Mapping the transcriptional regulatory network of a fungal pathogen by exploiting transcription factor perturbation

Dhoha Abid[1], Holly Brown[1], Chase Mateusiak[1], Tamara L. Doering[4], and Michael R. Brent[1,2,3,4,5.]

1. Department of Computer Science and Engineering, Washington University, St. Louis, MO 63110
2. Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, MO 63110
3. Department of Genetics, Washington University School of Medicine, St. Louis, MO 63110
4. Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, MO 63110
5. Corresponding author

Michael Brent
brent@wustl.edu
http://mblab.wustl.edu
314-660-2205
Campus Box 8510
Washington University
Saint Louis, MO 63130

Abstract

Cryptococcus neoformans is a deadly fungal pathogen. Upon entering a mammalian host, it deploys a voluminous polysaccharide capsule that is necessary for it to survive host defenses and maintain an infection. Capsule expansion is regulated transcriptionally, as deletion of many transcription factors alters capsule. Thus, we set out to map the transcriptional regulatory network of *C. neoformans* – that is, to identify the TFs that directly regulate each gene in the genome. First, we carried out RNA-seq of 120 single-TF-deletion strains, together with wild-type controls. We then applied NetProphet3, a TF network mapping algorithm, to predict the direct functional targets of each TF. Unexpectedly, analysis of this network indicated that there are no TFs that primarily regulate genes involved in capsule formation. Rather, the TFs that play a role in deploying capsule also regulate many other genes and processes. Comparison to a TF network map we built for *Saccharomyces cerevisiae*, a distantly related model yeast, identified pairs of TFs that are functionally orthologous – that is, their targets are enriched for orthologous genes. In many cases, these pairs are different from the ones identified by sequence homology alone. We suggest that network analyses should be used to complement sequence comparison when searching for functionally orthologous transcription factors. Our network map can be searched and visualized at http://cryptococcus.net.

Introduction

Cryptococcus neoformans is a fungal pathogen that occurs in the environment and may infect mammals, including humans and mice. When infectious particles of *C. neoformans* are inhaled, they can initiate a lung infection. In immunocompetent individuals this is generally cleared, although latent infection may occur. In severely immunocompromised individuals, however, the infection persists in the lungs and frequently disseminates to the brain, where it causes potentially lethal meningitis (Meya and Williamson, 2024; Tugume, et al., 2023). *C. neoformans* is estimated to be the second most common cause of AIDS-related mortality (Rajasingham, et al., 2017) and responsible for 147,000 deaths annually (Denning, 2024; Rajasingham, et al., 2017) .

C. neoformans is a basidiomycete, belonging to a phylum that diverged approximately 400 million years ago from the ascomycete lineage, which includes the model yeast *Saccharomyces cerevisiae* and other fungal pathogens like *Candida* and *Aspergillus* (Taylor and Berbee, 2006). It is distinguished by a large polysaccharide capsule that surrounds the cell and is central to its virulence (Boodwa-Ko and Doering, 2024). Upon entry into a host or exposure to conditions that mimic the host environment, the capsule expands considerably (Agustinho, et al., 2018; Wang, et al., 2018). Acapsular cells are avirulent (Fromtling, et al., 1982; Kwon-Chung and Rhodes, 1986), while larger capsules increase resistance to engulfment by phagocytes during infection (Bulmer and Sans, 1967; Kozel, 1977; Srikanta, et al., 2011).

We know that capsule production is regulated at the level of transcription (Gish, et al., 2016; Granger, et al., 1985), yet our understanding of this regulation is limited. One way to define regulatory processes is to generate a transcription factor (TF) network map (Bonneau, et al., 2006; Greenfield, et al., 2010; Huynh-Thu and Geurts, 2018; Jackson, et al., 2020; Madar, et al., 2010; Margolin, et al., 2006; Roy, et al., 2013; Siahpirani and Roy, 2017). Such maps describe the relationships between each TF in the genome and the genes that it regulates. A map consists of nodes and edges: nodes represent TFs and genes, and an edge exists between a TF and a target gene if the TF regulates that gene. Edges are described as: TF-target.

82 Many gene regulatory networks are intended to represent functional regulation rather than
83 molecular interactions. In other words, edges indicate relationships between genes (or TFs and
84 genes) that could be either direct or indirect. Here, we aim to generate a map that represents
85 relationships in C. *neoformans* that are both direct and functional. In other words, the TF must
86 physically bind to the regulatory DNA of the target gene (so it is direct), and as a result, it must
87 modulate the transcription rate of that target (so it is functional). To this end, we used
88 NetProphet3, a state-of-the-art method for mapping TF networks that we recently developed
89 (Abid and Brent, 2023). NetProphet3 uses as input gene expression profiles of cells with or
90 without perturbation of a TF by, for example, deletion or overexpression. Because they are
91 based on expression data, NetProphet3 maps are functional. By studies of the model yeast
92 *Saccharomyces cerevisiae*, for which genome-wide datasets on TF binding locations are
93 available, we have also shown that these maps are highly enriched for direct interactions (Abid
94 and Brent, 2023).

95 In this work, we generated a comprehensive dataset of gene expression changes in
96 response to TF perturbation in *C. neoformans*. Specifically, we carried out RNA-Seq on 120 TF
97 deletion strains and used the resulting data as input for NetProphet3 to create a TF network
98 map. The map assigns scores to all possible TF-target interactions that represent our confidence
99 in the existence of a direct, functional relationship. After validating this map, we used it to
100 identify functional orthologs of Cryptococcus TFs in *S. cerevisiae*. We also searched the network
101 for TFs that primarily regulate genes involved in capsule formation, but found none, suggesting
102 that capsule is not deployed independently but rather as part of an arsenal of cellular defense
103 and stress response systems. Our new TF perturbation response dataset and the network map
104 we constructed from it will be made available to the research community.

Results

Overview of the transcriptome

We generated gene expression data for 120 strains, each lacking one TF (Supplementary File S1), which we made previously (Maier, et al., 2015) or obtained from the Madhani *C. neoformans* deletion collection (Boucher et al, bioRXiv). We grew these, along with wild-type (WT) controls, in standard laboratory conditions for rich medium growth (YPD, room air, 30°C). We then transferred them to capsule-inducing conditions for 90 minutes and measured gene expression levels by RNA-Seq. The capsule-inducing conditions we used, which are designed to mimic the host environment, include mammalian tissue culture medium (DMEM), 37 °C, and 5% CO_2; these conditions are known to result in cells with large capsules (Maier, et al., 2015). We chose a 90-minute time point to reflect early changes in transcription and because it has yielded insights into cryptococcal gene regulation in our previous studies of TF deletion strains (Gish, et al., 2016; Maier, et al., 2015).

To enable rigorous analyses, we carried out RNA-Seq in a strictly controlled manner and with multiple biological replicates for each TF deletion strain (average = 4, Fig. 1A). We also included WT control cultures with each experiment, totaling 122 samples. Under our conditions, we detected expression of 6746 genes, which included 163 of the 165 TFs in *C. neoformans*. Figure 1B shows the distribution of expression levels for all TFs in WT cells, while Supplementary Figure S1A provides the values for individual TFs. The most highly expressed TF genes were *RIM101* (O'Meara, et al., 2014) and *BZP1* (Jung, et al., 2015). Most TFs were measurably expressed in capsule-inducing conditions (Figure 1B), so we expected that deleting the corresponding genes could alter the expression of their targets.

Next, for each TF deletion strain we calculated the number of differentially-expressed (DE) genes, which may be functional targets of the TF. A gene was defined as DE if the adjusted P-value of expression in the deletion versus WT was below 0.01 and its absolute $\log_2$ fold-change (FC) relative to WT was greater than 0.4 (Fig. 1C and Supplementary Fig. S1B). Half of the TF-deletion strains had 1-20 DE genes, suggesting that they have relatively specific functions in

these conditions, while the other half had more than 20. Overall, the expression level of a TF in the WT strain was not correlated with the number of genes that were DE in the TF deletion (Pearson correlation $\rho = 0$). However, when we plotted the number of genes that responded to a TF deletion against that TF's WT expression level, TFs clustered into three groups (Fig. 1D). TFs in the blue cluster regulated few to no genes, despite the broad range of expression levels of those TFs in WT cells. In the green cluster, the expression level of each TF correlated with the number of genes that were DE when that TF was deleted (Pearson correlation $\rho = 0.57$, $P < 0.01$). TFs in the red cluster were expressed at a relatively high level and had an intermediate number of DE genes (54-381). This group was enriched for TFs where the corresponding deletion strain has an abnormal capsule phenotype (Hypergeometric p = 0.0007), whereas the other clusters were not (hypergeometric p = 0.97 and p = 0.98; see Supplemental Methods for details).

