# 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 (TFs) 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.

Many gene regulatory networks are intended to represent functional regulation rather than molecular interactions. In other words, edges indicate relationships between genes (or TFs and genes) that could be either direct or indirect. Here, we aim to generate a map that represents relationships in C*. neoformans* that are both direct and functional. In other words, the TF must physically bind to the

84 regulatory DNA of the target gene (so it is direct), and as a result, it must modulate the transcription
85 rate of that target (so it is functional). Importantly, this is not always the case: genes whose promoters
86 are bound by a TF in ChIP-Seq or other binding location assays are frequently unaffected by deleting or
87 over-expressing the TF, which shows that direct interaction does not always indicate a functional
88 relationship (Kang, et al., 2022; Mahendrawada, et al., 2025).

89 To generate our network, we used NetProphet3, a state-of-the-art method for mapping TF
90 networks that we recently developed (Abid and Brent, 2023). NetProphet3 uses as input gene
91 expression profiles of cells with or without perturbation of a TF by, for example, deletion or
92 overexpression. Because they are based on expression data, NetProphet3 maps are functional, in the
93 sense that TFs are likely to affect the expression of their predicted target genes. Here, we show that
94 the functional TF-target relationships predicted by NetProphet3 in *Cryptococcus* are highly enriched for
95 direct interactions, consistent with our findings in *Saccharomyces cerevisiae* (Abid and Brent, 2023).
96 Indeed, we show that NetProphet3 predicts the probability that a TF will bind a target gene in a ChIP-
97 seq experiment with reasonable accuracy, using only gene expression data as input.

98 In this work, we generated a comprehensive dataset of gene expression changes in response to TF
99 perturbation in *C. neoformans*. Specifically, we carried out RNA-Seq on 120 TF deletion strains (selected
100 as described below) and used the resulting data as input for NetProphet3 to create a TF network map.
101 The map assigns scores to all possible TF-target interactions that represent our confidence in the
102 existence of a direct, functional relationship. After validating this map, we used it to identify functional
103 orthologs of Cryptococcus TFs in *S. cerevisiae*. We also searched the network for TFs that primarily
104 regulate genes involved in capsule formation, but found none, suggesting that capsule is not deployed
105 independently but rather as part of an arsenal of cellular defense and stress response systems. Our
106 new TF perturbation response dataset is available from GEO (GSE297962), the network map we
107 constructed from it may be found at https://doi.org/10.5281/zenodo.17193243, and a visualization is
108 available at https://cryptococcus.net.

# 109 Results

## 110 Overview of the transcriptome

111 We generated gene expression data for all of the single-TF deletion strains that we could either
112 generate (Maier, et al., 2015) or obtain from the Madhani *C. neoformans* deletion collection (Boucher,

et al., 2024). We grew these 120 strains, each lacking one TF (Supplementary File S1), 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 (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 , File S1). We also included WT control cultures along with TF deletion strains in each experiment, totaling 122 WT 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

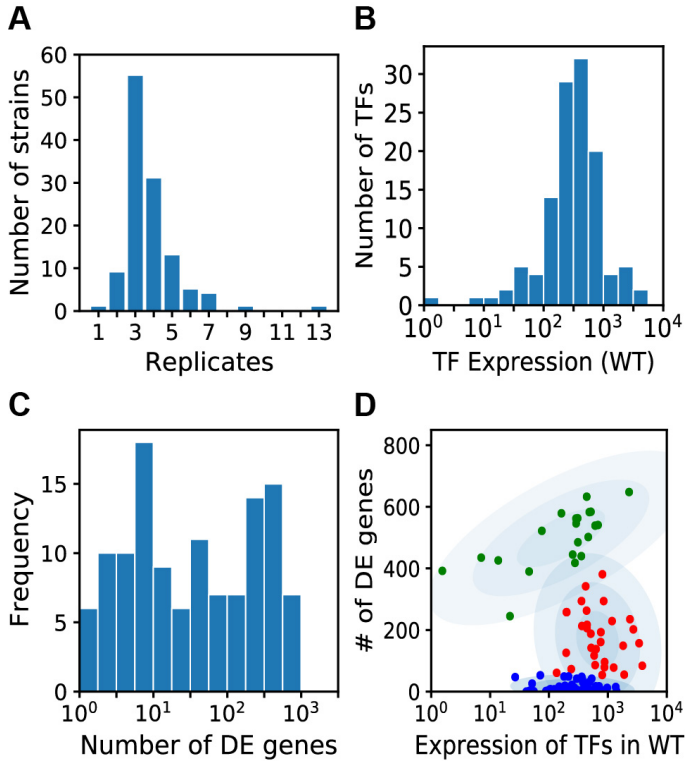


Fig 1. (A) Histogram of the number of RNA-seq biological replicates performed for TF deletion strains. (B) Histogram of the expression levels of TFs (normalized read counts) in WT cells. Only the 120 TFs that were deleted were included. (C) Histogram of the number of DE genes in TF deletion strains. (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.

126 Supplementary Figure S1A provides the values for individual TFs. The most highly expressed TF genes
127 were *RIM101* (O'Meara, et al., 2014) and *BZP1* (Jung, et al., 2015). Most TFs were measurably expressed
128 in capsule-inducing conditions (Figure 1B), so we expected that deleting the corresponding genes could
129 alter the expression of their targets.

130 Next, for each TF deletion strain we calculated the number of differentially expressed (DE) genes,
131 which may be functional targets of the TF (Fig. 1C and Supplementary Fig. S1B). A gene was defined as
132 DE if the adjusted P-value of expression in the deletion versus WT was below 0.01 and the absolute
133 $\log_2$ fold-change (FC) relative to WT was greater than 0.4 (https://doi.org/10.5281/zenodo.17193388
134 for non-TF-encoding genes and supplementary File S3 for TF-encoding genes). Half of the TF-deletion
135 strains had 1-20 DE genes, suggesting that they have relatively specific functions in our conditions,
136 while the other half had more than 20. Overall, the expression level of a TF in the WT strain was not
137 correlated with the number of genes that were DE in the TF deletion (Pearson correlation $\rho = 0$).
138 However, when we plotted the number of genes that responded to a TF deletion against that TF's WT
139 expression level, TFs clustered into three groups (Fig. 1D). TFs in the blue cluster regulated few to no
140 genes, despite the broad range of expression levels of those TFs in WT cells. In the green cluster, the
141 expression level of each TF correlated with the number of genes that were DE when that TF was deleted
142 (Pearson correlation $\rho = 0.57$, $P < 0.01$). TFs in the red cluster were expressed at a relatively high
143 level and had an intermediate number of DE genes (54-381). This group was enriched for TFs where
144 the corresponding deletion strain has an abnormal capsule phenotype (Hypergeometric p = 0.0007),
145 whereas the other clusters were not (hypergeometric p = 0.97 and p = 0.98; see Supplemental Methods
146 for details).

147

## 148 A TF network map for *Cryptococcus neoformans*

149 We used NetProphet3 (Abid and Brent, 2023) to construct a TF network map for *C. neoformans*.
150 The goal of NetProphet3 is to predict direct and functional TF-target relationships, in other words,
151 those in which a TF binds the promoter of a target gene and thereby influences its expression. It
152 achieves this by using a machine learning (ML) algorithm to learn patterns in gene expression data that
153 are indicative of direct regulation. ML algorithms are first trained to make predictions on cases where
154 the true answer is known. They are then applied to make predictions where the true answer may not

155 be known. During training, NetProphet3 uses available TF binding location data, such as from ChIP-Seq
156 experiments, to indicate whether binding of the TF to a gene promoter is known to occur; this is
157 assigned as a label, with a value of 1 (binding) or 0 (no binding) (Fig. 2A). In the prediction phase this
158 information is no longer available, so only the expression levels of a gene after a TF perturbation are
159 used to predict whether the TF binds its promoter region. In both phases, the gene expression data
160 used to make predictions is first processed in various ways, producing the actual inputs to NetProphet3
161 (called features). Specifically, the features include the fold-change in expression of the target when the
162 TF is perturbed (DE feature) and the relationships between the expression of TFs and the expression of
163 target genes across all gene expression samples (LASSO and BART features; see Methods).

