RESEARCH ARTICLE

CryptoCEN: A Co-Expression Network for Cryptococcus neoformans reveals novel proteins involved in DNA damage repair

Matthew J. O'Meara [1]*, Jackson R. Rapala [2], Connie B. Nichols[3], A. Christina Alexandre [2], R. Blake Billmyre [4], Jacob L Steenwyk [5], J. Andrew Alspaugh [3], Teresa R. O'Meara [2]*

1 Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, United States of America, **2** Department of Microbiology and Immunology, University of Michigan, Ann Arbor, Michigan, United States of America, **3** Departments of Medicine and Molecular Genetics/Microbiology; and Cell Biology, Duke University School of Medicine, Durham, North Carolina, United States of America, **4** Departments of Pharmaceutical and Biomedical Sciences/Infectious Disease, College of Pharmacy/ College of Veterinary Medicine, University of Georgia, Athens, Georgia, United States of America, **5** Howard Hughes Medical Institute and the Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, California, United States of America

* maom@umich.edu (MJO); tromeara@umich.edu (TRO)

OPEN ACCESS

Citation: O'Meara MJ, Rapala JR, Nichols CB, Alexandre AC, Billmyre RB, Steenwyk JL, et al. (2024) CryptoCEN: A Co-Expression Network for Cryptococcus neoformans reveals novel proteins involved in DNA damage repair. PLoS Genet 20(2): e1011158. https://doi.org/10.1371/journal.pgen.1011158

Editor: Xiaorong Lin, University of Georgia, UNITED STATES

Received: January 5, 2024

Accepted: January 30, 2024

Published: February 15, 2024

Data Availability Statement: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper and its Supporting information files or uploaded to figshare at DOI: 10.6084/m9.figshare.23979342, https://figshare.com/articles/dataset/CryptoCEN_data_v1_0_0/23979342.

Funding: This work was supported by NIH (R35GM147894 to TRO), NSURP (to ACA), NIH (AI175711 to JAA), NIH (T32GM007544 to JRR).

Abstract

Elucidating gene function is a major goal in biology, especially among non-model organisms. However, doing so is complicated by the fact that molecular conservation does not always mirror functional conservation, and that complex relationships among genes are responsible for encoding pathways and higher-order biological processes. Co-expression, a promising approach for predicting gene function, relies on the general principal that genes with similar expression patterns across multiple conditions will likely be involved in the same biological process. For *Cryptococcus neoformans*, a prevalent human fungal pathogen greatly diverged from model yeasts, approximately 60% of the predicted genes in the genome lack functional annotations. Here, we leveraged a large amount of publicly available transcriptomic data to generate a *C*. *neoformans* Co-Expression Network (CryptoCEN), successfully recapitulating known protein networks, predicting gene function, and enabling insights into the principles influencing co-expression. With 100% predictive accuracy, we used CryptoCEN to identify 13 new DNA damage response genes, underscoring the utility of guilt-by-association for determining gene function. Overall, co-expression is a powerful tool for uncovering gene function, and decreases the experimental tests needed to identify functions for currently under-annotated genes.

Author summary

A central problem in genetics is the connection between genotype and phenotype. Computational approaches to predict gene function can be especially useful for non-model organisms where extensive functional testing has not yet been performed. Co-

JLS is a Howard Hughes Medical Institute Awardee of the Life Sciences Research Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: JLS is an advisor for ForensisGroup Incorporated.

expression to predict gene function is based on the principle that genes that share similar expression patterns across multiple environmental conditions or perturbations are likely to be involved in the same biological process. Here, we collected transcriptomic data from the *Cryptococcus neoformans* field and built a robust co-expression network for predicting gene function, especially biological process information. Not only are we able to use this network for retrospective analysis of known gene clusters, but we are also able to make prospective predictions about gene function, including the well-studied processes of capsule and ergosterol biosynthesis. We also discovered a new role for 13 genes in the response to DNA damaging agents, showing that co-expression can reveal new players in conserved biological processes.