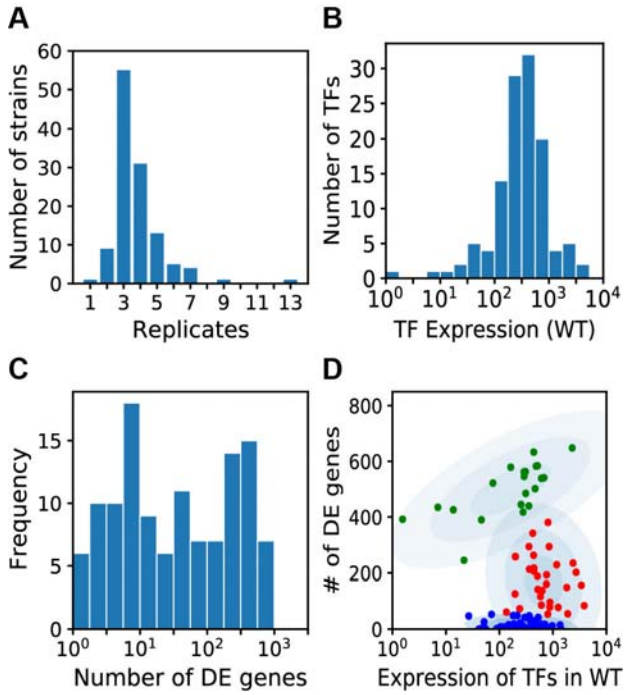

Fig 1. (A) Histogram of the number of biological replicates obtained for TF deletion strains. (B) Histogram of normalized read counts of the expression levels of TFs in WT cells. Only the 120 TFs that were deleted were included. (C) Histogram of the number of DE genes in TF deletion strains. (D). The number of genes differentially expressed in response to TF deletions as a function of WT expression levels of the deleted TFs. The red cluster was enriched for TFs implicated in capsule formation.

A TF network map for *Cryptococcus neoformans*

We used NetProphet3 (Abid and Brent, 2023) to construct a TF network map for *C. neoformans*. The goal of NetProphet3 is to predict direct and functional TF-target relationships, in other words, those where a TF binds the promoter of a gene and thereby influences its expression. It achieves this by using a machine learning (ML) algorithm (XGBoost) to learn patterns in gene expression data that are indicative of direct regulation. ML algorithms are first trained to make predictions on cases where the true answer is known. They are then applied to make predictions where the true answer may not be known. During training, NetProphet3 uses available TF binding location data, such as from ChIP-Seq experiments, to indicate whether binding of the TF to a gene promoter is known to occur; this is assigned as a label, with a value

155 of 1 (binding) or 0 (no binding) (Fig. 2A). In the prediction phase this information is no longer
156 available, so only the expression levels of a gene after a TF perturbation are used to predict
157 whether the TF binds its promoter region. In both phases, the gene expression data used to
158 make predictions is first processed in various ways, producing the actual inputs to NetProphet3
159 (called features). Specifically, the features include the fold-change in expression of the target
160 when the TF is perturbed (DE feature) and the relationships between the expression of TFs and
161 the expression of target genes (LASSO and BART features). LASSO and BART are regression
162 algorithms that produce models for predicting the expression of one target at a time from the
163 expression of all TFs. We use these predictive models to extract the features that represent the
164 relationship between the expression of a TF and the expression of a potential target, across all
165 perturbation samples.

166 When the trained NetProphet3 model is applied to gene expression data, it produces a
167 binding probability for each TF-gene pair (Fig. 2A, Predictions), which is interpreted as the
168 probability that the gene is a direct, functional target of that TF. The predicted probability is
169 high if NetProphet3 identifies patterns in the gene expression data that suggest the TF binds in
170 the gene's promoter.

171 NetProphet3 has different modes, depending on whether there is experimental evidence of
172 TF binding. For TFs with binding location data, we constructed a network using the NetProphet3
173 integration mode (Fig. 2A). This predicts a score for a TF-gene edge by integrating *C.*
174 *neoformans* binding data and gene expression (Abid and Brent, 2023). This integration is
175 important because many physical interactions are not functional, i.e. a TF may bind to the
176 promoter of a gene without affecting its expression.

177 For TFs without binding data in *C. neoformans*, we used a cross-trained mode. In this case,
178 we trained NetProphet3 on *S. cerevisiae* binding and expression data (see Methods for details).
179 We then used this trained model on our data from gene expression in *C. neoformans* TF
180 deletion strains. Finally, we combined the predicted scores from both modes into a single
181 network that includes all TF-gene pairs (Fig. 2A, 'Predictions'; Supplementary File S3).

The *C. neoformans* network map consists of direct, functional interactions

We used two evaluation metrics to assess whether high scoring TF-target edges are direct and functional. One of these is based on binding data and one is based on gene function annotation. We defined a binding evaluation metric as the fraction of predicted TF-target edges that are supported by physical binding data (Abid and Brent, 2023) and used it to evaluate the cross-trained network. We performed this evaluation for the 10 *C. neoformans* TFs with associated physical binding data: Cir1 (Do, et al., 2020), Cuf1 (Garcia-Santamarina, et al., 2018), Gat201 (Homer, et al., 2016), Gat204 (Homer, et al., 2016), Hap-X (Do, et al., 2020), Nrg1 (Gish, et al., 2016; Maier, et al., 2015), Liv3 (Homer, et al., 2016), Pdr802 (Reuwsaat, et al., 2021), Sre1 (PRJNA557210), and Usv101 (Gish, et al., 2016; Maier, et al., 2015).

True TF-target edges represent a small fraction of the universe of possible TF-gene edges. For this reason, we evaluated only the top scoring TF-gene edges. To do this we first ranked all TF-gene where the TF was one of the 10 that had binding data, from highest to lowest scores. We then generated networks that included a total number of edges equal to various multiples of the number of TFs. The most stringently thresholded network retained only 100 edges, equal to ten times the number of TFs with binding data (on average 10 per TF, although different TFs have different numbers of targets). Finally, we plotted the fraction of edges with binding support as a function of the network threshold (Figure 2B).

Because we ranked edges by score, the networks with fewer targets per TF had higher average edge scores and were expected to have greater binding support. Indeed, for the smallest network, over 50% of predicted edges were supported by binding data, compared to a random expectation of 4% (Fig. 2B). Since ChIP-seq studies can miss true interactions, this is an excellent outcome and gives us high confidence in the remaining predictions.

As expected, increasing the network size led to reduction of the fraction of edges supported by binding data (Fig. 2B). This shows that higher scoring edges are more likely to be direct. Furthermore, the performance of the *C. neoformans* network with the binding metric is comparable to that of the *S. cerevisiae* network produced by NetProphet3 (Abid and Brent,

210 2023), indicating that the NetProphet3 method developed using *S. cerevisiae* is equally
211 applicable to *C. neoformans*.

212 The second evaluation metric we applied was based on gene ontology (GO) biological
213 process annotations. This GO metric reflects the degree to which each TF's targets are enriched
214 for genes that share a common functional annotation and are therefore biologically coherent.
215 For each TF and each biological process annotation, we calculated a hypergeometric p-value
216 (see Methods for details). A significant p-value indicates that more of the TF's target genes are
217 annotated with the function than would be expected by chance, and hence the TF's targets are
218 functionally coherent. We downloaded *C. neoformans* annotations from the UniProt website,
219 most of which were transferred from *S. cerevisiae* based on sequence similarity (see Methods
220 for details). Overall, 799 GO biological processes were transferred to *C. neoformans*, annotating
221 over 3,000 genes, and the target sets of 30 TFs in our network were significantly enriched for at
222 least one GO biological process term (Supplementary Table S1). To evaluate statistical
223 significance, we created 30 random networks by permuting TFs and targets and evaluated them
224 using the same metric. Although there is some fluctuation in absolute level, the performance of
225 our network significantly exceeded those of the random networks at every threshold, indicating
226 that each TF's targets in our network were much more functionally coherent than would be
227 expected by chance (Fig. 2C). This further supports the validity of the network.