164 When the trained NetProphet3 model is applied to gene expression data, it produces a binding
165 probability for each TF-gene pair (Fig. 2A, Predictions), which is interpreted as the probability that the
166 gene is a direct, functional target of that TF. The predicted probability is high if NetProphet3 identifies
167 patterns in the gene expression data that suggest the TF binds in the gene's promoter. Since very little
168 ChIP-seq binding location data is available for *Cryptococcus*, we trained NetProphet3 on *S. cerevisiae*
169 binding and expression data (see Methods for details). We refer to this as NetProphet3's *cross trained*
170 mode. We then used this trained model on our gene expression data from *C. neoformans* TF deletion
171 strains (without binding location data). The model is trained to combine evidence from features to
172 make reasonable probability estimates. Importantly, the features do not include the identity of specific
173 TFs or target genes. We therefore hypothesized that it would perform as well on *C. neoformans* data
174 as it did on *S. cerevisiae* data.

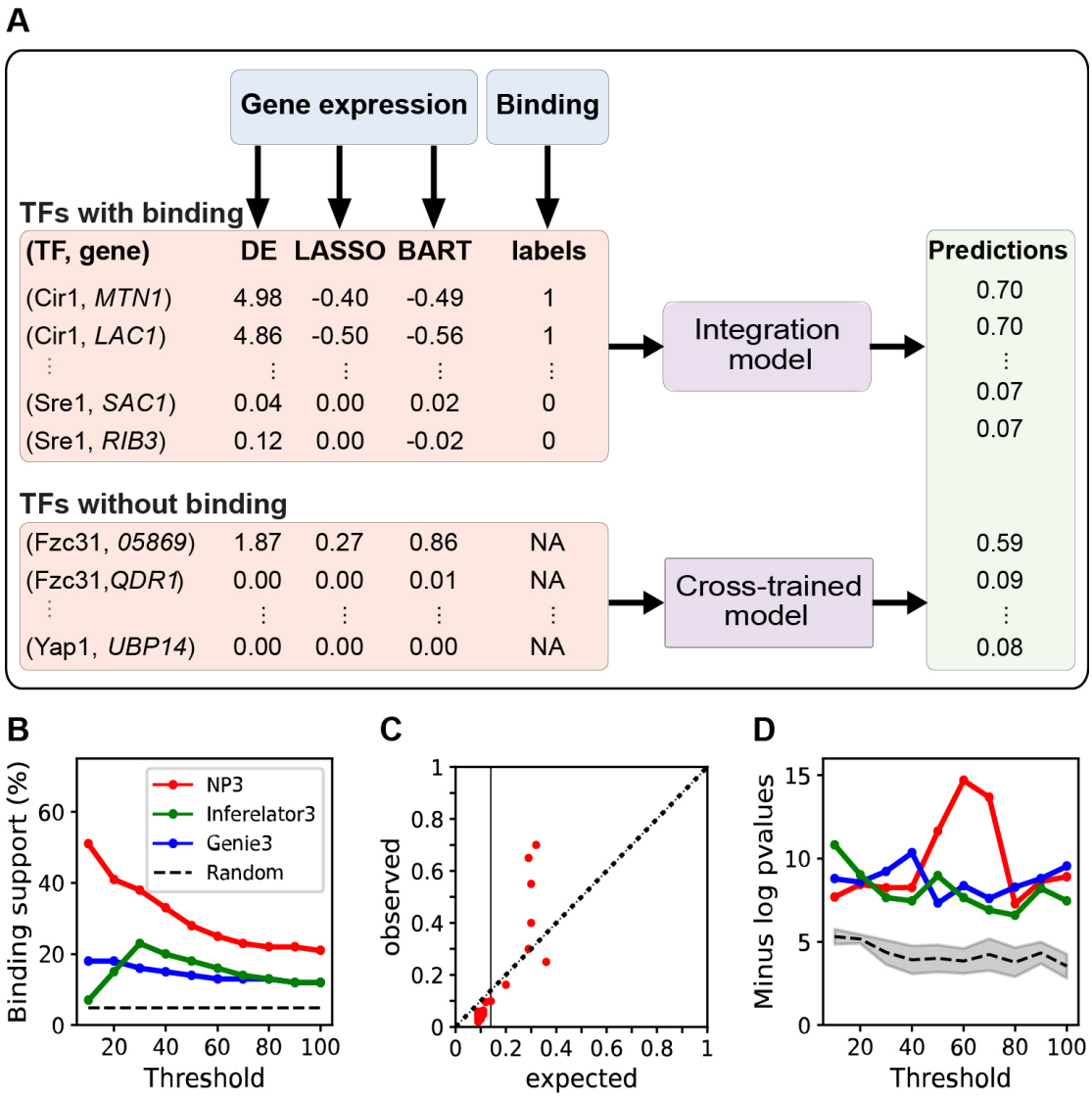


Fig 2. (A) Diagram of the NetProphet3 data flow, as described in the text. (B) Evaluation of the cross-trained mode, where Cryptococcus binding location data is not used in training, for agreement with binding data. Only TF-target edges where binding data is available for the TF were evaluated. Shown is the performance of NetProphet3 (red) versus random expectation (black dashed line) and two other established methods for inferring TF networks from gene expression data. Threshold, the number of highest-scoring edges included in the network, divided by the number of TFs. (C) ChIP-Seq validation of NetProphet predictions. Expected, NetProphet score (predicted probability); observed, actual validation of predictions by ChIP-seq data. (D) Evaluation of the NetProphet3 network after integrating binding data, using the GO metric; y-axis values plotted are the median across all TFs of the -log p value of the most significant GO term for that TF. The other algorithms indicated in panel (B) are shown for comparison. Shaded area, 95% confidence interval.

## The *C. neoformans* network map consists of direct interactions

We next used two evaluation metrics to assess whether high scoring TF-target edges in our map are direct and functional. One of these is based on binding data and one is based on gene function annotation. For the first one, we defined a binding evaluation metric as the fraction of predicted TF-

179 target edges that are supported by physical binding data (Abid and Brent, 2023). We used this to
180 evaluate the network, which had been trained using only *S. cerevisiae* data and then applied to our
181 *Cryptococcus* gene expression data. We performed this evaluation for 10 *C. neoformans* TFs with
182 associated physical binding data: Cir1 (Do, et al., 2020), Cuf1 (Garcia-Santamarina, et al., 2018), Gat201
183 (Homer, et al., 2016), Gat204 (Homer, et al., 2016), Hap-X (Do, et al., 2020), Nrg1 (Gish, et al., 2016;
184 Maier, et al., 2015), Liv3 (Homer, et al., 2016), Pdr802 (Reuwsaat, et al., 2021), Sre1 (PRJNA557210),
185 and Usv101 (Gish, et al., 2016; Maier, et al., 2015).