Introduction

The availability of genome sequences from non-model organisms has greatly increased with the development of high-throughput sequencing techniques; however, insights into gene function have not kept pace. Without functional information, it is difficult to interpret the results of genome-scale experiments, hindering our ability to model and predict cellular function. There are several computational methods for predicting protein function, including machine-learning approaches based on sequence [1,2], structure [3], homology, literature co-reference, and integrated bioinformatics [4]. Co-expression, the coordinated regulation of two genes across various conditions, is observed among genes with similar function [5], and the principle of guilt-by-association has proven to be a successful approach for predicting gene function based on co-expression patterns [6ï¿½8]. Co-expression has consistently been a strong source for gene function prediction; for example, a global effort aiming to improve gene function predictionï¿½ the Critical Assessment of Function Annotation 3 challengeï¿½ found that expression-based approaches outperformed all others [9].

Fungal pathogens pose a growing threat to human welfare. *Cryptococcus neoformans*, an opportunistic human fungal pathogen that causes lethal meningitis in immunocompromised individuals if left untreated, was recently given the highest priority of public importance by the World Health Organization [10]. Of the major human fungal pathogens, *C. neoformans* is also notable for being a basidiomycete, having diverged from the ascomycete lineage, which contains the model yeast *Saccharomyces cerevisiae* and the majority of human fungal pathogens approximately 400 million years ago [11]. Much of the current gene annotation information comes from inferred orthology with *S. cerevisiae*, but only 17% of *C. neoformans* genes have *S. cerevisiae* ortholog annotations due to the large evolutionary distance between these organisms. Thus, nearly 60% of the genes in the *C. neoformans* genome remain ï¿½ hypothetical ï¿½ unspecified ï¿½ and lack any computed or curated biological process Gene Ontology (GO) term information [12,13]. Additionally, there have been numerous examples where sequence conservation between *C. neoformans* proteins and model yeast failed to predict gene function [14ï¿½16].

As *C. neoformans* is less genetically tractable than model yeast, researchers have taken advantage of transcriptomic approaches to help understand how the cells respond to various environmental and genetic perturbations. For example, differential expression analysis helped identify multiple stages of *C. neoformans* infection of the host [17], including the extensive cell wall remodeling during host infection. Other studies have performed transcriptional profiling across mutant strains to identify regulatory relationships between transcription factors that control capsule formation, the key virulence factor in *C. neoformans* [18], or the response to environmental pH [19]. Previous investigators have used computational approaches to predict

gene function in *C. neoformans* through the generation of CryptoNetV1 [20]. However, advances in transcriptomics, predictive algorithms, and structural modeling now allow for more rigorous strategies in assigning potential function to unannotated genes.

Here, we leveraged the extensive publicly available transcriptional profiling data generated by the *C. neoformans* research community to build a *C. neoformans* co-expression network (termed CryptoCEN) that captures multiple dimensions of genetic and cellular function. For example, CryptoCEN uncovered the capsule co-expression network that encompassed cell cycle processes, revealed a co-expression signature between ergosterol and translation, and identified a role for the Cdc42 and Cdc420 paralogs in kinetochore assembly. Importantly, we were also able to demonstrate the high predictive value of co-expression by prospectively identifying 13 new proteins involved in DNA damage responses, a set of pathways responsible for maintaining genome integrity that are thought to be promising targets for therapeutics [21]. We also identified new proteins involved in capsule, a critical virulence factor, and those that contribute to fluconazole susceptibility. Together, this demonstrates the functional utility of co-expression in understanding gene function.

Results

Generating CryptoCEN, a global co-expression network that captures genomic function

We generated a co-expression network for *C. neoformans* based on the general principles established in our previous implementation of a co-expression network for *C. albicans* using our CalCEN R package [7]. First, we collected RNA sequencing (RNAseq) data from the NCBI Sequence Read Archives (SRA). We chose studies that included at least 8 samples and filtered for those examining the *C. neoformans* H99 type strain and its derivatives, including KN99. This resulted in 1,524 runs across 34 studies (Fig 1A and S1 Table). The conditions for these experiments included a wide range of environmental perturbations, such as differences in nutrient source, cell cycle, chemical perturbations, and genetic mutation [18,19,22ï¿½27]. To ensure uniformity in analysis and data processing, the raw reads from each study were re-mapped to the 9,189 transcripts from FungiDB release-49 of the *C. neoformans* H99 reference transcriptome using RSEM with bowtie2 [28,29]. Similar to our work in *C. albicans*, to ensure sufficient coverage, we removed runs where greater than 50% of the genes had zero expression, yielding 1,523 samples in total (S1 Table). We used fragments per kilobase of transcript per million mapped reads (FPKM) as the estimated expression for each gene under each condition, and then used Spearman rank correlation to measure the correlation between gene expression profiles with the EGAD R package [7,30]. Rather than pooling all expression profiles for all studies, we built separate co-expression networks for each study and combined the average across all networks, following best practices described by Ballouz and colleagues [6]. For studies designed to explore specific biology, the within-study expression variation may be more meaningful than the between-study variation, and therefore grouping by study can increase predictive accuracy. For each pair of *C. neoformans* genes, we then generated a value between 0 and 1 to represent the rank of co-expression among all pairs of genes (Fig 1B and S2 Table). To show the data underlying representative pairs of gene co-expression, we chose four levels of co-expression (0.4, 0.6, 0.8, 0.9) and four pairs of genes at each of these co-expression levels. The FPKM values for the genes from each of the 1523 samples, colored by study, are plotted (Fig 1C).