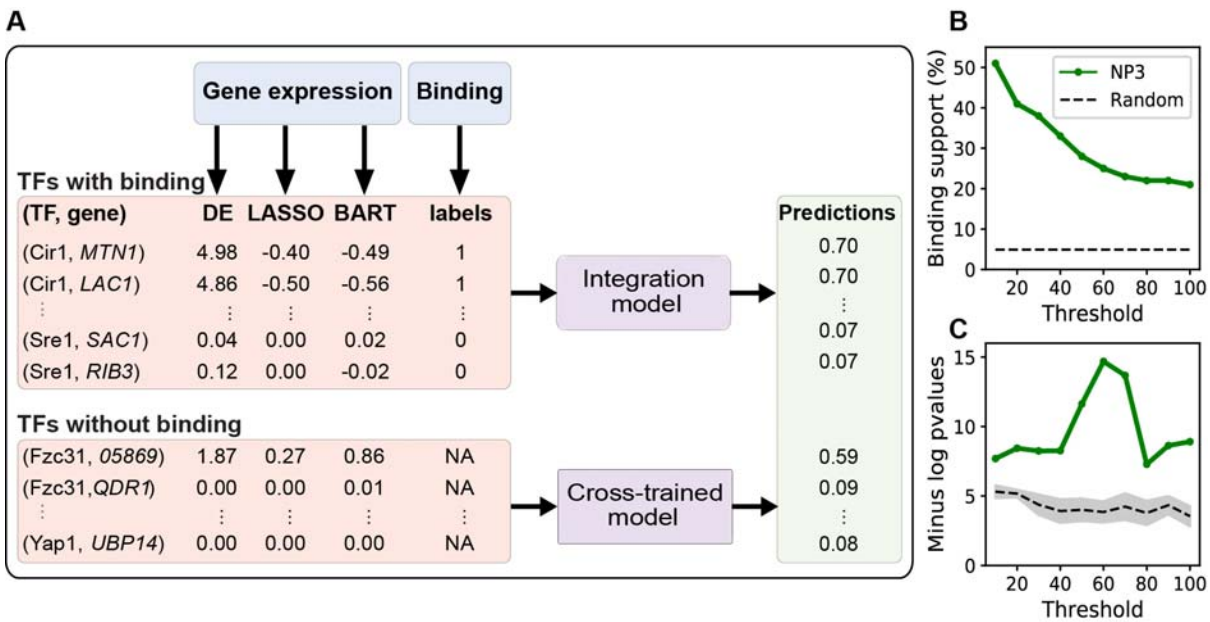

Fig 2. (A) Diagram of the NetProphet3 data flow, as described in the text. Output, final edge scores. (B-C) Performance of NetProphet3 (green) versus random expectation (black dashed line). Threshhold, average number of target genes per TF. (B) Evaluation of the cross-trained mode using the binding metric. Only TF-target edges whose TFs have binding data were evaluated. (C) Evaluation using the GO metric; values plotted are the median across all TFs for the most significant GO term. Shaded area, 95% confidence interval for 30 random networks.

The most influential TFs are enriched for those that are known to regulate capsule.

NetProphet3 outputs probability scores for all possible TF-gene edges, most of which have very low scores. When selecting a set of predicted edges for further analysis, there is a tradeoff between including only edges predicted with high confidence and including enough edges to characterize the organism's network comprehensively. As a reasonable compromise, for further analysis we chose a set that included on average 100 targets per TF. This set consists of all edges scoring above 0.14 (Fig. 3A, dashed line).

We generated or obtained deletion strains lacking 120 of the 165 TFs in the Cryptococcus genome. For the other TFs, which may be essential, NetProphet3 still predicts targets. It does this by using non-DE features obtained by analyzing all gene expression profiles together (Fig. 2A, LASSO and BART). We therefore had predictions for all 163 TFs that were expressed in our dataset; our network included the 16,300 highest-scoring TF-target edges. 161 of the TFs had at least one target gene under the conditions assayed. The total number of regulated genes

243 (4,103) corresponds to 60% of those that were measurably expressed in our studies and 49% of
244 all *C. neoformans* genes. This TF-target network, provided in Supplemental File S3, is the basis
245 for our analyses below.

246 The more targets a TF has, the more influence it can potentially exert on the transcriptional
247 state of the cell. In our network, the number of direct targets of each TF ranged from 7 to 884
248 (Fig. 3B). When we ranked TFs by the number of targets, the top 16 TFs (10%; Supplementary
249 File S8) covered 6,709 (41%) of the 16,300 TF-target edges in our network. Surprisingly, these
250 16 TFs regulate 81% of the 4,103 genes in the network, showing their disproportionate impact.
251 We were interested to notice that 10 of these TFs are known to regulate the capsule, based on
252 the phenotypes of corresponding deletion strains. This is significantly enriched over the fraction
253 of all TFs that are known to regulate the capsule (63% versus 31%; hypergeometric $p = 0.006$;
254 Supplementary File S8; see Discussion).

255 In addition to targets per TF (Fig. 3B), we examined the distribution of TFs per target gene
256 (Fig. 3C). The distribution peaked at one regulator per target gene and dropped off
257 exponentially, with one outlier gene having 55 regulators (CNAG_04126, not plotted in Fig. 3C).
258 When we ranked genes by this metric, we found that the top 10% of targets (410) participated
259 in 33% (5,315) of the TF-target edges. On average these targets have 13 regulators; together,
260 these regulators include all 161 TFs in the network (Supplementary File S8). Also, these 410
261 targets were generally enriched for capsule-implicated genes (those whose deletion perturbs
262 capsule; hypergeometric $p < 3 \times 10^{-6}$); they were specifically enriched for genes that encode
263 capsule-implicated TFs (hypergeometric $p < 10^{-6}$).

264 Many TFs are both powerful regulators and highly regulated, thus occupying central
265 positions in transcriptional networks. We defined hub TFs as those that are in the top 10% both
266 by the number of TFs that regulate them and by the number of genes they regulate. Our
267 network contained 14 hub TFs (Table 1), of which eight are known to regulate capsule
268 (hypergeometric $p = 0.007$; see Discussion).

The *C. neoformans* network map is biologically coherent

We took advantage of the abundant available data on capsule phenotypes to assess whether the network makes biological sense. For this, we first created a capsule subnetwork composed of all TF-target edges for which both the TF and target are known to influence capsule phenotype. This subnetwork includes 113 target genes, of which 49 encode TFs, linked by 318 TF-target edges.

TF	# of targets	# of regulators	Capsule
Bzp5	884	18	
Nrg1	692	11	✓
Gat201	567	19	✓
Swi6	509	11	✓
05375	475	13	
Fzc51	402	14	
Usv101	402	12	✓
00514	383	11	
Cir1	334	12	✓
Rim101	328	15	✓
Ccd6	305	11	
Rds2	284	16	✓
Sp1	282	21	✓
Ert1	231	15	

Table 1. Hub TFs, defined as the intersection of the top 10% of TFs ranked by number of direct targets with the top 10% of genes ranked by the highest number of direct TFs. The table includes the common name of the TF or CNAG identifier when it is not available, the number of direct targets, the number of direct regulators, and whether the TF is known to regulate the phenotype of capsule.

Focusing on the capsule subnetwork, we investigated whether the direction of regulation (activation or repression) was consistent with the capsule phenotypes of a TF and its target gene. We considered an edge consistent when (1) the TF activates the target, and both have the same capsule phenotype (hypercapsular or hypocapsular) or (2) the TF represses the target and they have opposite capsule phenotypes. In Figure 3D, we show the 10 highest-scoring TF-target edges of the subnetwork; notably, many of the targets themselves encode TFs. All the edges shown were consistent with the capsule phenotype (binomial $p = 0.02$) except for two: Gat201-*PDR802* and Gat201-*TDA10*. In the case of *TDA10*, the negative regulation by Cir1 may be stronger than the positive regulation by Gat201; it cannot conform with both.

We speculated that target genes regulated by the highest number of capsule TFs (those whose corresponding deletion strain has altered capsule) were most likely to have a capsule phenotype upon deletion. To test this, we ranked genes by the number of capsule TFs that regulate them from highest to lowest and calculated the percentage of genes that have a capsule phenotype at different thresholds (Fig. 3E). We found that genes regulated by the

highest number of capsule TFs were much more likely to be capsule genes than those that were regulated by a lower number of capsule TFs: 9 out of the top 100 genes in this ranking were capsule genes (those whose deletion results in altered capsule), while only 2% of all genes have a known capsule phenotype (hypergeometric $p = 0.02$).

TF-TF pairs that share many targets.