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

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

199 As expected, increasing the network size led to a reduction in the fraction of edges supported by
200 binding data (Fig. 2B). This shows that higher scoring edges are more likely to be direct. Furthermore,
201 the performance of the *C. neoformans* network with the binding metric is comparable to that of the *S.*
202 *cerevisiae* network produced by NetProphet3 (Abid and Brent, 2023), indicating that the NetProphet3
203 method developed using *S. cerevisiae* is equally applicable to *C. neoformans*. Since no *Cryptococcus*
204 binding data were used at any point in the construction of the network, these are effectively
205 prospective experiments, providing a good estimate of accuracy.

206 For comparison, we also ran two popular network inference algorithms from other labs, Genie3
207 (Huynh-Thu and Geurts, 2018) and Inferelator3 (Skok Gibbs, et al.) on our data (Fig. 2B). The results
208 showed that NP3 substantially outperformed both of these in terms of ChIP-Seq support.

209 The TF-target edge scores output by NP3 are intended to be estimates of the probability that the
210 target's promoter would be bound by the TF in a ChIP-seq experiment. To test whether these scores
211 were reasonable estimates, we selected sets of edges within various probability score ranges and
212 calculated both the expected validation rate for each set (the average of its edge scores) and its actual
213 validation rate, based on ChIP-Seq data (Fig. 2C). The vertical line shows the minimum score for
214 inclusion in the network of 100 targets per TF. Above this score threshold, the observed rate was either
215 very close to the expected rate or above it, indicating that the scores are reasonable estimates of the
216 validation rate and in some cases are even too pessimistic.

## 217 Incorporating Cryptococcus binding data into the network

218 Once evaluation with Cryptococcus binding data was complete, we produced a second network
219 that incorporated the binding data into the network by using NP3's integration mode (Abid and Brent,
220 2023) (see Supplemental Methods). The two networks differ only in the edges emanating from the 10
221 TFs for which we had ChIP-Seq data. The second network is used for the remainder of this paper and is
222 also available at https://doi.org/10.5281/zenodo.17193243. This network map can be searched and
223 visualized at http://cryptococcus.net.

## 224 The *C. neoformans* network map is functionally coherent

225 The second evaluation metric we applied was based on gene ontology (GO) biological process
226 annotations. This GO metric reflects the degree to which each TF's targets are enriched for genes that
227 share a common functional annotation and are therefore biologically coherent. For this evaluation, we
228 downloaded *C. neoformans* annotations from the UniProt website, most of which were transferred
229 from *S. cerevisiae* based on sequence similarity (see Methods for details). Overall, 799 GO biological
230 processes were transferred to *C. neoformans*, annotating over 3,000 genes, and the target sets of 30
231 TFs in our network were significantly enriched for at least one GO biological process term
232 (Supplementary Table S1). For each TF and each biological process annotation, we then calculated a
233 hypergeometric p-value (see Methods for details). A significant p-value indicates that more of the TF's
234 target genes are annotated with the function than would be expected by chance, and hence the TF's

235 targets are functionally coherent. To evaluate statistical significance, we created 30 random networks
236 by permuting TFs and targets and evaluated them using the same metric. Although there is some
237 fluctuation in absolute level, the performance of our network significantly exceeded those of the
238 random networks at every threshold, indicating that each TF's targets in our network were much more
239 functionally coherent than would be expected by chance (Fig. 2D). This further supports the validity of
240 the network. The Genie3 and Inferelator3 networks had similar functional coherence, although
241 NetProphet3 outperformed them on networks of intermediate size.

## 242 The most influential TFs are enriched for those that are known to regulate 243 capsule

244 For this study, we generated or obtained deletion strains lacking 120 of the 165 TFs in the
245 Cryptococcus genome. We were unable to delete the other 45. However, NetProphet3 still predicts
246 targets for these TFs. It does this by using non-DE features obtained by analyzing all gene expression
247 profiles together (Fig. 2A, LASSO and BART; see Methods). We therefore had predictions for all 163 TFs
248 that were expressed in our dataset, so our network included the 16,300 highest-scoring TF-target
249 edges. 161 of the TFs had at least one target gene under the conditions assayed. The total number of
250 regulated genes (4,103) corresponds to 60% of those that were measurably expressed in our studies
251 and 49% of all *C. neoformans* genes. This network is the basis for our analyses below.

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

258 The more targets a TF has, the more influence it can potentially exert on the transcriptional state
259 of the cell. In our network, the number of direct targets of each TF ranged from 7 to 884 (Fig. 3B). When
260 we ranked TFs by the number of targets, the top 16 TFs (10%; Supplementary File S5) covered 6,709
261 (41%) of the 16,300 TF-target edges in our network. Surprisingly, these 16 TFs regulate 81% of the 4,103
262 genes in the network, showing their disproportionate impact. We were interested to notice that 10 of

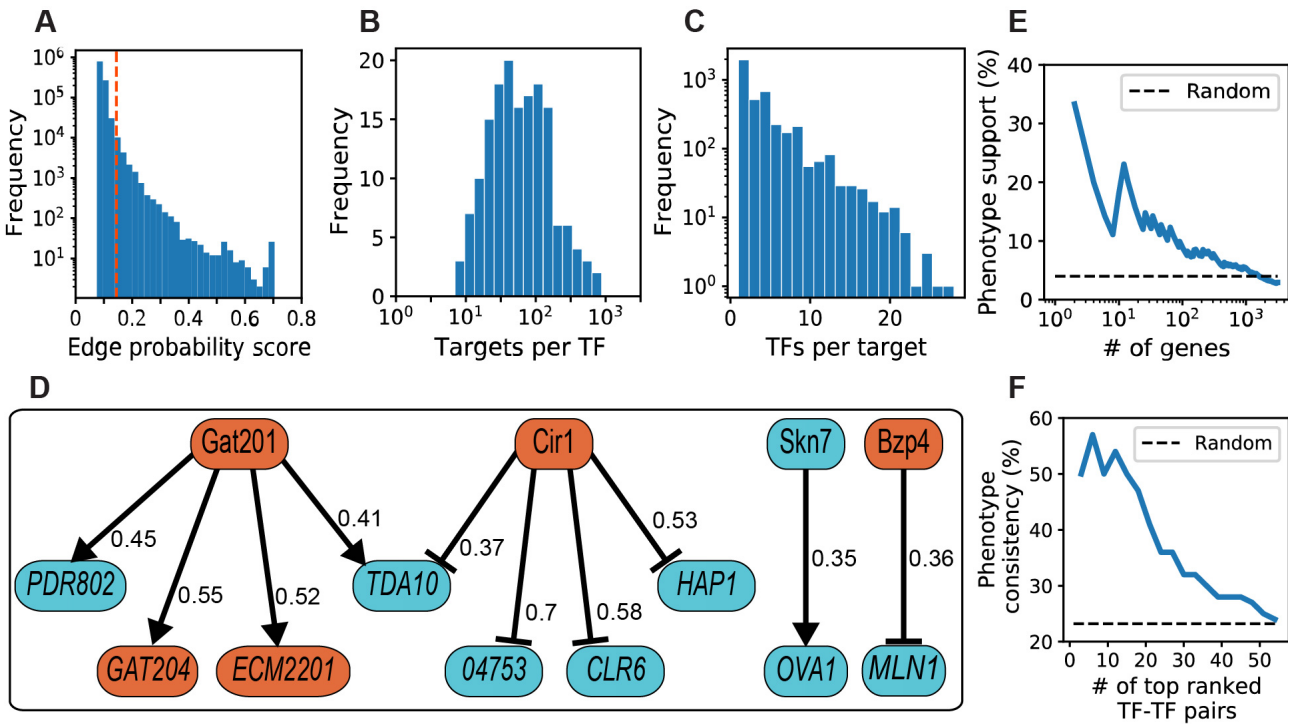


Fig 3. (A) NetProphet3 probability scores for all TF-gene edges. The vertical dashed orange line indicates the score threshold corresponding to 100 TF-target edges per TF on average. (B) The number of targets per TF. (C) The number of direct TF regulators per gene. (D) The top 10 high scoring TF-target edges in the capsule subnetwork. Colors indicate the phenotype of strains deleted for the corresponding genes: orange for hypocapsular and cyan for hypercapsular. 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.

these TFs are known to regulate the capsule, based on the phenotypes of corresponding deletion strains (File S2). This is significantly enriched over the fraction of all TFs that are known to regulate the capsule (63% versus 31%; hypergeometric p = 0.006; see Discussion).