We then used UMAP dimensionality reduction and Louvain clustering methods [31,32] to obtain a global visualization of the overall *C. neoformans* gene by expression matrix (Fig 1D). To test the sensitivity of the embedding to UMAP parameters, we re-embedded the gene by

Fig 1. ***Cryptococcus*** Co-Expression analysis identifies both known and unknown gene clusters. A) A gene-by-environment heatmap generated from collected *C. neoformans* RNAseq experiments from the SRA. The genes are on the y axis, and conditions are on the x axis. B) A gene-by-gene heatmap generated from Spearman rank correlation. C) Representative gene-by-gene expression patterns for pairs of genes at four different co-expression scores. Each dot represents the expression in a single RNAseq run, and dots are colored by study. D) Embedding of the *C. neoformans* Co-Expression Network (CryptoCEN) using UMAP for dimensionality reduction. Twenty-two clusters were identified using Louvain modularity clustering. The genes in each cluster were analyzed for biological process GO term enrichment using Fisher's exact test through FungiDB, and the most significantly enriched specific term after Bonferroni correction for multiple testing was used for labelling. Clusters without significant GO term enrichment after multiple hypothesis testing correction were labeled only by cluster number. E) Genes in cluster 6 sub-clusters were analyzed for both biological process and cellular component GO term enrichment using Fisher's exact test through FungiDB. F) Genes in cluster 16 sub-clusters were analyzed for both biological process and cellular component GO term enrichment using Fisher's exact test through FungiDB.

https://doi.org/10.1371/journal.pgen.1011158.g001

expression matrix over a range of the key embedding parameters umap_a and umap_b. We then plotted the resulting embeddings as a lattice of scatter plots. The embeddings show that as umap_a and umap_b increase, the embeddings become more compact but largely show similar overall global and local structure (S1A Fig).

In the overall gene by expression UMAP, we identified 22 clusters (S3 Table) and used GO-term enrichment analysis using Fisher's exact test run through FungiDB of biological process to predict functional signatures for each group. There was a clear cluster for cell cycle proteins and DNA replication (cluster 11) and proteins involved in carbohydrate transport (cluster 14). We also observed sub-structures within each cluster, such as in cluster 6, where there were sub-clusters for the large and small ribosome proteins and for oxidative response proteins (Fig 1E). Additionally, in cluster 16, which was enriched for proteins involved in respiration, we could see a distinct sub-cluster for those proteins encoded within the mitochondrial genome (Fig 1F). Recapitulating known functional groups of genes demonstrate the efficacy of our approach. However, we also identified 8/22 clusters for which there was no enrichment for any biological process GO term annotation, highlighting the lack of functional information for many genes in the *C*. *neoformans* genome.

Benchmarking CryptoCEN via Retrospective Gene Function Analysis

There are 8,334 genes currently identified in the *C*. *neoformans* genome, and 2,754 of them are uncharacterized with no functional annotation across homology or gene function prediction tools (Fig 2A). A goal of the co-expression network is to predict gene function for these under-annotated genes. Therefore, we first benchmarked the ability of the CryptoCEN to retrospectively predict all of the current *C*. *neoformans* GO term annotations without the NOT qualifier collected from FungiDB [33]. Since GO terms form an ontological hierarchy, we propagated annotations from more specific to more general terms and then filtered for terms having between 20 and 1,000 annotations. For this benchmarking, we used guilt-by-association, where the strength of the GO term prediction is based on the fraction of neighbors (or weighted sum for a network where edges are weighted e.g., by co-expression score) in the network with a given term. We then compared our predictions with the current set of *C*. *neoformans* predictions using the area under the receiver operating characteristic curve (AUROC) score and averaging over a 10-fold cross-validation. True-positive results were defined when the CryptoCEN predicted GO term matched the known GO term from FungiDB and the AUROC score provided an unbiased measure of enrichment and network quality. Using this dataset, we found that the CryptoCEN network had a neighbor-voting AUROC of 0.74 (+/- 0.093) (Fig 2B), suggesting that CryptoCEN was not making predictions based on multifunctional genes [7,34].

