation of a STE12 homolog. *Science* **266**, 1723–1726.
- Lo, H.J., Köhler, J.R., DiDomenico, B., Loebenberg, D., Cacciapuoti, A., and Fink, G.R. (1997). Nonfilamentous *C. albicans* mutants are avirulent. *Cell* **90**, 939–949.
- Lohse, M.B., and Johnson, A.D. (2010). Temporal anatomy of an epigenetic switch in cell programming: the white-opaque transition of *C. albicans*. *Mol. Microbiol.* **78**, 334–434.
- Martchenko, M., Levitin, A., Hogues, H., Nantel, A., and Whiteway, M. (2007). Transcriptional rewiring of fungal galactose-metabolism circuitry. *Curr. Biol.* **17**, 1007–1013.
- Mitrovich, Q.M., Tuch, B.B., Guthrie, C., and Johnson, A.D. (2007). Computational and experimental approaches double the number of known introns in the pathogenic yeast *Candida albicans*. *Genome Res.* **17**, 492–502.
- Nett, J.E., Lepak, A.J., Marchillo, K., and Andes, D.R. (2009). Time course global gene expression analysis of an in vivo *Candida* biofilm. *J. Infect. Dis.* **200**, 307–313.
- Nobile, C.J., Solis, N., Myers, C.L., Fay, A.J., Deneault, J.S., Nantel, A., Mitchell, A.P., and Filler, S.G. (2008). *Candida albicans* transcription factor Rim101 mediates pathogenic interactions through cell wall functions. *Cell. Microbiol.* **10**, 2180–2196.
- Noble, S.M., and Johnson, A.D. (2005). Strains and strategies for large-scale gene deletion studies of the diploid human fungal pathogen *Candida albicans*. *Eukaryot. Cell* **4**, 298–309.
- Noble, S.M., French, S., Kohn, L.A., Chen, V., and Johnson, A.D. (2010). Systematic screens of a *Candida albicans* homozygous deletion library decouple morphogenetic switching and pathogenicity. *Nat. Genet.* **42**, 590–598.
- Pappas, P.G., Rex, J.H., Sobel, J.D., Filler, S.G., Dismukes, W.E., Walsh, T.J., and Edwards, J.E.; Infectious Diseases Society of America. (2004). Guidelines for treatment of candidiasis. *Clin. Infect. Dis.* **38**, 161–189.
- Remm, M., Storm, C.E., and Sonnhammer, E.L. (2001). Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J. Mol. Biol.* **314**, 1041–1052.
- Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., et al. (2000). Genome-wide location and function of DNA binding proteins. *Science* **290**, 2306–2309.
- Reuss, O., Vik, A., Kolter, R., and Morschhäuser, J. (2004). The SAT1 flipper, an optimized tool for gene disruption in *Candida albicans*. *Gene* **341**, 119–127.
- Saldanha, A.J. (2004). Java Treeview—extensible visualization of microarray data. *Bioinformatics* **20**, 3246–3248.
- Schweizer, A., Rupp, S., Taylor, B.N., Rölinghoff, M., and Schröppel, K. (2000). The TEA/ATTS transcription factor CaTec1p regulates hyphal development and virulence in *Candida albicans*. *Mol. Microbiol.* **38**, 435–445.
- Sellam, A., Askew, C., Epp, E., Tebbji, F., Mullick, A., Whiteway, M., and Nantel, A. (2010). Role of transcription factor CaNdt80p in cell separation, hyphal growth, and virulence in *Candida albicans*. *Eukaryot. Cell* **9**, 634–644.
- Sophianopoulou, V., and Diallynas, G. (1995). Amino acid transporters of lower eukaryotes: regulation, structure and topogenesis. *FEMS Microbiol. Rev.* **16**, 53–75.
- Staab, J.F., Ferrer, C.A., and Sundstrom, P. (1996). Developmental expression of a tandemly repeated, proline- and glutamine-rich amino acid motif on hyphal surfaces on *Candida albicans*. *J. Biol. Chem.* **271**, 6298–6305.
- Stoldt, V.R., Sonneborn, A., Leuker, C.E., and Ernst, J.F. (1997). 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.* **16**, 1982–1991.
- Storey, J.D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* **100**, 9440–9445.
- Teixeira, M.C., Monteiro, P., Jain, P., Tenreiro, S., Fernandes, A.R., Mira, N.P., Alenquer, M., Freitas, A.T., Oliveira, A.L., and Sá-Correia, I. (2006). The YEASTRACT database: a tool for the analysis of transcription regulatory associations in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **34** (*Database issue*), D446–D451.
- Tuch, B.B., Galgoczy, D.J., Hernday, A.D., Li, H., and Johnson, A.D. (2008). The evolution of combinatorial gene regulation in fungi. *PLoS Biol.* **6**, e38.
- Unal, E., Kinde, B., and Amon, A. (2011). Gametogenesis eliminates age-induced cellular damage and resets life span in yeast. *Science* **332**, 1554–1557.
- van het Hoog, M., Rast, T.J., Martchenko, M., Grindle, S., Dignard, D., Hogues, H., Cuomo, C., Berriman, M., Scherer, S., Magee, B.B., et al. (2007). Assembly of the *Candida albicans* genome into sixteen supercontigs aligned on the eight chromosomes. *Genome Biol.* **8**, R52.
- Wang, L., Feng, Z., Wang, X., Wang, X., and Zhang, X. (2010). DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* **26**, 136–138.
- Wapinski, I., Pfeffer, A., Friedman, N., and Regev, A. (2007). Automatic genome-wide reconstruction of phylogenetic gene trees. *Bioinformatics* **23**, i549–i558.
- Ward, M.P., Gimeno, C.J., Fink, G.R., and Garrett, S. (1995). SOK2 may regulate cyclic AMP-dependent protein kinase-stimulated growth and pseudohyphal development by repressing transcription. *Mol. Cell. Biol.* **15**, 6854–6863.
- Zhao, R., Lockhart, S.R., Daniels, K., and Soll, D.R. (2002). Roles of TUP1 in switching, phase maintenance, and phase-specific gene expression in *Candida albicans*. *Eukaryot. Cell* **1**, 353–365.