In addition to targets per TF, we examined the distribution of TFs per target gene (Fig. 3C). The distribution peaked at one regulator per target gene and dropped off exponentially, with one outlier gene having 55 regulators (CNAG_04126, not plotted in Fig. 3C). When we ranked genes by this metric, we found that the top 10% of targets (410) participated in 33% (5,315) of the TF-target edges. On average these top 10% targets have 13 regulators (Supplementary File S5). Also, these 410 targets were generally enriched for capsule-implicated genes (those whose deletion perturbs capsule; hypergeometric $p < 3 \times 10^{-6}$; Supplementary File S5).

273 Many TFs are both powerful regulators and
274 highly regulated, thus occupying central positions in
275 transcriptional networks. We defined hub TFs as
276 those that are in the top 10% both by the number
277 of TFs that regulate them and by the number of
278 genes they regulate. Our network contained 14 hub
279 TFs (Table 1), of which eight are capsule implicated
280 (hypergeometric p = 0.007; see Discussion).

| TF | # of targets | # of regulators | Capsule-implicated |
|---|---|---|---|
| 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. We define hub TFs as those that are in the top 10% of TFs ranked by number of direct targets and are encoded by a gene in the top 10% of genes ranked by number of regulators. CNAG_05375 is an unnamed gene.

## 281 The *C. neoformans* network map is 282 biologically coherent

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

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

299 We speculated that target genes regulated by the highest number of capsule-implicated TFs were
300 most likely to have a capsule phenotype upon deletion. To test this, we ranked genes by the number
301 of capsule TFs that regulate them from highest to lowest and calculated, at different thresholds, the

302 percentage of genes that have a capsule phenotype (Fig. 3E). We found that genes regulated by the
303 highest number of capsule TFs were much more likely to be capsule genes than those that were
304 regulated by a lower number of capsule TFs: 9 out of the top 100 genes in this ranking were capsule-
305 implicated, while only 2% of all genes are capsule-implicated (hypergeometric $p = 0.02$).

## 306 TF-TF pairs that share many targets

307 We previously showed that TFs that share large fractions of their targets are more likely to work
308 together as a physical protein complex than those that do not (Abid and Brent, 2023). TF-TF pairs that
309 share many targets but do not form a physical complex may cooperate by other mechanisms, such as
310 being co-regulated by a third factor or binding the same DNA sequences. To gain insight into the TF-TF
311 relationships of capsule TFs, we calculated the Jaccard similarity between target sets of all capsule-
312 implicated TF pairs in the network (Supplementary File S6). The Jaccard similarity is the number of
313 targets in common divided by the total number of individual targets regulated by the two TFs. It is equal
314 to one if all targets are shared by both TFs, and zero if none are shared. For each TF-TF pair, we
315 determined whether this similarity was significantly higher than would be expected by chance
316 (hypergeometric adjusted $p < 0.05$), indicating that the TFs are likely to cooperate (as defined above).
317 For this analysis, we used only TF-target edges that had high NetProphet3 probability scores (greater
318 or equal to 0.2). These TF-TF interactions form a second subnetwork, the capsule TF subnetwork, that
319 includes 43 capsule TFs linked by 56 TF-TF edges.

320 We observed that these TF pairs tended to regulate their targets in the same direction if the two
321 TFs had the same capsule phenotype and in the opposite direction if the TFs had opposite capsule
322 phenotypes. We considered the TF-TF pair to display phenotype consistency if both TFs had the same
323 capsule phenotype and at least 80% of their shared targets were regulated in the same direction, or if
324 they had opposite phenotypes and at least 80% of their shared targets were regulated in the opposite
325 direction. We ranked these TF-TF pairs by the significance of their Jaccard similarity and calculated the
326 fraction of pairs above different rank thresholds that display phenotype consistency (Fig. 3F). This
327 fraction was much higher than would be expected at random and tended to decline as the significance
328 of their target-set overlap declined, consistent with the notion that pairs with the highest target overlap
329 are most likely to work in a coordinated fashion.

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., 2016;
332 Haynes, et al., 2011; Kumar, et al., 2011; Maier, et al., 2015). However, these have generally been
333 limited to relatively small groups of genes and focus on direct regulation of one TF by another. Here
334 we focus on the overlap of targets among all capsule TFs. Figure 4 shows all pairs of capsule TFs that
335 share many more targets in common than would be expected by chance ($p < 10^{-5}$), along with the
336 targets that they share. Notably, half of these TFs are hub TFs, highlighting their central role in
337 transcriptional regulation. While most of the regulatory relationships are activating, several TFs
338 (Usv101 and Cir1) are uniformly repressive in this context.

339

## 340 Comparison of the *Cryptococcus neoformans* and *Saccharomyces cerevisiae* 341 regulatory networks

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

352 Currently, the best available resource on homology between *S. cerevisiae* and *C. neoformans*
353 proteins was compiled by Kelliher et al, who published a list of 4,572 homologous protein pairs based
354 primarily on amino acid sequence similarity (Kelliher, et al., 2016). In this list, a few *C. neoformans*
355 genes were mapped to more than one *S. cerevisiae* gene because *S. cerevisiae* genes include paralogs
356 derived from a whole genome duplication event. We were curious as to whether the 63 pairs of TFs in
357 this list, assigned by sequence homology, also showed functional similarity based on overlap of their