In comparison, using information from orthologous systems, such as BLAST or CryptoNetV1 [20], resulted in AUROCs of 0.78 (+/- 0.14) and 0.72 (+/- 0.11). Using orthologous physical or genetic associations for *S*. *cerevisiae* had a predictive power roughly on par with random chance (AUROCs of 0.56 ± 0.095 and 0.55 ± 0.87, respectively) (S1B Fig). In contrast with BLAST or CryptoNetV1, which had better relative prediction over biological process and molecular function, the CryptoCEN network had a relatively better prediction for the cellular component terms. To quantify if orthologous systems are complementary, we evaluated AUROC over all combinations of networks by summing over edge weights. The predictive performance of CryptoCEN in combination with BLAST and CryptoNetV1 offered a small but measurable increase compared to the combined predictive power of the two established models. For both BLAST and CryptoNetV1, adding CryptoCEN substantially increased performance to AUROCs of 0.86 (+/- 0.071) and 0.88 (+/- 0.076) respectively, and

Fig 2. CryptoCEN can retrospectively predict biological process GO terms. A) The *C. neoformans* genome contains many unannotated genes. UpSet plot of gene annotation information from different sources. Each bar in the upper region shows the number of gene nodes in the intersection of the set of databases indicated by the rows with filled circles in the lower region. The first column represents genes that are not included in any of the current annotation databases, including orthology to the model yeast *S. cerevisiae*. B) Combining CryptoCEN with other sources of information increases retrospective prediction accuracy. UpSet plot of the retrospective prediction accuracy, as determined by the neighbor voting guilt-by-association (GBA) area under the ROC curve (AUROC). AUROCs values range between 0.5 for random predictor and 1 for a perfect predictor. As data sources are combined, the prediction accuracy increases. Each annotated GO term is colored by ontology biological process (BP), cellular component (CC), or molecular function (MF). C) Prediction accuracy increases as the number of studies included in the network increases. Mean neighbor voting GBA performance for the CryptoCEN built over random subsets of RNAseq studies. The blue curve represents a mean of a nonparametric locally estimated scatterplot scattering (LOESS) fit.

https://doi.org/10.1371/journal.pgen.1011158.g002

all three combined lead to the highest overall performance of 0.91 (+/- 0.06), compared with 0.90 (+/- 0.64) for BLAST and CryptoNetV1 only. Using all three networks specifically benefited biological process and cellular component prediction (Fig 2B). This demonstrates that CryptoCEN captures information undetected in previous prediction tools and that adding RNAseq-based co-expression can increase the quality of gene function predictions, at least retrospectively.

A potential limitation of CryptoCEN is that the transcriptome space has not been fully explored, and additional RNAseq studies may be needed to break spurious correlated expression by measuring under new environmental contexts. To address this question, we assessed how the predictive accuracy of the network changes as we remove studies from the network. For each *k* in the range [1, 33], we sampled different combinations of studies of size *k* and estimated their ability to predict gene function annotations. We found that as we increased the studies from 1ï¿½10, there was a rapid rise in average performance from ~0.64 to ~0.71. After this point, performance increases steadily to 0.74 when approaching the full set of 34 studies (Fig 2C). This suggests that we have sufficient studies of the transcriptome to build a robust co-expression network, but that the accuracy will improve with the integration of additional studies. Although the number of RNAseq studies needed to generate a co-expression network will vary based on genetic background and the fraction of transcriptome space represented, this finding may also help inform the experimental design to generate co-expression networks in other organisms.