We previously showed that TFs that share large fractions of their targets are more likely to work together as a physical protein complex than those that do not (Abid and Brent, 2023). TF-TF pairs that share many targets but do not form a physical complex may cooperate by other mechanisms, such as being co-regulated by a third factor or binding the same DNA sequences. To gain insight into the TF-TF relationships of capsule TFs, we calculated the Jaccard similarity between target sets of all capsule-implicated TF pairs in the network (Supplementary File S9). The Jaccard similarity is the number of targets in common divided by the total number of individual targets regulated by the two TFs. It is equal to one if all targets are shared by both TFs, and zero if none are shared. For each TF-TF pair, we determined whether this similarity was significantly higher than would be expected by chance (hypergeometric adjusted $p < 0.05$), indicating that the TFs are likely to cooperate (as defined above). For this analysis, we used TF-target edges that had high NetProphet3 probability scores (greater or equal to 0.2). From these, we generated a new interaction map that links TF pairs that are enriched for shared targets. This TF-TF interaction map includes 43 capsule TFs that are linked by 56 TF-TF edges.

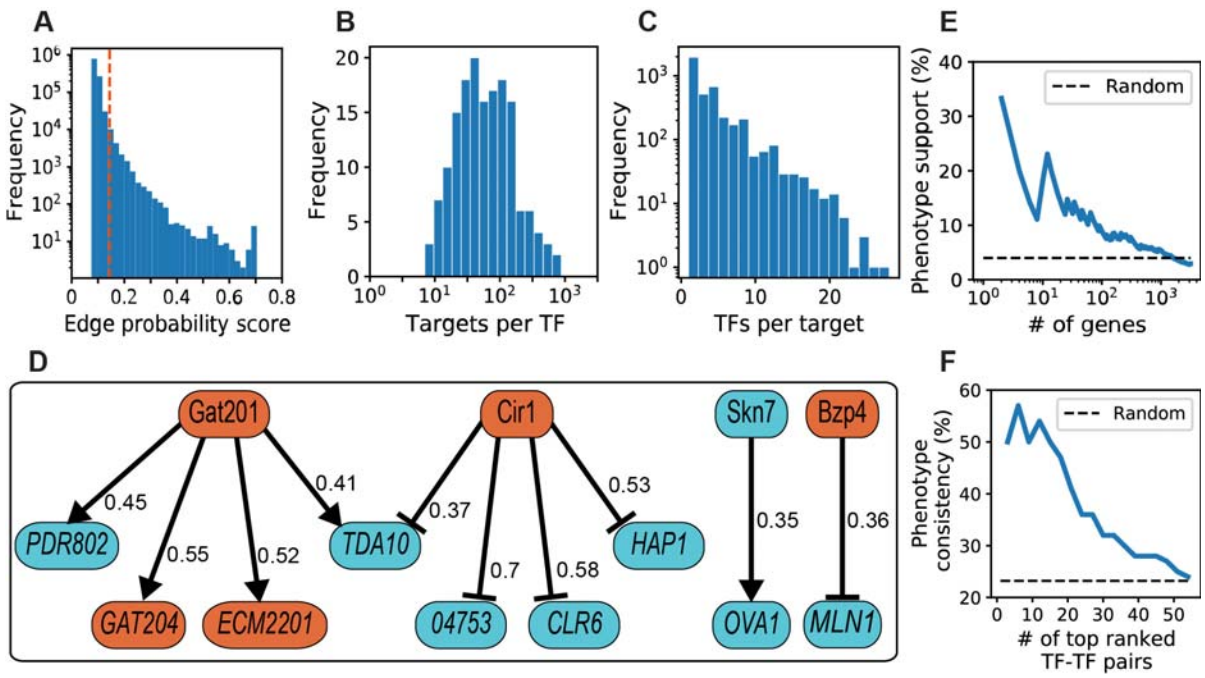

Fig 3. (A) NetProphet3 probability scores for all TF-gene edges. The vertical dashed orange line indicates the score threshold corresponding to 100 TF-target edges per TF on average. (B) The number of targets per TF. (C) The number of direct TF regulators per gene. (D) The top 10 high scoring TF-target edges in the capsule subnetwork. Colors indicate the phenotype of strains deleted for the corresponding genes: orange for hypocapsular and cyan for hypocapsular. Arrows denote activation and T-heads denote repression. All target genes except for *TDA10* and *CNAG_04753* encode DNA-binding proteins. The Gat201, Gat204 edge has been previously detailed (Chun, et al., 2011). (E) Genes were ranked by the number of capsule TFs that regulate them and the percentage of genes where the deletion strain exhibits a capsule phenotype was plotted for various cutoffs. (F) Percent of TF-TF pairs ranked above the indicated threshold that share a target and show phenotype consistency. Dashed line, percent of all TF-TF pairs that show phenotype consistency.

Next, we investigated how capsule TF-TF pairs with the most significant Jaccard similarity (P-adj < 10^{-5}) regulated their shared targets. We observed that these TF pairs tended to regulate their targets in the same direction if the two TFs had the same capsule phenotype and in the opposite direction if the TFs had opposite capsule phenotypes. If two TFs had the same capsule phenotype and at least 80% of their shared targets were regulated in the same direction, or if they had opposite phenotypes and at least 80% of their shared targets were regulated in the opposite direction, we considered the TF-TF pair to display phenotype consistency. We ranked TF-TF pairs by the significance of their Jaccard similarity and calculated the fraction of pairs above different rank thresholds that display phenotype consistency (Fig. 3F). This fraction was much higher than would be expected at random and tended to decline as the Jaccard similarity

327 of their target sets declined, consistent with the notion that pairs with the highest target
328 overlap work in a coordinated fashion. A list of these TF-TF pairs is presented in Supplementary
329 File S9.

330 Relationships between cryptococcal TFs implicated in capsule have been reported in several
331 contexts, including some efforts to map regulatory interactions (Chun, et al., 2011; Gish, et al.,
332 2016; Haynes, et al., 2011; Kumar, et al., 2011; Maier, et al., 2015). However, these have
333 generally been limited to relatively small groups of genes and focus on direct regulation of one
334 TF by another. Here we focus on the overlap of targets among capsule TFs. Figure 4 shows all
335 pairs of capsule TFs that share many more targets in common than would be expected by
336 chance ($p < 10^{-05}$), along with the targets that they share. Notably, half of these TFs are hub TFs,
337 highlighting their central role in transcriptional regulation. While most of the regulatory
338 relationships are activating, several TFs (Usv101 and Cir1) are uniformly repressive in this
339 context. Fine tuning in this system appears to be enabled by multiple regulators on single
340 genes, for example in the case of 00735 (lower right), which is regulated by five independent
341 TFs, or *DIP5, CT1* (top right), and 02198 (upper left), which are each regulated by four.