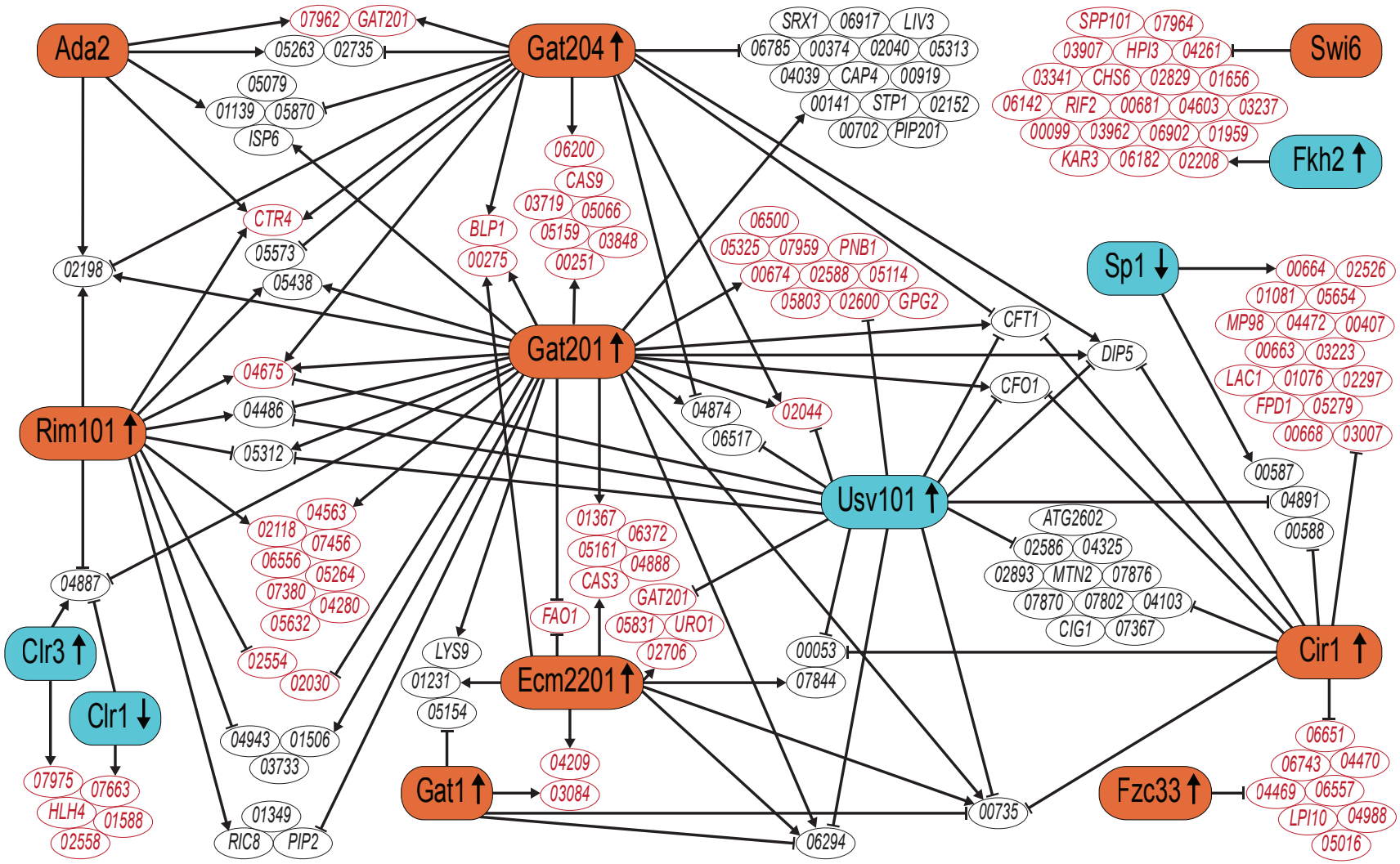


Fig 4. A TF network map of only shared targets of the 12 TF-TF pairs with the most significant hypergeometric p-values for their shared targets. These pairs are Gat201-Gat204, Usv101-Gat201, Swi6-Fkh2, Usv101-Cir1, Gat201-Rim101, Sp1-Cir1, Ecm2201-Gat201, Clr1-Clr3, Gat204-Ada2, Ecm2201-Gat1, Usv101-Ecm2201, Cir1-Fzc33. Orange and blue shading indicates that cells lacking the TF are hypocapsular or 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 labeled in red.

358

359 target genes. For this purpose, we returned to the full NetProphet3 network for *S. cerevisiae*
360 ((Abid and Brent, 2023); https://doi.org/10.5281/zenodo.17196637) and *C. neoformans*.

361 First, we calculated a Jaccard similarity for each ($TF_{Cn}$, $TF_{Sc}$) pair based on their homologous
362 target genes. As above, higher Jaccard similarity for a ($TF_{Cn}$, $TF_{Sc}$) pair means that a higher fraction
363 of the TFs' targets encode homologous proteins. To determine whether the Jaccard similarity was
364 significantly greater than what would be expected by chance, we calculated a permutation-based
365 empirical P-value (see Methods). Surprisingly, only three ($TF_{Cn}$, $TF_{Sc}$) pairs from the Kelliher et al
366 sequence-based list had significantly similar target gene sets: ($Mig1_{Cn}$, $Mig1_{Sc}$), ($Skn7_{Cn}$, $Skn7_{Sc}$),
367 and ($Gat1_{Cn}$, $Gat1_{Sc}$). None of the other pairs on this list had significant target-set overlap, despite
368 solid sequence similarity (Supplementary File S7-A). Next, we examined the seven homologous
369 pairs from the Kelliher list where both the *C. neoformans* and *S. cerevisiae* TF targets were
370 enriched for a GO biological process (Table 2). Only one pair, ($Fkh2_{Cn}$, $Fkh2_{Sc}$), had overlap of their
371 most significant GO terms, suggesting that the remaining pairs primarily regulate distinct
372 biological processes.

373

| $TF_{Cn}$ | $TF_{Sc}$ | Identity | *Cryptococcus* GO term | *Cerevisiae* GO term |
|---|---|---|---|---|
| Fkh2 | Fkh2 | 47% | cell cycle and DNA replication | regulation of DNA replication |
| Fzc1 2 | 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 to 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 term. 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.

374

375 In the previous analyses, we began with TF sequence similarity and looked for evidence of
376 functional similarity. Next, we began with TF pairs that had similar target sets and asked whether

377 the TFs had similar protein sequences. For this we identified ($TF_{Cn}$, $TF_{Sc}$) pairs based solely on
378 overlap of their targets and selected those with the highest Jaccard similarity Index (>0.1).
379 Surprisingly, BLASTP did not identify sequence similarity between most of these TFs, with only
380 one pair ($Gat201_{Cn}$, $Gat4_{Sc}$), showing a significant alignment (evalue<1.0; Supplementary File S7-
381 B). In general, pairs that shared homologous targets, and hence are more likely have similar
382 functions, did not have similar protein sequences.

383 We next looked for ($TF_{Cn}$, $TF_{Sc}$) pairs that had a balance of both sequence and functional
384 similarity, using less stringent cutoffs. First, we ran BLASTP to align all possible ($TF_{Cn}$, $TF_{Sc}$) pairs.
385 161 Cryptococcus TFs were in at least one pair with amino acid identity greater than 30%, 162
386 were in at least one pair with BLASTP e-value less or equal than 1, and 24 were in a pair that had
387 a significant Jaccard similarity index (corrected empirical $P < 0.05$; see Methods). Seventeen
388 Cryptococcus TFs were in a pair that met all these criteria. For each of these, we calculated a
389 score that combined their sequence and functional similarity (SF score, see Methods). For each
390 *Cryptococcus* TF, we selected the ($TF_{Cn}$, $TF_{Sc}$) pair with the highest SF score. If more than one $TF_{Cn}$
391 mapped to the same $TF_{Sc}$, the $TF_{Cn}$ with the higher SF score was paired with that $TF_{Sc}$, while the
392 $TF_{Cn}$ with the lower SF score was mapped to its $TF_{Sc}$ with the second-highest score.

393 The 17 TF pairs that we identified across these widely diverged species are derived from the
394 same common ancestor (based on sequence similarity) and retain shared functions (based on
395 target sets). Interestingly, the cryptococcal TFs in nine of these pairs (Table 3) do not appear in
396 the sequence-based list of Kelliher et al. The remaining eight were included in that list, but our
397 analysis suggests that they pair best to different partner TFs in *S. cerevisiae* with high sequence
398 and functional similarity (Table 4). For only two TFs, Gat1 and Mig1, do both approaches agree
399 (see Discussion).

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

404 (FDR = 0.05; Supplementary File S8), thus highlighting the unique biology of *C. neoformans*
405 compared to model yeast.

| $TF_{Cn}$ | $TF_{Sc}$ | 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 identified by our method based on sequence and functional similarity, but not by Kelliher et al., who used sequence similarity alone. Identity and e-value from BLASTP; Jaccard similarity and empirical P-values calculated as described in Methods.