While the embedding (Fig 1D) and co-expression network are derived from the same data, due to the different processing strategies, they may diverge. To examine this, for each embedding cluster, we computed how much the intra-cluster associations are enriched over the inter-cluster associations for closeness in the co-expression network. Specifically, for each cluster, we selected 1000 query genes independently at random from the cluster and a second gene independently at random from within the cluster and labeled these edges as intra-cluster associations. Then, for each of the query genes, we selected a gene independently at random from all genes not in the cluster and labeled these edges inter-cluster associations. We then computed the area under the receiver operator characteristic (AUROC) for the enrichment of the intra-cluster associations over the inter-cluster associations based on the co-expression score. Recall that if there was no statistical enrichment, we would expect the AUROC to be ~0.5 and if there was perfect enrichment (all intra-cluster associations had higher co-expression scores than inter-cluster associations) then the AUROC would be 1. We found that the embedding of cluster enrichment scores ranged from ~0.5 to ~0.9, with higher enrichment for clusters on the left (S1C Fig). This suggests that the co-expression provides additional functional enrichment beyond just the gene by expression matrix embedding.

Evolutionary constraints can inform co-expression

Among proteins that form physical interactions, we hypothesized that those participating in obligate complexes may have stronger selective pressure for co-expression. For example, if all the sub-units need to be expressed at stoichiometric levels for the complex be functional, then the metabolic cost of expressing isolated subunits is wasted, leading to a selective pressure for co-expression [35]. Moreover, incorrect expression outside of stoichiometric ratios can generate proteotoxic stress, as the uncomplexed subunits are actively detrimental to the cell, as demonstrated during cases of aneuploidy [36].

Given that there is not a well-curated list of complexes in *C. neoformans*, we leveraged the better annotated complexes in *S. cerevisiae* and used sequence homology to infer candidate complexes. To support this strategy, we reasoned that complexes that are highly evolutionarily conserved may have robust cooperative function. We collected *S. cerevisiae* complexes from EBI (2021-10-13) and filtered out sub-units that were nucleic acid, small molecule, mitochondrially encoded, or unrecognized, yielding 616 multi-subunit complexes. Of these, 408 had at least 2 distinct subunits with a one-to-one ortholog in *C. neoformans* identified by OrthoMCL [37]. Within these complexes, there are 13,950 pair-wise co-complex interactions of *C. neoformans* proteins (S4 Table). Of these, over half (7,142) are interactions within the 17 annotated

ribosomal or ribonucleic protein complexes, with an average co-expression score of 0.84. The other half of the candidate interactions (6,808) had an average co-expression score of 0.75 (Fig 3A). This finding supports the hypothesis that there is selective pressure to co-express members of protein complexes.

A recently developed approach to predict function is to use co-evolutionary networks; beyond being part of the same complex, co-evolution is based on the principle that functionally related genes will show similar rates of evolution across speciation events [38]. To calculate this, the evolutionary rate values are estimated for each branch of an orthologous gene's phylogeny and compared with each other gene to determine the rate of co-evolution [38]. We hypothesized that combining a co-evolutionary network would increase the accuracy of the co-expression network. Therefore, we generated a co-evolutionary network for *C. neoformans* using 15 related species (S2A Fig), with the caveat that the species tree of related organisms is much sparser for *C. neoformans* and relatives compared to the 331 available Saccharomycotina yeast, making this analysis less robust that the established *S. cerevisiae* co-evolution information [38]. We then examined whether combining the co-evolutionary network would increase the accuracy of the co-expression network. Here, we subset the analysis to those genes with clear ortholog groups (S5 Table). However, we identified very little correlation between co-expression and co-evolution networks (Fig 3B and S2B Fig), limiting our ability to integrate this information. This can be potentially attributed to the low density of closely related species genome sequences, which limits our ability to identify co-evolution signatures.

It is also possible to examine the evolutionary pressures on co-expression in the context of gene duplication. Gene duplication provides an opportunity for either shared functionalization, as in the case of the histone proteins, or neofunctionalization and sub-functionalization [39,40]. Among genes with clear ortholog groups (S5 Table), we examined the co-expression of each protein with every other protein in each orthogroup and plotted the distribution (Fig 3C). Given that the co-expression scores are ranked, a random distribution should be flat. However, we observed that the co-expression is less than random, suggesting that there is pressure to be differentially transcribed from other genes within an orthogroup. To characterize the distribution, we fit the distribution with a skew-normal curve, and the alpha parameter is significantly above zero (4.4 with a standard deviation of 0.56), indicating the extent of the skew. The outlier with high co-expression between orthologous genes were the H2A and H2B histone genes. This bias towards a shift in expression between paralogs is consistent with previous results in multiple systems, where changes in transcription are required for evolutionary divergence in function [41ï¿½43].