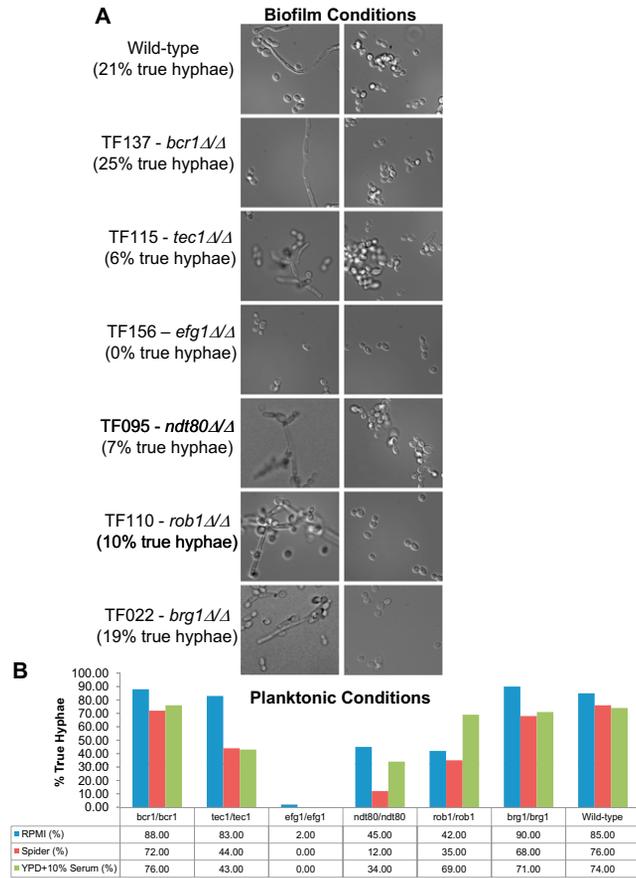
**Figure S1. Characterization of In Vitro Biofilm-Defective Mutants over Time, Related to Figure 1 and Table S1**

The wild-type reference strain SN250 and the six transcription regulator mutant strains were grown under biofilm conditions at the following time points: 0 hr, 8 hr, 24 hr, and 48 hr postadherence, and visualized by CSLM. The upper panel for each image shows the top view and the lower panel shows the side view. Scale bars represent 20  $\mu\text{m}$ .