406

407

| $TF_{Cn}$ | $TF_{Sc}$ | Identity | e-value | Jaccard similarity | Empirical P-value | Our method | Kelliher et al. |
|---|---|---|---|---|---|---|---|
| Ert1 | OAF1 | 41% | 1.86E-06 | 0.07 | 0.00 | ✓ | |
| Ert1 | Ert1 | 49% | 2.82E-09 | 0.01 | 1.00 | | ✓ |
| Fzc33 | Rgt1 | 33% | 0.046 | 0.07 | 0.01 | ✓ | |
| Fzc33 | Sef1 | 43% | 7.69E-07 | 0 | N/A | | ✓ |
| Gat1 | Gat1 | 70% | 3.12E-22 | 0.09 | 0.00 | ✓ | ✓ |
| Mig1 | Mig1 | 68% | 8.48E-27 | 0.09 | 0.00 | ✓ | ✓ |
| 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 | 1.00 | | ✓ |
| Skn7 | Skn7 | 51% | 5.18E-31 | 0.06 | 0.03 | ✓ | ✓ |
| Skn7 | Hms2 | 28% | 5.48E-08 | 0.03 | 1.00 | | ✓ |
| Usv101 | Msn2 | 59% | 1.41E-11 | 0.1 | 0.00 | ✓ | |
| Usv101 | Usv1 | 62% | 3.59E-20 | 0.07 | 0.09 | | ✓ |
| Usv101 | Rgm1 | 58% | 2.15E-21 | 0.05 | 0.99 | | ✓ |

Table 4. ($TF_{Cn}$, $TF_{Sc}$) pairs identified using both our method based on sequence and functional similarity and the Kelliher et al. method based only on sequence similarity alone. Identity, e-value, Jacquard similarity, and empirical p-value derived as in Table 3.

408

# Methods

## Cell growth

To maximize reproducibility of RNA-seq, we followed strictly controlled protocols for cell recovery from frozen stocks, initial culture in YPD, inoculation into preconditioned media, and growth (see Supplementary Methods) and (Kang, et al., 2025).

## RNA-Seq

RNA isolation, library preparation, and sequencing were by standard methods as detailed in the Supplementary Methods and (Kang, et al., 2025).

## NetProphet3

We processed counts of measured gene expression levels into normalized $\log_2$ fold changes using DESeq2 (Love, et al., 2014). Samples for each TF deletion strain were compared to all 122 WT samples taken together. We then 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, which did not have DE data, we assigned all potential target genes a value of 0 (https://doi.org/10.5281/zenodo.17193620).

LASSO (Efron, et al., 2004; Friedman, et al., 2008) and BART (Chipman, et al., 2010; He, et al., 2019) are regression algorithms which were used to derive additional features for each potential TF-target edge by training models to predict the expression of each gene from the expression levels of all TFs. The LASSO feature for a TF-gene edge is the linear regression coefficient learned for that TF in the model for that gene. The BART feature for a TF-gene edge is the predicted change in expression of the gene when the TF's expression changes from its highest to lowest value found in the data (see (Abid and Brent, 2023) for details).

434 As described in the text, we first trained NetProphet3 on binding and expression data from *S.*
435 *cerevisiae* (NetProphet3 cross-trained mode, described in (Abid and Brent, 2023)). This training
436 is not TF-specific; NetProphet3 simply learns patterns of gene expression features that are
437 characteristic edges supported by binding location data. For the ten TFs that have binding
438 location data in Cryptococcus, we trained a separate, TF-specific model that integrates binding
439 and expression data (NetProphet3 integration mode, described in (Abid and Brent, 2023)).Finally,
440 we combined networks produced by the cross-trained mode and the integration mode into one
441 final network (Figure 2A).

## 442 Evaluation metrics

443 Binding metric:

444 We used available binding data, which was for only 10 TFs (see Results and Table S2), which
445 contained 3,308 supported TF-target edges. Since these datasets and the information provided
446 with them were so heterogeneous, we used different P-values for thresholds for each dataset
447 (Supplementary Table S2).

448 GO metric:

449 We used the GO-Term-Finder package (Boyle, et al., 2004) to calculate whether the targets of
450 each TF were enriched for any GO biological process terms (Abid and Brent, 2023). GO
451 annotations were downloaded from the UniProt website (Supplementary File S4) and transferred
452 from *S. cerevisiae* annotations based on sequence homology. We then calculated a
453 Hypergeometric P-value for enrichment of each TF's targets with each GO biological process
454 term. For TFs with at least one significant GO term, we calculated the median, across all
455 significant GO terms, of their minus log P-value.

## Empirical P-value for Jaccard similarity between Cryptococcus and Saccharomyces 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.

## Sequence-functional (SF) similarity score calculation

For Cryptococcus TFs that 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 ensures that the fraction is still defined when the empirical P-value is zero.

472 # Discussion

473 A TF network map represents regulatory relationships by linking each TF to its direct,
474 functional targets – the genes it regulates by binding to their regulatory DNA. This mechanistic
475 approach stands in contrast to co-expression networks, which link pairs of genes that have
476 correlated expression patterns. Co-expression networks have been constructed for Cryptococcus
477 and used to identify new genes with a particular deletion phenotype, such as altered capsule
478 thickness (O'Meara, et al., 2024). TF network maps have also been used to identify novel genes
479 with a particular phenotype by using methods such as PhenoProphet (Maier, et al., 2015). In
480 addition, TF network maps can be used for inferring changes in TF activity (Ma and Brent, 2021)
481 and for determining which TFs regulate many of the same target genes, and hence may form a
482 physical complex (Abid and Brent, 2023).

483 For *S. cerevisiae*, a wealth of comprehensive genomic resources has contributed to our
484 understanding of gene regulation. There are multiple binding location datasets for most yeast
485 TFs, obtained by using ChIP-chip (Harbison, et al., 2004), ChIP-exo (Rossi, et al., 2021), ChEC-seq
486 (Lupo, et al., 2023; Mahendrawada, et al., 2023) and Transposon Calling Cards (Kang, et al., 2020).
487 These are complemented by multiple comprehensive datasets, in which gene expression was
488 profiled before and after each yeast TF was perturbed by gene deletion (Hu, et al., 2007;
489 Kemmeren, et al., 2014), gene over-expression (Hackett, et al., 2020), or rapid protein
490 degradation (Mahendrawada, et al., 2025). We previously used some of these data sets and our
491 mapping algorithm NetProphet3 to build a comprehensive TF network map for *S. cerevisiae* (Abid
492 and Brent, 2023). We wished to do the same for *C. neoformans*, but no comprehensive binding
493 or TF perturbation datasets existed. To fill this gap, we created a comprehensive TF perturbation
494 response dataset by carrying out RNA-Seq on strains in which the gene encoding a single TF had
495 been deleted, a resource that we expect will be useful to the Cryptococcus research community.
496 This dataset is completely independent of our large RNA-Seq dataset on responses to host-like
497 conditions (Kang, et al., 2024).

498 With this dataset in hand, we used NetProphet3 to build a TF network map for Cryptococcus.
499 To determine its reliability, we used two primary metrics. The first tested whether high-scoring
500 edges (those NetProphet3 is most confident of) tend to be supported by ChIP-seq data where

501 such data are available (Fig. 2B, 2C). By this metric, NP3 outperformed two other popular network
502 inference algorithms that we applied to our gene expression data. The second metric tested
503 whether TFs' target sets are enriched for genes with shared biological functions at a rate higher
504 than would be expected by chance (Fig. 2D). Both assessments indicated that our network is
505 reliable.