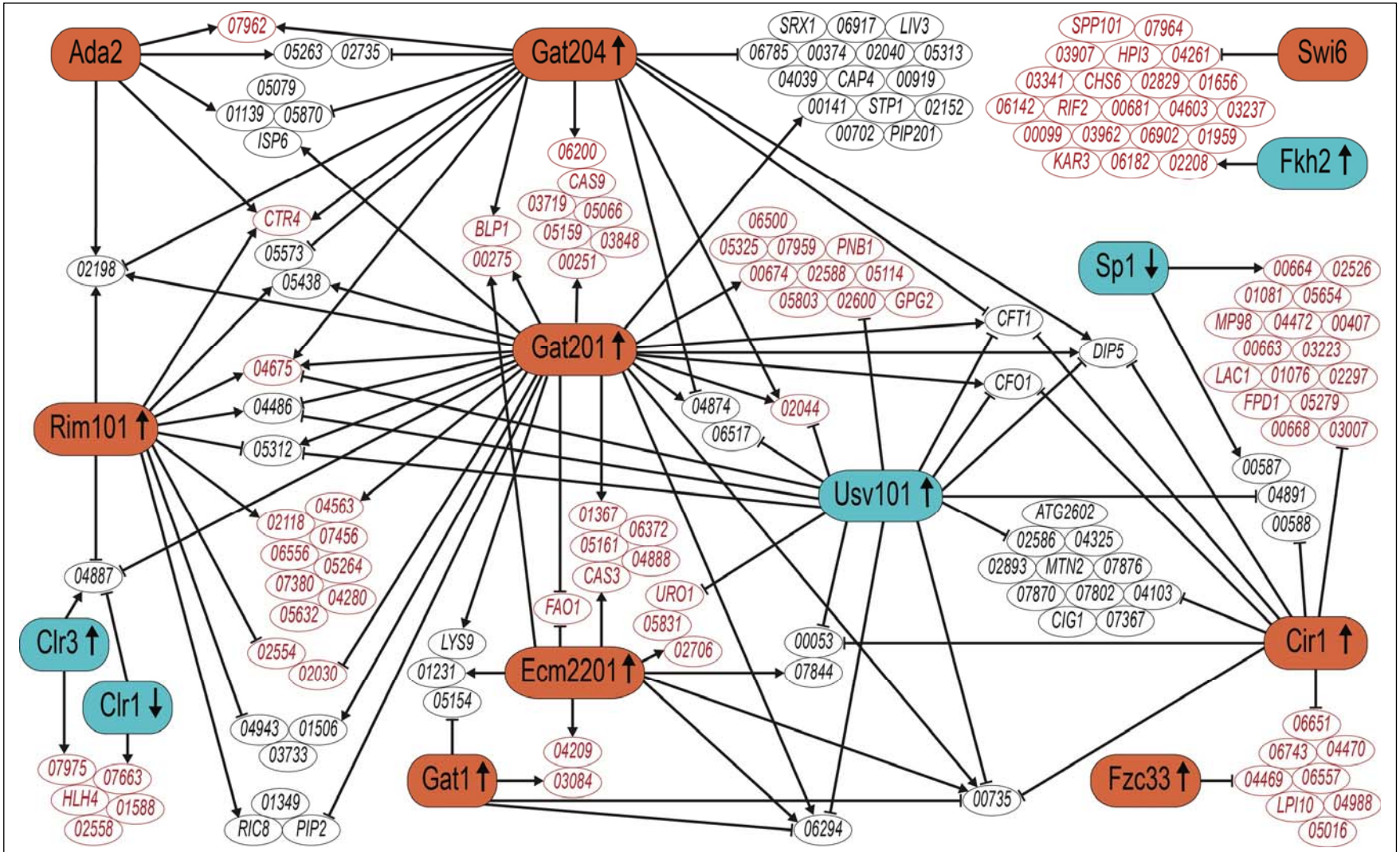

Fig 4. A TF network map of only shared targets of the 12 TF-TF pairs with the most significant hypergeometric p-values for their shared targets. These pairs are Gat201-Gat204, Usv101-Gat201, Swi61-Fkh2, Usv101-Cir1, Gat201-Rim101, Sp1-Cir1, Ecm2201-Gat201, Clr1-Clr3, Gat204-Ada2, Ecm2201-Gat1, Usv101-Ecm2201, Cir1-Fzc33.Orange and blue shading indicates that cells lacking the TF are hypocapsular and hypercapsular, respectively. TFs that are up-regulated and down-regulated in capsule inducing conditions are indicated by upward and downward arrows, respectively. Targets whose regulators display phenotype consistency (defined in main text) are colored red.

342

Comparison of the *Cryptococcus neoformans* and *Saccharomyces cerevisiae* regulatory networks

Much of what we know about the functions and interactions of proteins in *C. neoformans* comes from studies of their orthologs in the model yeast *S. cerevisiae*. However, reconstruction of phylogenetic gene trees by sequence similarity comes with substantial uncertainty, and therefore identification of truly orthologous genes between species does, too. This is especially problematic for genes belonging to families that have substantial sequence similarity within each species, as is the case for many TFs. Even when two TFs are true historical orthologs, their functions may have diverged during evolution, so non-orthologous TFs may have more functional similarity. We therefore combined sequence similarity with functional similarity, defined as regulating putatively orthologous target genes, to identify TF pairs with both types of similarity. We denote these pairs as (TF_{Cn}, TF_{Sc}).

Currently, the best available resource on homology between *S. cerevisiae* and *C. neoformans* proteins was compiled by Kelliher et al, who published a list of 4,572 homologous protein pairs based primarily on amino acid sequence similarity (Kelliher, et al., 2016). A few *C. neoformans* genes were mapped to more than one *S. cerevisiae* gene because *S. cerevisiae* genes include paralogs derived from a whole genome duplication event. We were curious as to whether the 63 pairs of TFs in this list, assigned by sequence homology, also showed functional similarity based on overlap of their target genes. For this purpose, we returned to the global NetProphet3 networks for *S. cerevisiae* ((Abid and Brent, 2023); Supplementary File S10) and *C. neoformans*, which relate all TFs to their targets.

First, we calculated a Jaccard similarity for each (TF_{Cn}, TF_{Sc}) pair based on their homologous target genes. As above, higher Jaccard similarity for a (TF_{Cn}, TF_{Sc}) pair means that a higher fraction of their targets encode homologous proteins. To determine whether the Jaccard similarity was significantly greater than what would be expected by chance, we calculated a permutation-based empirical P-value (see Methods). Surprisingly, only three (TF_{Cn}, TF_{Sc}) pairs from the Kelliher et al sequence-based list had significantly similar target gene sets: ($Mig1_{Cn}$, $Mig1_{Sc}$), ($Skn7_{Cn}$, $Skn7_{Sc}$), and ($Gat1_{Cn}$, $Gat1_{Sc}$). None of the other sequence-based (TF_{Cn}, TF_{Sc})

371 pairs on this list had significant target set overlap, despite solid sequence similarity between the
372 TFs (Supplementary File S11-A). Next, for the seven homologous (TF_{Cn}, TF_{Sc}) pairs from the
373 Kelliher list where both the *C. neoformans* and *S. cerevisiae* TF targets were enriched for a GO
374 biological process (Table 2), we looked for concordance between these process assignments.
375 Only one pair had overlap of their most significant GO terms, suggesting that the remaining
376 pairs regulate distinct biological processes.

377

TF_{Cn}	TF_{Sc}	Identity	GO_{Cn}	GO_{Sc}
Fkh2	Fkh2	47%	cell cycle and DNA replication	regulation of DNA replication
Fzc12	Mal33	40%	regulation of cell cycle and DNA replication	carbohydrate transport
Mln1	Rtg3	30%	carbohydrate transport	TCA cycle, amino acid and nitrogen metabolism
Rsc8	Swi3	36%	transcriptional regulation	transport of ubiquitinated proteins
Swi6	Swi6	25%	regulation of cell cycle	cell wall and meiosis
Yap1	Yap1	48%	transcriptional regulation	response oxidative stress, sterol biosynthesis, response to metal ion, sulfur metabolism
Zfc2	Usv1	46%	peptidyl-lysine modification to peptidyl-hypusine	metabolism of glycogen and other carbohydrates

Table 2. (TF_{cn}, TF_{sc}) pairs proposed by Kelliher et al where the targets of each TF are enriched for a GO biological process. Identity, amino acid identity from BLASTP; GO_{cn} and GO_{sc}, most significant GO terms for the target sets in *C. neoformans* and *S. cerevisiae* networks, respectively.

378

379 In the previous analyses, we began with TF sequence similarity and looked for evidence of
380 functional similarity. Next, we began with TF pairs that had similar target sets and asked
381 whether the TFs had similar protein sequences. For this we identified (TF_{Cn}, TF_{Sc}) pairs based
382 solely on overlap of their targets and selected those with the highest Jaccard similarity Index
383 (>0.1). Surprisingly, BLASTP did not identify sequence similarity between most of these (TF_{Cn},
384 TF_{Sc}) pairs, with only one ($Gat201_{Cn}$, $Gat4_{Sc}$) showing a significant alignment (Supplementary
385 File. S11-B). Thus, (TF_{Cn}, TF_{Sc}) pairs that shared homologous targets, and hence are more likely
386 have similar functions, did not have similar protein sequences.

387 When we investigated pairs with high scores for either sequence similarity or target
388 overlap, they rarely satisfied our criteria for the other characteristic. We next looked for (TF_{Cn},
389 TF_{Sc}) pairs that had a balance of both sequence and functional similarity, using less stringent
390 cutoffs. First, we ran BLASTP to align all possible (TF_{Cn}, TF_{Sc}) pairs. 161 Cryptococcus TFs were in
391 at least one pair with amino acid identity greater than 30%, 162 were in at least one pair with e-
392 value less or equal than 1, and 24 were in a pair that had a significant Jaccard similarity index
393 (corrected empirical $P < 0.05$; see Methods). Seventeen Cryptococcus TFs were in a pair that
394 met all these criteria. For each of these, we calculated a score that combined sequence and
395 functional similarity (SF score, see Methods). For each TF_{Cn}, we selected the (TF_{Cn}, TF_{Sc}) pair with
396 the highest SF score. If more than one TF_{Cn} mapped to the same TF_{Sc}, the TF_{Cn} with the higher SF
397 score was mapped to that TF_{Sc}, while the TF_{Cn} with the lower SF score was mapped to its TF_{Sc}
398 with the second-highest score.

399 The 17 TF pairs that we identified across these widely diverged species are derived from the
400 same common ancestor (based on sequence similarity) and retain shared functions (based on
401 target sets). Interestingly, nine of these (Table 3) do not appear in the sequence-based list of
402 Kelliher et al. The remaining eight of them were included in that list, but our analysis suggests
403 mapping to a different partner TF with high sequence and functional similarity (Table 4). For
404 only two TFs, Gat1 and Mig1, do both approaches agree (see Discussion).