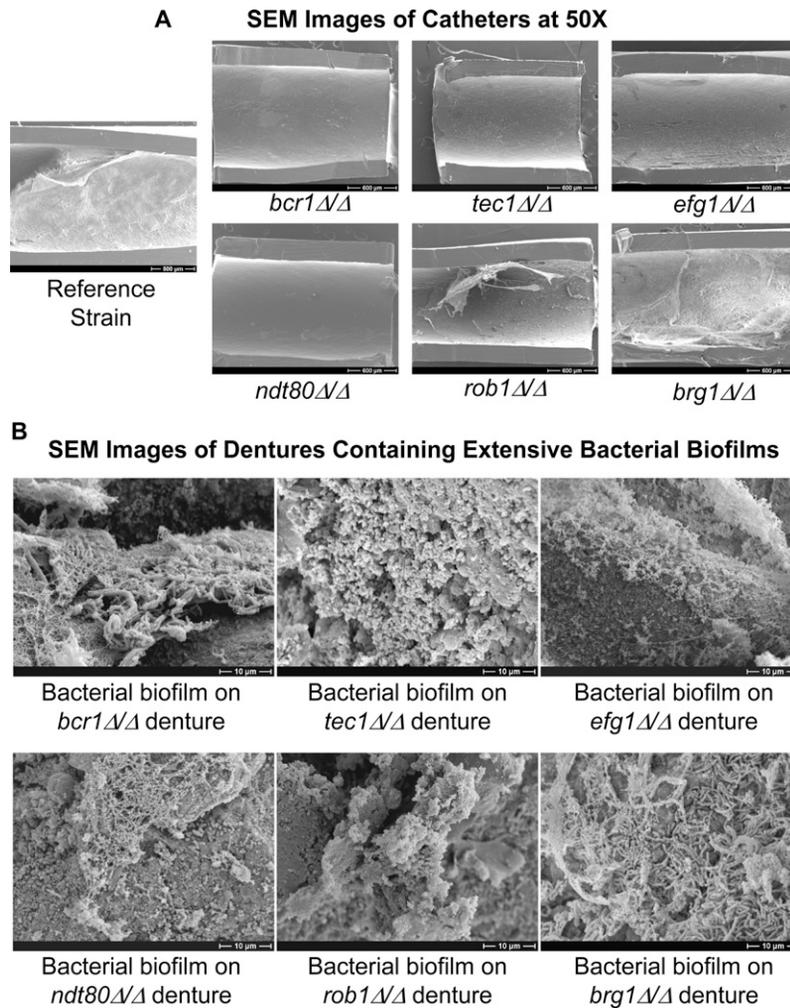
**Figure S2. Screening and Characterization of In Vitro Biofilm-Defective Transcription Regulator Mutants, Related to Figure 1 and Table S1**

Nine regulator mutant strains: TF022, TF091, TF095, TF103, TF110, TF115, TF117, TF137, and TF156 were initially found to have biofilm defects. Visual phenotypic characterization of the mutants is shown in panel A. Of these combined nine mutants, we did not follow up on TF091 (*orf19.1685Δ/Δ*) because its biofilm defect is due to a growth defect specific to Spider medium, TF103 (*orf19.3063Δ/Δ*) because its biofilm defect is the result of an overall growth defect, and TF117 (*tup1Δ/Δ*) because this mutant is highly pleiotropic and its phenotype interfered with the biofilm assay. Panel B shows CSLM side view images of the six complemented strains where a wild-type allele was added back into the six core regulator mutant strains. Reintroduction of an ectopic copy of the wild-type allele back into each mutant reversed the biofilm-formation defect of each mutant to near wild-type levels by CSLM (panel B, upper images), and to complete wild-type levels by the visual plate assay (panel B, bottom images). Scale bars represent 20  $\mu\text{m}$ .