506 Our previous *Cryptococcus* TF network map (Maier, et al., 2015) was built using gene
507 expression data on only 41 TF-deletion strains. The current network uses data from 79 additional
508 TF deletion strains, so it has more high-scoring edges from the new TFs. By our measures, the
509 new network is significantly better (Fig. S2). We also built a website for the new network where
510 users can visualize subnetworks comprising TFs and genes of interest (https://cryptococcus.net).
511 We anticipate this will be of use to researchers in the field as they develop and test mechanistic
512 hypotheses about cryptococcal TFs.

513 Capsule-implicated TFs played a central role in our network, comprising 8 of the 14 hub
514 TFs (Table 1), significantly more than expected by chance. Surprisingly, no TF had capsule-
515 implicated genes as more than 36% of its targets. Even for major capsule regulators such as
516 Gat201, Cir1, and Nrg1, only 4-5% of their targets were capsule-implicated genes. This suggests
517 the unexpected idea that there may not be any capsule-specific TFs. Instead, capsule genes
518 appear to be regulated by TFs that also regulate many other processes. Large polysaccharide
519 capsules are phylogenetically restricted among fungi, presumably evolved to manage species-
520 specific challenges. Thus, one might expect that they could be deployed specifically in response
521 to such stresses, independent of phylogenetically ubiquitous stress responses. However, our
522 network suggests that this is not the case. Instead, it may be better to think of capsule as one of
523 a large set of tools that are deployed together in response to certain stressful conditions.
524 Nonetheless, genes that are coordinately regulated by multiple capsule-implicated TFs (Fig. 4)
525 may be involved in capsule development or other key stress-response functions and hence of
526 interest for follow-up studies.

527 Cryptococcus genes are often named after their putative orthologs in *S. cerevisiae* and
528 implicitly assumed to have similar function. Putative orthologs are identified by protein sequence
529 similarity (Kelliher, et al., 2016). However, TFs typically consist of a small DNA binding domain

530 (DBD) and large intrinsically disordered regions that contain no identifiable functional domains.
531 These disordered regions tend to have little sequence conservation across long evolutionary
532 distances. Therefore, TF alignments between distantly related species are often possible only
533 within the DBD, which comprises a small fraction of TF protein sequence. Furthermore, it is
534 increasingly understood that, while DBDs may define a core DNA binding motif, binding locations
535 are significantly impacted by the disordered domains (Brodsky, et al., 2020; Mindel, et al., 2024).
536 These observations suggest that inferring functional homology between TFs from sequence alone
537 may be difficult and error prone. To complement sequence-based approaches, we attempted to
538 determine functional homology by comparing *C. neoformans* and *S. cerevisiae* TFs according to
539 how many putatively orthologous target genes they shared. By combining sequence homology
540 and functional homology, we identified 17 pairs consisting of a Cryptococcus TF and an *S.*
541 *cerevisiae* TF that we consider true functional orthologs. Cryptococcus TFs that are not paired to
542 *S. cerevisiae* orthologs may have evolved species-specific functions and therefore regulate many
543 target genes that do not have *S. cerevisiae* orthologs. Alternatively, they may have retained their
544 function while their target genes diverged from *S. cerevisiae* to the point where their orthologs
545 cannot be identified by sequence comparison. Interestingly, sequence homology alone had
546 suggested no orthologs for nine of the cryptococcal TFs in our 17 pairs in a previous study (Table
547 3), while for the other eight, our method suggested different orthologs from those suggested by
548 sequence similarity alone. Overall, the sequence-based method tended to agree with our method
549 only when protein identity was very high (~70%). This suggests that TF network maps may, in
550 general, be important for identifying functional orthologs between highly diverged species.

551 We were able to construct a high-quality TF network map using a large new dataset we
552 generated of gene expression profiles from strains lacking TFs (perturbation response data).
553 However, more data resources would make the network map even better. Although our genome
554 sequence analysis identified 165 putative transcription factors in the genome of *Cryptococcus*
555 *neoformans*, we were able to obtain or construct deletion mutants for only 120 of them. Some
556 of the remaining 45 have been added to the Madhani deletion collection since this work was
557 carried out. The remainder may be essential (Billmyre, et al., 2024). For those that are essential,
558 newly implemented methods in the field, such as inducible protein degradation, (Huang, et al.,

2025), may be useful. 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 (Mayhew, 2016; Yen, 2023). 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).

# Data availability

All RNA-Seq data on Cryptococcus TF deletion strains used in this paper are available through the NCBI Gene Expression Omnibus under accession GSE297962.

576 # References

577

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

580 Billmyre, R.B.*, et al.* Saturation transposon mutagenesis enables genome-wide identification of
581 genes required for growth and fluconazole resistance in the human fungal pathogen
582 Cryptococcus neoformans. *bioRxiv* 2024.

583 Bonneau, R.*, et al.* The Inferelator: an algorithm for learning parsimonious regulatory networks
584 from systems-biology data sets de novo. *Genome Biol.* 2006;7(5):R36.

585 Boodwa-Ko, D. and Doering, T.L. A Quick reCAP: Discovering *Cryptococcus neoformans* Capsule
586 Mutants. *J Fungi (Basel)* 2024;10(2).

587 Boucher, M.J.*, et al.* Phenotypic landscape of a fungal meningitis pathogen reveals its unique
588 biology. *bioRxiv* 2024.

589 Boyle, E.I.*, et al.* GO::TermFinder--open source software for accessing Gene Ontology
590 information and finding significantly enriched Gene Ontology terms associated with a list of
591 genes. *Bioinformatics* 2004;20(18):3710-3715.

592 Brodsky, S.*, et al.* Intrinsically Disordered Regions Direct Transcription Factor In Vivo Binding
593 Specificity. *Mol. Cell* 2020;79(3):459-471 e454.

594 Bulmer, G.S. and Sans, M.D. *Cryptococcus neoformans*. II. Phagocytosis by human leukocytes. *J*
595 *Bacteriol* 1967;94(5):1480-1483.

596 Chun, C.D., Brown, J.C.S. and Madhani, H.D. A major role for capsule-independent
597 phagocytosis-inhibitory mechanisms in mammalian infection by *Cryptococcus neoformans*. *Cell*
598 *Host Microbe* 2011;9(3):243-251.

599 Denning, D.W. Global incidence and mortality of severe fungal disease. *Lancet Infect Dis*
600 2024;24(7):e428-e438.

601 Do, E.*, et al.* A Transcriptional Regulatory Map of Iron Homeostasis Reveals a New Control
602 Circuit for Capsule Formation in Cryptococcus neoformans. *Genetics* 2020;215(4):1171-1189.

603 Fromtling, R., Shadomy, H. and Jacobson, E. Decreased virulence in stable, acapsular mutants of
604 *Cryptococcus neoformans*. *Mycopathologia* 1982;79:23-29.

605 Garcia-Santamarina, S.*, et al.* Genome-wide analysis of the regulation of Cu metabolism in
606 Cryptococcus neoformans. *Mol. Microbiol.* 2018;108(5):473-494.

607 Gish, S.R.*, et al.* Computational Analysis Reveals a Key Regulator of Cryptococcal Virulence and
608 Determinant of Host Response. *mBio* 2016;7(2):e00313-00316.

609 Gish, S.R.*, et al.* Computational Analysis Reveals a Key Regulator of Cryptococcal Virulence and
610 Determinant of Host Response. *MBio* 2016;7(2).

611 Granger, D.L., Perfect, J.R. and Durack, D.T. Virulence of *Cryptococcus neoformans*. Regulation
612 of capsule synthesis by carbon dioxide. *J Clin Invest* 1985;76(2):508-516.