TF_{Cn}	TF_{Sc}	Identity (%)	e-value	Jaccard similarity	Empirical P-value
Gat201	Gat4	41	2.16E-05	0.11	0
01069	Ppr1	50	6.36E-05	0.10	0.01
Fzc45	Pip2	34	0.007	0.09	0.01
Stb4	War1	32	0.026	0.09	0
Cir1	Gzf3	59	2.09E-18	0.08	0
Fzc19	Tbs1	30	0.83	0.08	0.03
Sip401	Dal81	32	0.76	0.08	0
Fzc44	Mal33	52	0.17	0.07	0.04
06097	Pdr3	48	7.40E-02	0.07	0.02

Table 3. Nine (TF_{Cn}, TF_{Sc}) pairs whose *C. neoformans* TFs_{Cn} were mapped to their potential *S. cerevisiae* homologous TFs_{Sc} using our method based on sequence and functional similarity, but were not mapped in the Kelliher et al list, which is based on sequence similarity only. Identity and e-value from BLASTP; Jaccard similarity and empirical P-values calculated as described in the Methods using our NetProphet3 maps for *C. neoformans* and *S. cerevisiae* and pairs are listed by Jaccard similarity.

405

TF_{Cn}	TF_{Sc}	Identity	e-value	Jaccard similarity	Empirical P-value	Our method	Kelliher et al.
Ert1	OAF1	41%	1.86E-06	0.07	0	□	
Ert1	Ert1	49%	2.82E-09	0.01	1		□
Fzc33	Rgt1	33%	0.046	0.07	0.01	□	
Fzc33	Sef1	43%	7.69E-07	0	NA		□
Gat1	Gat1	70%	3.12E-22	0.09	0	□	□
Mig1	Mig1	68%	8.48E-27	0.09	0	□	□
Ppr1	Mal13	47%	0.061	0.07	0.05	□	
Ppr1	Ppr1	25%	1.28E-07	0.04	0.61		□
Rim101	ZAP1	31%	5.93E-12	0.08	0.01	□	
Rim101	Rim101	40%	7.69E-07	0.04	1		□
Skn7	Skn7	51%	5.18E-31	0.06	0.03	□	□
Skn7	Hms2	28%	5.48E-08	0.03	1		□
Usv101	Msn2	59%	1.41E-11	0.1	0	□	
Usv101	Usv1	62%	3.59E-20	0.07	0.09		□
Usv101	Rgm1	58%	2.15E-21	0.05	0.99		□

Table 4. Shown are (TF_{Cn}, TF_{Sc}) pairs whose *C. neoformans* TFs_{Cn} were mapped to potential *S. cerevisiae* homologous TFs_{sc} using both our method based on sequence and functional similarity and the Kelliher et al. method based only on sequence similarity. The last two columns indicate whether the pair was identified using our method, the Kelliher et al method, or both. Identity, e-value, Jacquard similarity, and empirical p-value derived as in Table 3.

406

Of the 165 Cryptococcus TFs, fewer than half (72) have been mapped to *S. cerevisiae* TFs based on either our method or sequence similarity alone. The remaining 93 have apparently diverged and may serve functions that do not exist in *S. cerevisiae*. 29 of these 93 TFs have target sets enriched for genes that do not have homologs in *S. cerevisiae* (FDR = 0.05; Supplementary File S12).

Methods

Cell growth

To maximize reproducibility of RNA-seq, cell recovery from frozen stocks, initial culture in YPD, inoculation into preconditioned media, and growth followed a strictly controlled protocol. These methods are detailed in (Kang, et al., 2025).

RNA-Seq

RNA-Seq methods are as described in(Kang, et al., 2025).

NetProphet3

We processed counts of measured gene expression levels into $\log_2$ fold changes (FC) using DESeq2 (Love, et al., 2014) (Supplementary File. S4 and S5). Next, we ran NetProphet3. See (Abid and Brent, 2023) for details of how NetProphet3 works. Our dataset included perturbation profiles for 120 TFs. For those TFs, the differential expression (DE) feature for each gene was the $\log_2$ fold change of the gene's expression in the TF deletion mutant relative to its expression in the wild-type strain. For Gat204 and Liv3 we used TF perturbation data from Homer et al (Homer, et al., 2016). For the remaining TFs, we did not have differential expression data, so we assigned all genes a DE feature of 0. As described in the text, we trained NetProphet3 differently depending on whether TF binding data was available (Fig. 2A). For TFs that did not have binding data, we used models trained on *S. cerevisiae* data (NetProphet3 cross-trained mode; described in (Abid and Brent, 2023). Finally, we combined networks generated from the NetProphet3 cross-trained mode and the NetProphet3 integration mode into one final network (Figure 2A).

Evaluation metrics

<u>Binding metric:</u>

We used available binding data, which was for only 10 TFs: Cir1 (Do, et al., 2020), Cuf1 (Garcia-Santamarina, et al., 2018), Gat201 (Homer, et al., 2016), Gat204 (Homer, et al., 2016), Hap-X (Do, et al., 2020), Nrg1 (Gish, et al., 2016; Maier, et al., 2015), Liv3 (Homer, et al., 2016), Pdr802(Reuwsaat, et al., 2021), Sre1 (PRJNA557210), and Usv101 (Gish, et al., 2016; Maier, et

al., 2015). We selected TF-target bound edges based on their P-values, which yielded a total of 3,308 TF-target edges. Since these datasets and the information provided with them were so heterogeneous, we used different P-values for thresholds for each dataset (Supplementary Table S2).

GO metric:

We used the GO-Term-Finder package (Boyle, et al., 2004) to calculate whether the targets of each TF were enriched for any GO biological process terms (Abid and Brent, 2023). GO annotations were downloaded from the UniProt website (Supplementary File S7) and transferred from *S. cerevisiae* annotations based on sequence homology. We then calculated a Hypergeometric P-value for enrichment of each TF's targets with each GO biological process term. For TFs with at least one significant GO term, we calculated the median, across all significant GO terms, of their minus log P-value.

Empirical P-value for Jaccard similarity between Cryptococcus and Cerevisiae TFs

We constructed 1,000 randomized Cryptococcus networks by shuffling confidence scores of TF-target edges. For each randomized network, we calculated the Jaccard similarity index for each TF-TF pair. For each TF-TF pair in the original network, the fraction of TF-TF pairs in randomized networks with a greater or equal Jaccard similarity is the empirical P-value.

SF score calculation

For all cases where a *C. neoformans* TF aligned to an *S. cerevisiae* TF with at least 30% identity and an E-value of 1.0 or less, we compared their target sets and calculated an empirical P-value for the null hypothesis that the overlap is no larger than expected by chance (see previous paragraph). We then combined the alignment E-value with the empirical P-value for target overlap to obtain a sequence-functional similarity (SF) score via the following equation:

$$SF = \frac{-log_{10}(evalue)}{(empirical\ Pvalue * 100 + 1)}$$

The SF is large when the alignment E-value is small (hence its negative logarithm is large) and the empirical P-value is small. The +1 in the denominator is ensures that the fraction is still defined when the empirical P-value is zero.

Discussion

A TF network map represents regulatory relationships by linking each TF to its direct, functional targets – the genes it regulates by binding to their regulatory DNA. This mechanistic

471 approach stands in contrast to co-expression networks, which link pairs of genes that have
472 correlated expression patterns. Co-expression networks have been constructed for
473 Cryptococcus and used to identify new genes with a particular deletion phenotype, such as
474 altered capsule thickness (O'Meara, et al., 2024). TF network maps have also been used to
475 identify novel genes with a particular phenotype, by using analytic methods such as
476 PhenoProphet (Maier, et al., 2015). In addition, TF network maps can be used for inferring
477 changes in TF activity (Ma and Brent, 2021) and for determining which TFs regulate many of the
478 same target genes, and hence may form a physical complex (Abid and Brent, 2023).