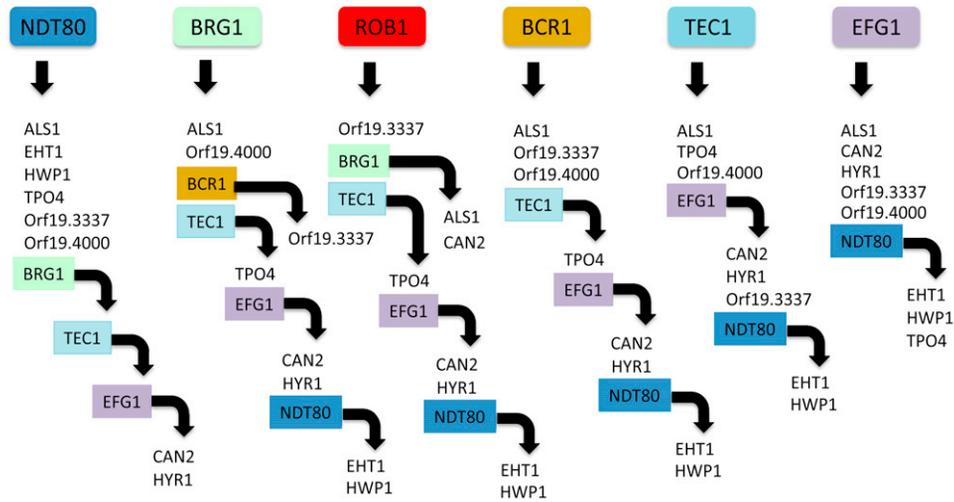
**Figure S3. Assessment of Hyphal Formation Abilities for the Six Core Biofilm-Defective Regulator Mutants, Related to Figure 1**

Images of true hyphae found suspended in the medium (surrounding the biofilm) under biofilm-forming conditions are shown in panel A. Panels on the left show evidence of hyphal formation; panels on the right are more representative of the entire suspended cell population. The percentage of true hyphae produced by each strain under the three indicated planktonic conditions is shown quantitatively in panel B. Hyphal defects have been previously reported for *efg1Δ/Δ*, *tec1Δ/Δ*, and *ndt80Δ/Δ* *in vitro* (Lo et al., 1997; Ramage et al., 2002; Schweizer et al., 2000; Sellam et al., 2010; Stoldt et al., 1997).



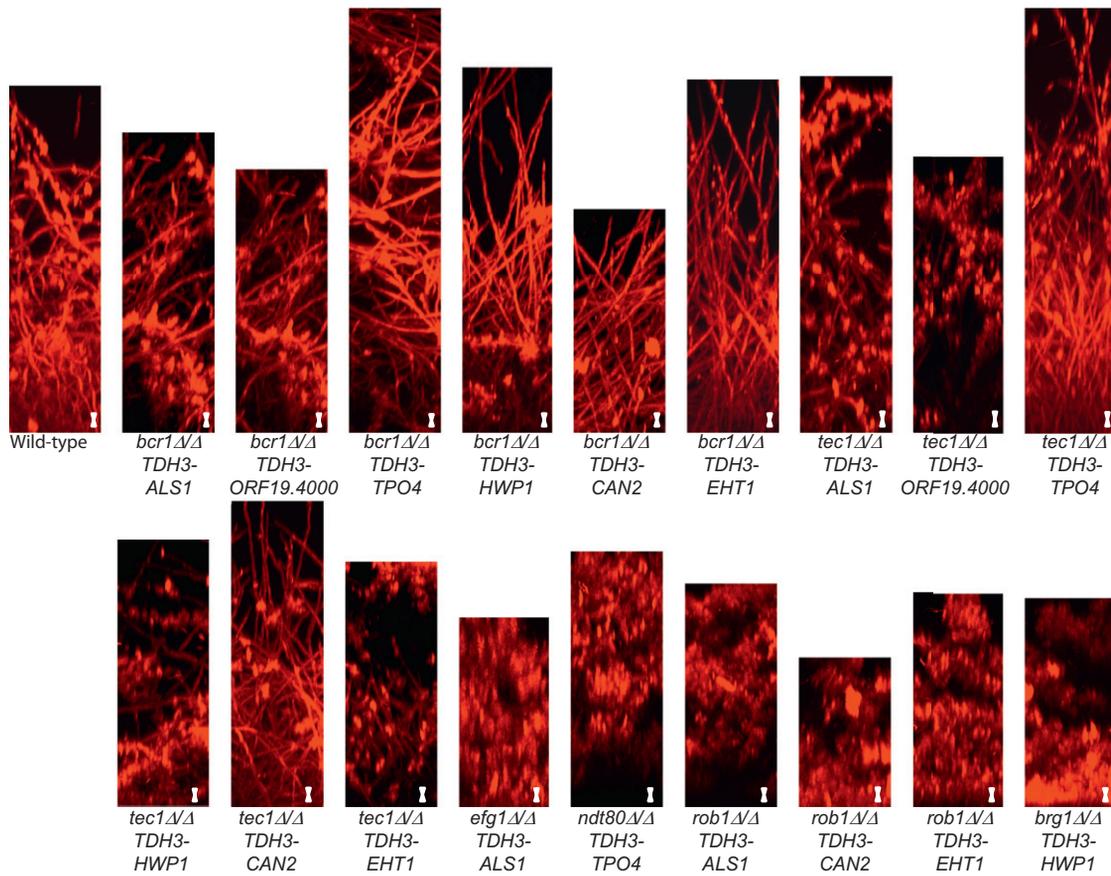
**Figure S4. Biofilm Formation In Vivo, Related to Figure 2**