613 Greenfield, A.*, et al.* DREAM4: Combining genetic and dynamic information to identify biological
614 networks and dynamical models. *PLoS One* 2010;5(10):e13397.

615 Hackett, S.R.*, et al.* Learning causal networks using inducible transcription factors and
616 transcriptome-wide time series. *Mol. Syst. Biol.* 2020;16(3):e9174.

617 Harbison, C.T.*, et al.* Transcriptional regulatory code of a eukaryotic genome. *Nature*
618 2004;431(7004):99-104.

619 Haynes, B.C.*, et al.* Toward an Integrated Model of Capsule Regulation in *Cryptococcus*
620 *neoformans*. *PLoS Pathog.* 2011;7(12):e1002411.

621 Homer, C.M.*, et al.* Intracellular Action of a Secreted Peptide Required for Fungal Virulence. *Cell*
622 *Host Microbe* 2016;19(6):849-864.

623 Hu, Z., Killion, P.J. and Iyer, V.R. Genetic reconstruction of a functional transcriptional regulatory
624 network. *Nat. Genet.* 2007;39(5):683-687.

625 Huang, M.Y.*, et al.* An auxin-inducible degron system for conditional mutation in the fungal
626 meningitis pathogen Cryptococcus neoformans. *G3 (Bethesda)* 2025.

627 Huynh-Thu, V.A. and Geurts, P. dynGENIE3: dynamical GENIE3 for the inference of gene
628 networks from time series expression data. *Sci. Rep.* 2018;8(1):3384.

629 Jackson, C.A.*, et al.* Gene regulatory network reconstruction using single-cell RNA sequencing of
630 barcoded genotypes in diverse environments. *Elife* 2020;9.

631 Jung, K.W.*, et al.* Systematic functional profiling of transcription factor networks in
632 *Cryptococcus neoformans*. *Nat Commun* 2015;6:6757.

633 Kang, Y., Jung, W.J. and Brent, M.R. Predicting which genes will respond to transcription factor
634 perturbations. *G3 (Bethesda)* 2022.

635 Kang, Y.*, et al.* Dual threshold optimization and network inference reveal convergent evidence
636 from TF binding locations and TF perturbation responses. *Genome Res.* 2020;30(3):459-471.

637 Kang, Y.S.*, et al.* Leveraging a new data resource to define the response of *Cryptococcus*
638 *neoformans* to environmental signals. *Genetics* 2025;229(1):1-29.

639 Kelliher, C.M.*, et al.* Investigating Conservation of the Cell-Cycle-Regulated Transcriptional
640 Program in the Fungal Pathogen, Cryptococcus neoformans. *PLoS Genet* 2016;12(12):e1006453.

641 Kemmeren, P.*, et al.* Large-scale genetic perturbations reveal regulatory networks and an
642 abundance of gene-specific repressors. *Cell* 2014;157(3):740-752.

643 Kozel, T.R. Non-encapsulated variant of *Cryptococcus neoformans*. II. Surface receptors for
644 cryptococcal polysaccharide and their role in inhibition of phagocytosis by polysaccharide.
645 *Infect Immun* 1977;16(1):99-106.

646 Kumar, P.*, et al.* Emerging themes in cryptococcal capsule synthesis. *Curr. Opin. Struct. Biol.*
647 2011;21(5):597-602.

648 Kwon-Chung, K.J. and Rhodes, J.C. Encapsulation and melanin formation as indicators of
649 virulence in *Cryptococcus neoformans*. *Infect Immun* 1986;51(1):218-223.

650 Love, M.I., Huber, W. and Anders, S. Moderated estimation of fold change and dispersion for
651 RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.

652 Lupo, O.*, et al.* The architecture of binding cooperativity between densely bound transcription
653 factors. *Cell Syst* 2023;14(9):732-745 e735.

654 Ma, C.Z. and Brent, M.R. Inferring TF activities and activity regulators from gene expression
655 data with constraints from TF perturbation data. *Bioinformatics* 2021;37(9):1234-1245.

656 Madar, A.*, et al.* DREAM3: network inference using dynamic context likelihood of relatedness
657 and the inferelator. *PLoS One* 2010;5(3):e9803.

658 Mahendrawada, L.*, et al.* Surprising connections between DNA binding and function for the
659 near-complete set of yeast transcription factors. *bioRxiv* 2023.

660 Mahendrawada, L.*, et al.* Low overlap of transcription factor DNA binding and regulatory
661 targets. *Nature* 2025.

662 Maier, E.J.*, et al.* Model-driven mapping of transcriptional networks reveals the circuitry and
663 dynamics of virulence regulation. *Genome Res.* 2015;25(5):690-700.

664 Margolin, A.A.*, et al.* ARACNE: an algorithm for the reconstruction of gene regulatory networks
665 in a mammalian cellular context. *BMC Bioinformatics* 2006;7 Suppl 1:S7.

666 Meya, D.B. and Williamson, P.R. Cryptococcal Disease in Diverse Hosts. *N Engl J Med*
667 2024;390(17):1597-1610.

668 Mindel, V.*, et al.* Intrinsically disordered regions of the Msn2 transcription factor encode
669 multiple functions using interwoven sequence grammars. *Nucleic Acids Res.* 2024;52(5):2260-
670 2272.

671 O'Meara, M.J.*, et al.* CryptoCEN: A Co-Expression Network for Cryptococcus neoformans reveals
672 novel proteins involved in DNA damage repair. *PLoS Genet* 2024;20(2):e1011158.

673 O'Meara, T.R.*, et al.* The *Cryptococcus neoformans* Rim101 transcription factor directly
674 regulates genes required for adaptation to the host. *Mol Cell Biol* 2014;34(4):673-684.

675 Rajasingham, R.*, et al.* Global burden of disease of HIV-associated cryptococcal meningitis: an
676 updated analysis. *Lancet Infect. Dis.* 2017;17(8):873-881.

677 Reuwsaat, J.C.V.*, et al.* The Transcription Factor Pdr802 Regulates Titan Cell Formation and
678 Pathogenicity of Cryptococcus neoformans. *mBio* 2021;12(2).

679 Rossi, M.J.*, et al.* A high-resolution protein architecture of the budding yeast genome. *Nature*
680 2021;592(7853):309-314.

681 Roy, S.*, et al.* Integrated module and gene-specific regulatory inference implicates upstream
682 signaling networks. *PLoS Comput. Biol.* 2013;9(10):e1003252.

683 Siahpirani, A.F. and Roy, S. A prior-based integrative framework for functional transcriptional
684 regulatory network inference. *Nucleic Acids Res* 2017;45(4):e21.

685 Skok Gibbs, C.*, et al.* High-performance single-cell gene regulatory network inference at scale:
686 the Inferelator 3.0. *Bioinformatics* 2022;38(9):2519-2528.

687 Srikanta, D.*, et al.* A sensitive high-throughput assay for evaluating host-pathogen interactions
688 in *Cryptococcus neoformans* infection. *PLoS One* 2011;6(7):e22773.

689 Taylor, J.W. and Berbee, M.L. Dating divergences in the Fungal Tree of Life: review and new
690 analyses. *Mycologia* 2006;98(6):838-849.

691 Tugume, L.*, et al.* Cryptococcal meningitis. *Nat Rev Dis Primers* 2023;9(1):62.

692 Wang, Z.A., Li, L.X. and Doering, T.L. Unraveling synthesis of the cryptococcal cell wall and
693 capsule. *Glycobiology* 2018;28(10):719-730.

694