479 For *S. cerevisiae*, a wealth of comprehensive genomic resources has contributed to our
480 understanding of gene regulation. Relevant datasets in this system include multiple binding
481 location datasets for most TFs, obtained by using ChIP-chip (Harbison, et al., 2004), ChIP-exo
482 (Rossi, et al., 2021), ChEC-seq (Lupo, et al., 2023; Mahendrawada, et al., 2023) and Transposon
483 Calling Cards (Kang, et al., 2020). These are complemented by multiple comprehensive TF
484 perturbation datasets, in which gene expression was profiled before and after each TF was
485 perturbed by gene deletion (Hu, et al., 2007; Kemmeren, et al., 2014), gene over-expression
486 (Hackett, et al., 2020), or TF degradation (Mahendrawada, et al., 2025). We used some of these
487 data sets and our mapping algorithm NetProphet3 to build a comprehensive TF network map
488 for *S. cerevisiae(Abid and Brent, 2023)*. We wished to do the same for *C. neoformans*, but no
489 comprehensive binding or TF perturbation datasets existed. We therefore opted to create a
490 comprehensive TF perturbation response dataset by carrying out RNA-Seq on strains in which
491 the gene encoding a single TF had been deleted, a resource that we expect will be useful to the
492 Cryptococcus research community.

493 With this dataset in hand, we used NetProphet3 to build a TF network map for
494 Cryptococcus. To determine its reliability, we used two primary metrics. The first tested
495 whether high-scoring edges (those NetProphet3 is most confident of) tend to be supported by
496 ChIP-seq data, for cases where such data are available (Fig. 2B). The second metric tested
497 whether TFs' target sets are enriched for genes with shared biological functions at a rate much
498 higher than would be expected by chance (Fig. 2C). Both assessments indicated that our
499 network was reliable, with metrics similar to those of our *S. cerevisiae* network.

500 The availability of a reliable network allowed us to examine biologically relevant behaviors
501 of cryptococcal TFs. For example, we found that when two TFs whose deletion alters capsule
502 thickness share a large fraction of their targets, they tend to regulate those targets in the same
503 direction (both activating or both repressing) when they have the same phenotype (both
504 hypocapsular or both hypercapsular); conversely, they tend to regulate their targets in opposite
505 directions when they have opposite phenotypes (Fig 3F). The network therefore makes
506 biological sense. We have also established a web interface where users can search for the
507 regulators of genes or targets of TFs and visualize the resulting subnetwork (under
508 construction). We anticipate this will be of use to researchers in the field as they develop and
509 test mechanistic hypotheses about cryptococcal TFs.

510 We were surprised by the prominent role in the network played by TFs whose deletion
511 alters capsule thickness. Ten capsule TFs are among the top 16 TFs in *C. neoformans* with the
512 greatest number of targets, highlighting their importance in regulation. We also designated TFs
513 as hubs if they were in the top 10% of TFs ranked both by the number of their targets and by
514 the number of regulators of the gene that encodes them. Eight of the 14 hub TFs in our
515 network had capsule phenotypes (Table 1). Both these fractions were significantly more than
516 would be expected by chance.

517 Notably, the TFs whose deletion alters capsule thickness tended to have some capsule-
518 implicated genes as targets, but they also had many other targets. No TF had capsule-
519 implicated genes as more than 36% of its targets; even for major capsule regulators such as
520 Gat201, Cir1, and Nrg1 capsule-implicated genes made up only 4-5% of their targets. This
521 suggests the unexpected idea that there may not be any capsule-specific TFs. Instead, capsule
522 genes appear to be regulated by TFs that also regulate many other processes. Large
523 polysaccharide capsules are phylogenetically restricted among fungi, presumably evolved to
524 manage species-specific challenges. Thus, one might expect that they could be deployed
525 specifically in response to such stresses, independent of phylogenetically ubiquitous stress
526 responses. However, our network suggests that this is not the case. Instead, it may be better to
527 think of capsule as one of a large set of tools that are deployed together in response to certain
528 stressful conditions.

529 Although there were no TFs that selectively regulated capsule-involved genes, there were
530 several pairs of capsule-involved TFs that regulated many genes in common. We were
531 interested to see which genes were coordinately regulated by multiple such TFs (Fig. 4), as they
532 may be involved in either capsule development or other stress-response functions. Most of
533 these genes are uncharacterized. Those that are regulated by three or even four capsule-
534 involved genes may be of particular interest for follow-up studies.

535 Cryptococcus genes are often named after their putative orthologs in *S. cerevisiae* and
536 implicitly assumed to have similar function. Putative orthologs are identified by protein
537 sequence similarity (Kelliher, et al., 2016). However, TFs typically consist of a small DNA binding
538 domain (DBD) and large intrinsically disordered regions, which contain no identifiable
539 functional domains. These disordered regions tend to have little sequence conservation across
540 long evolutionary distances. Therefore, TF alignments between distantly related species are
541 often possible only within the DBD, which comprises a small fraction of TF protein sequence.
542 Furthermore, it is increasingly understood that, while DBDs may define a core DNA binding
543 motif, binding locations are significantly impacted by the disordered domains (Brodsky, et al.,
544 2020; Mindel, et al., 2024). These observations suggest that inferring functional homology
545 between TFs from sequence alone may be difficult and error prone. To complement sequence-
546 based approaches, we attempted to determine functional homology by comparing *C.*
547 *neoformans* and *S. cerevisiae* TFs according to how many orthologous target genes they shared.
548 By combining sequence homology and functional homology, we identified 17 pairs of a
549 Cryptococcus TF and a corresponding *S. cerevisiae* TF that we consider true functional
550 orthologs. Cryptococcus TFs that are not paired to *S. cerevisiae* orthologs may have evolved
551 species-specific functions and therefore regulate many target genes that do not have *S.*
552 *cerevisiae* orthologs. Alternatively, they may have retained their function even as their target
553 genes diverged from *S. cerevisiae* to the point where their orthologs cannot be identified by
554 sequence comparison. Interestingly, sequence homology alone had suggested no orthologs for
555 nine of the cryptococcal TFs in our 17 pairs in a previous study (Table 3), while for the other
556 eight, our method suggested different orthologs from those suggested by sequence similarity
557 alone. Overall, the sequence-based method tended to agree with our method only when

protein identity was very high (~70%). This suggests that using TF network maps may, in general, be a useful way to identify functional orthologs between highly diverged species where they may otherwise remain obscure.

We constructed a high-quality TF network map using a large new dataset we generated of gene expression profiles from cells lacking a TF (perturbation response data). However, more data resources would make the network map even better. Although our genome sequence analysis identified 165 putative transcription factors in the genome of *Cryptococcus neoformans*, we were able to obtain or construct deletion mutants for only 120 of them. The remaining 43 may be essential, although only 13 of them are predicted to be essential in YPD (Billmyre, et al., 2024). For those that are essential, newly implemented methods in the field, such as inducible protein degradation, could be employed (Huang, et al., 2025). Another, complementary resource that is needed to fully illuminate the regulatory network is a comprehensive TF binding location dataset, potentially generated using ChIP-exo, ChEC-seq, or Transposon Calling Cards. Future integration of such data into our TF network using NetProphet3 will yield an enhanced TF network map that will further advance research into this fascinating and important pathogen.

Acknowledgments

We are grateful to Cynthia Ma and Yu Sung Kang for their willingness to answer questions about their previous analyses of Cryptococcus RNA-Seq data.

Funding

This work was supported by NIH awards AI087794 (Doering) and GM141012 (Brent).

References

Abid, D. and Brent, M.R. NetProphet 3: a machine learning framework for transcription factor network mapping and multi-omics integration. *Bioinformatics* 2023;39(2).

Agustinho, D.P., *et al.* Peeling the onion: the outer layers of *Cryptococcus neoformans*. *Mem Inst Oswaldo Cruz* 2018;113(7):e180040.