Biofilm formation in the in vivo rat catheter model (panel A). The wild-type reference strain SN250, and the six transcription regulator mutant strains were inoculated into rat intravenous catheters, and the resulting biofilms were visualized after 24 hr of growth by SEM. These SEM images show catheter luminal surfaces at magnifications of 50X. Extensive bacterial biofilm formation in the in vivo rat denture model on the dentures of the six transcription regulator mutant strains (panel B). The regulator mutant strains were inoculated into rat dentures, and extensive bacterial biofilms were visualized on the denture surfaces after 24 hr of growth by SEM. SEM images show the denture surfaces at magnifications of 2,000x.



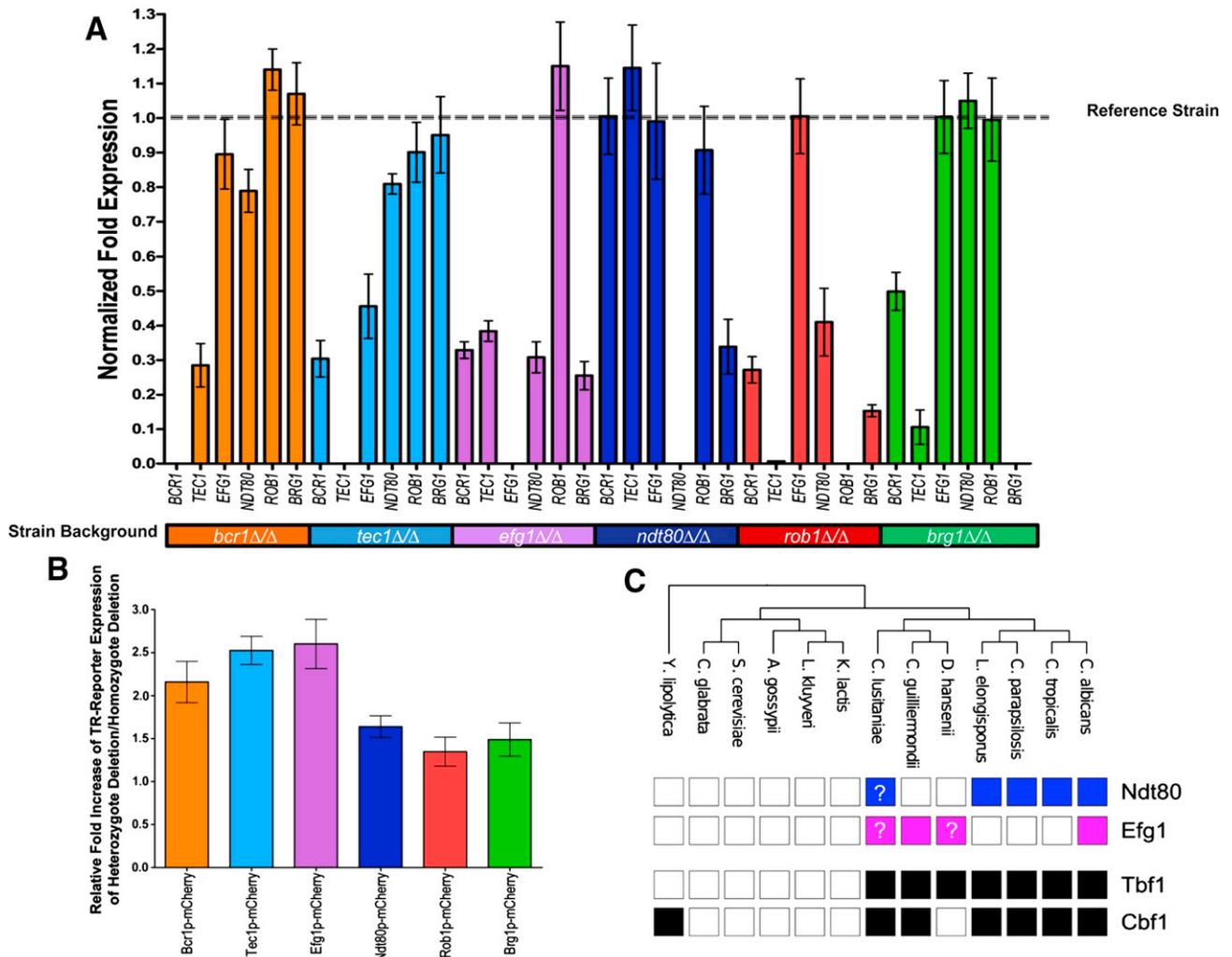
**Figure S5. Hierarchical Cascade of Target Gene Regulation, Related to Figure 5 and Table S3**