585 Billmyre, R.B., *et al.* Saturation transposon mutagenesis enables genome-wide identification of
586 genes required for growth and fluconazole resistance in the human fungal pathogen
587 Cryptococcus neoformans. *bioRxiv* 2024.
588 Bonneau, R., *et al.* The Inferelator: an algorithm for learning parsimonious regulatory networks
589 from systems-biology data sets de novo. *Genome Biol.* 2006;7(5):R36.
590 Boodwa-Ko, D. and Doering, T.L. A Quick reCAP: Discovering *Cryptococcus neoformans* Capsule
591 Mutants. *J Fungi (Basel)* 2024;10(2).
592 Boyle, E.I., *et al.* GO::TermFinder--open source software for accessing Gene Ontology
593 information and finding significantly enriched Gene Ontology terms associated with a list of
594 genes. *Bioinformatics* 2004;20(18):3710-3715.
595 Brodsky, S., *et al.* Intrinsically Disordered Regions Direct Transcription Factor In Vivo Binding
596 Specificity. *Mol. Cell* 2020;79(3):459-471 e454.
597 Bulmer, G.S. and Sans, M.D. *Cryptococcus neoformans*. II. Phagocytosis by human leukocytes. *J*
598 *Bacteriol* 1967;94(5):1480-1483.
599 Chun, C.D., Brown, J.C.S. and Madhani, H.D. A major role for capsule-independent
600 phagocytosis-inhibitory mechanisms in mammalian infection by *Cryptococcus neoformans*. *Cell*
601 *Host Microbe* 2011;9(3):243-251.
602 Denning, D.W. Global incidence and mortality of severe fungal disease. *Lancet Infect Dis*
603 2024;24(7):e428-e438.
604 Do, E., *et al.* A Transcriptional Regulatory Map of Iron Homeostasis Reveals a New Control
605 Circuit for Capsule Formation in *Cryptococcus neoformans*. *Genetics* 2020;215(4):1171-1189.
606 Fromtling, R., Shadomy, H. and Jacobson, E. Decreased virulence in stable, acapsular mutants of
607 *Cryptococcus neoformans*. *Mycopathologia* 1982;79:23-29.
608 Garcia-Santamarina, S., *et al.* Genome-wide analysis of the regulation of Cu metabolism in
609 Cryptococcus neoformans. *Mol. Microbiol.* 2018;108(5):473-494.
610 Gish, S.R., *et al.* Computational Analysis Reveals a Key Regulator of Cryptococcal Virulence and
611 Determinant of Host Response. *MBio* 2016;7(2).
612 Gish, S.R., *et al.* Computational Analysis Reveals a Key Regulator of Cryptococcal Virulence and
613 Determinant of Host Response. *mBio* 2016;7(2):e00313-00316.
614 Granger, D.L., Perfect, J.R. and Durack, D.T. Virulence of *Cryptococcus neoformans*. Regulation
615 of capsule synthesis by carbon dioxide. *J Clin Invest* 1985;76(2):508-516.
616 Greenfield, A., *et al.* DREAM4: Combining genetic and dynamic information to identify biological
617 networks and dynamical models. *PLoS One* 2010;5(10):e13397.
618 Hackett, S.R., *et al.* Learning causal networks using inducible transcription factors and
619 transcriptome-wide time series. *Mol. Syst. Biol.* 2020;16(3):e9174.
620 Harbison, C.T., *et al.* Transcriptional regulatory code of a eukaryotic genome. *Nature*
621 2004;431(7004):99-104.
622 Haynes, B.C., *et al.* Toward an Integrated Model of Capsule Regulation in *Cryptococcus*
623 *neoformans*. *PLoS Pathog.* 2011;7(12):e1002411.
624 Homer, C.M., *et al.* Intracellular Action of a Secreted Peptide Required for Fungal Virulence. *Cell*
625 *Host Microbe* 2016;19(6):849-864.
626 Hu, Z., Killion, P.J. and Iyer, V.R. Genetic reconstruction of a functional transcriptional regulatory
627 network. *Nat. Genet.* 2007;39(5):683-687.

628 Huang, M.Y., *et al.* An auxin-inducible degron system for conditional mutation in the fungal
629 meningitis pathogen Cryptococcus neoformans. *G3 (Bethesda)* 2025.
630 Huynh-Thu, V.A. and Geurts, P. dynGENIE3: dynamical GENIE3 for the inference of gene
631 networks from time series expression data. *Sci Rep* 2018;8(1):3384.
632 Jackson, C.A., *et al.* Gene regulatory network reconstruction using single-cell RNA sequencing of
633 barcoded genotypes in diverse environments. *Elife* 2020;9.
634 Jung, K.W., *et al.* Systematic functional profiling of transcription factor networks in
635 *Cryptococcus neoformans*. *Nat Commun* 2015;6:6757.
636 Kang, Y., *et al.* Dual threshold optimization and network inference reveal convergent evidence
637 from TF binding locations and TF perturbation responses. *Genome Res.* 2020;30(3):459-471.
638 Kang, Y.S., *et al.* Leveraging a new data resource to define the response of *Cryptococcus*
639 *neoformans* to environmental signals. *Genetics* 2025;229(1):1-29.
640 Kelliher, C.M., *et al.* Investigating Conservation of the Cell-Cycle-Regulated Transcriptional
641 Program in the Fungal Pathogen, Cryptococcus neoformans. *PLoS Genet* 2016;12(12):e1006453.
642 Kemmeren, P., *et al.* Large-scale genetic perturbations reveal regulatory networks and an
643 abundance of gene-specific repressors. *Cell* 2014;157(3):740-752.
644 Kozel, T.R. Non-encapsulated variant of *Cryptococcus neoformans*. II. Surface receptors for
645 cryptococcal polysaccharide and their role in inhibition of phagocytosis by polysaccharide.
646 *Infect Immun* 1977;16(1):99-106.
647 Kumar, P., *et al.* Emerging themes in cryptococcal capsule synthesis. *Curr. Opin. Struct. Biol.*
648 2011;21(5):597-602.
649 Kwon-Chung, K.J. and Rhodes, J.C. Encapsulation and melanin formation as indicators of
650 virulence in *Cryptococcus neoformans*. *Infect Immun* 1986;51(1):218-223.
651 Love, M.I., Huber, W. and Anders, S. Moderated estimation of fold change and dispersion for
652 RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.
653 Lupo, O., *et al.* The architecture of binding cooperativity between densely bound transcription
654 factors. *Cell Syst* 2023;14(9):732-745 e735.
655 Ma, C.Z. and Brent, M.R. Inferring TF activities and activity regulators from gene expression
656 data with constraints from TF perturbation data. *Bioinformatics* 2021;37(9):1234-1245.
657 Madar, A., *et al.* DREAM3: network inference using dynamic context likelihood of relatedness
658 and the inferelator. *PLoS One* 2010;5(3):e9803.
659 Mahendrawada, L., *et al.* Surprising connections between DNA binding and function for the
660 near-complete set of yeast transcription factors. *bioRxiv* 2023.
661 Mahendrawada, L., *et al.* Low overlap of transcription factor DNA binding and regulatory
662 targets. *Nature* 2025.
663 Maier, E.J., *et al.* Model-driven mapping of transcriptional networks reveals the circuitry and
664 dynamics of virulence regulation. *Genome Res.* 2015;25(5):690-700.
665 Margolin, A.A., *et al.* ARACNE: an algorithm for the reconstruction of gene regulatory networks
666 in a mammalian cellular context. *BMC Bioinformatics* 2006;7 Suppl 1:S7.
667 Meya, D.B. and Williamson, P.R. Cryptococcal Disease in Diverse Hosts. *N Engl J Med*
668 2024;390(17):1597-1610.
669 Mindel, V., *et al.* Intrinsically disordered regions of the Msn2 transcription factor encode
670 multiple functions using interwoven sequence grammars. *Nucleic Acids Res.* 2024;52(5):2260-
671 2272.

672 O'Meara, M.J., *et al.* CryptoCEN: A Co-Expression Network for Cryptococcus neoformans reveals
673 novel proteins involved in DNA damage repair. *PLoS Genet* 2024;20(2):e1011158.
674 O'Meara, T.R., *et al.* The *Cryptococcus neoformans* Rim101 transcription factor directly
675 regulates genes required for adaptation to the host. *Mol Cell Biol* 2014;34(4):673-684.
676 Rajasingham, R., *et al.* Global burden of disease of HIV-associated cryptococcal meningitis: an
677 updated analysis. *Lancet Infect. Dis.* 2017;17(8):873-881.
678 Reuwsaat, J.C.V., *et al.* The Transcription Factor Pdr802 Regulates Titan Cell Formation and
679 Pathogenicity of Cryptococcus neoformans. *mBio* 2021;12(2).
680 Rossi, M.J., *et al.* A high-resolution protein architecture of the budding yeast genome. *Nature*
681 2021;592(7853):309-314.
682 Roy, S., *et al.* Integrated module and gene-specific regulatory inference implicates upstream
683 signaling networks. *PLoS Comput. Biol.* 2013;9(10):e1003252.
684 Siahpirani, A.F. and Roy, S. A prior-based integrative framework for functional transcriptional
685 regulatory network inference. *Nucleic Acids Res* 2017;45(4):e21.
686 Srikanta, D., *et al.* A sensitive high-throughput assay for evaluating host-pathogen interactions
687 in *Cryptococcus neoformans* infection. *PLoS One* 2011;6(7):e22773.
688 Taylor, J.W. and Berbee, M.L. Dating divergences in the Fungal Tree of Life: review and new
689 analyses. *Mycologia* 2006;98(6):838-849.
690 Tugume, L., *et al.* Cryptococcal meningitis. *Nat Rev Dis Primers* 2023;9(1):62.
691 Wang, Z.A., Li, L.X. and Doering, T.L. Unraveling synthesis of the cryptococcal cell wall and
692 capsule. *Glycobiology* 2018;28(10):719-730.
693