Deletion of any one of the six biofilm regulators results in downregulation of eight target genes; however the promoters of these genes are not bound by every regulator. Indirect regulation by each biofilm regulator can be explained by direct regulation by at least one of the other regulators. The diagram shows potential regulation cascades from each regulator to each target gene. The arrows represent direct regulation (binding and differential regulation as determined by ChIP-chip and gene expression microarray).



**Figure S6. The Target Gene Ectopic Expression Strains that Rescued the Defects of Their Corresponding Transcription Regulator Deletion Mutant, Related to Figure 6 and Table S3**

CSLM side view images of the wild-type and the eighteen target gene ectopic expression strains that exhibited a rescuing phenotype are shown. Scale bars represent 20  $\mu\text{m}$ .



**Figure S7. Validation of Transcriptional Regulation of the Biofilm Regulators and Enrichment of the Regulator Motifs in Yeast, Related to Figure 7 and Table S3**

Validation of transcriptional regulation of the biofilm regulatory network by quantitative real time PCR (qPCR) (panel A). Expression levels of the six biofilm regulators (*BCR1*, *TEC1*, *EFG1*, *NDT80*, *ROB1*, and *BRG1*) in the background of each regulator mutant under biofilm conditions are shown in panel A. Normalized gene expression values were calculated using the  $\Delta\Delta Ct$  method using *TAF145* as a reference gene. Results are the means of three determinations. For ease of interpretation, the reference strain expression level values were set to 1.0 for each gene set, and the normalized expression level of each gene relative to *TAF145* expression is shown. Validation of transcriptional regulation of the biofilm regulators on their own promoters (panel B). Transcriptional reporters of the biofilm transcriptional regulators (TRs) were constructed where one copy of each regulator's promoter was fused to an mCherry transcriptional reporter in both a homozygous regulator mutant as well as a heterozygous regulator mutant. Normalized expression levels of these reporter strains under biofilm conditions were calculated using the  $\Delta\Delta Ct$  method using *TAF145* as a reference gene. Means of three determinations was determined. The relative fold increase of expression in the TRp-mCherry in the heterozygote deletion strain to its respective TRp-mCherry in the homozygote deletion strain is shown in panel B. Enrichment of regulator motifs across yeast species (panel C). Motifs for several biofilm regulators and two control regulators governing ribosomal biogenesis were scored across thirteen yeast species. Known targets of each regulator were mapped to orthologs in other species, and enrichment of the motif was determined relative to the rest of the genome for each species. Ndt80 and Efg1 were the only biofilm regulators with motifs containing sufficient predictive power in *C. albicans* to be scored across the other yeast species. As a positive control, we analyzed conservation of motifs upstream of two ribosomal protein genes (Tbf1 and Cbf1) from Lavoie et al., 2010. A cutoff for the motif score from MAST was selected to maximize the log-ratio of each regulator in *C. albicans* and was then used for the other species. Colored squares indicate significant enrichment of the regulator motif. Question marks indicate borderline significance.