Many paralogs, when duplicated, have differences in baseline expression levels that explain their differential function. To examine whether overall expression is predictive of co-expression, we performed a global analysis comparing co-expression levels to overall expression levels. This demonstrated a small bias for genes with low expression against having high co-expression partners (Fig 3D). In particular, 80% of co-expressed pairs (coexp score > 0.8) have a geometric mean expression greater than 13 FPKM, while only 40% of all pairs have a geometric mean expression above this threshold. Therefore, this may not be a major driver of co-expression patterns.

In *C. neoformans*, we have also observed the duplication of entire signaling cascades, as opposed to single gene duplications [44]. One notable example is the duplication of Ras1/Ras2, Cdc42/Cdc420 and Rac1/Rac2 proteins [44ï¿½46], with each paralog supporting specific cellular functions. For example, Cdc42 serves as a critical regulator of thermotolerance and virulence [46,47], while Cdc420 plays a minor role in the expression of virulence-associated phenotypes under basal conditions [48]; this is consistent with the higher baseline expression of Cdc42 compared with Cdc420. Importantly, Cdc42 is critical for septin localization, with Cdc10

Fig 3. Evolutionary constraints inform co-expression analyses. Distribution of co-expression scores for *C. neoformans* gene pairs across different types of evolutionary conservation. A) *C. neoformans* gene pairs with orthology to *S. cerevisiae* gene pairs that encode for proteins that are members of the same complex (13,950 pairs over 1,304 genes, coexp score mean: 0.80, IRQ50: [0.71, 0.90]), B) Significantly co-evolving gene pairs (140,592 pairs over 4,269 genes, coexp score mean: 0.50, interquartile-range at 50% (IRQ50): [0.40, 0.60]), and C) Paralogous gene pairs (1,056 pairs over 550 genes, coexp score mean: 0.58, IRQ50: [0.46, 0.67]). D) Scatter plot of the co-expression score by the geometric expression of the partners. E) The co-expressed partners at a 0.8 threshold for co-expression score of the duplicated genes Cdc42 and Cdc420 were compared and visualized in Cytoscape. Kinetochore proteins are highlighted in blue, septin proteins in purple, and unannotated or uncharacterized proteins are highlighted in yellow. Width of the lines indicates co-expression score. F) Tubulin is altered in the *cdc420*Δ and *cdc42*Δ mutant strains compared with the H99 wildtype. Tubulin was visualized by fusion of α-tubulin with GFP. Cells were incubated in liquid YPD at 30ï¿½C before imaging. Images taken at 40X magnification, scale = 5 microns.

https://doi.org/10.1371/journal.pgen.1011158.g003

completely mis-localized in the *cdc42*Δ mutant, likely explaining similar defects in thermotolerance in the *cdc42*Δ and septin mutant strains [48].

We analyzed the co-expression partners for the paralogous Cdc42 and Cdc420 proteins, with visualization using Cytoscape, where nodes represent genes and edges connect two co-expressed genes, with a co-expression score threshold of 0.8. From this, we identified entirely distinct networks for the two paralogs. Consistent with prior genetic and physiological data, Cdc42 was highly co-expressed with multiple septin proteins, including Cdc10 (Fig 3E). In contrast, Cdc420 was co-expressed with 10 kinetochore or spindle pole body proteins (Fig 3E), including parts of the middle layer, outer layer (Dam1/DASH), and spindle assembly checkpoint [49ï¿½52]. Previous work on physical interaction partners for kinetochore proteins revealed that Spc25 interacts with Cdc420 [53], giving additional evidence for the predicted relationship between Cdc420 and kinetochores.

Therefore, to test whether a loss-of-function mutation of Cdc420 would alter kinetochore function, we examined microtubule and nuclear dynamics in strains with either fluorescently-tagged α-tubulin (Tub1) or histone H4 proteins [49] in the WT, *cdc42*Δ, and *cdc420*Δ strains, using these as a proxy for kinetochore function. The H4-GFP protein highlighted a similarly well-defined nucleus in each strain and each growth condition (S2C Fig). In the wild-type strain, expression of the Tub1-GFP fusion protein demonstrated clear and elongated microtubules in cells undergoing cell division (Fig 3F). In contrast, both the *cdc42*Δ and *cdc420*Δ strains showed more cells with punctate rather than polymerized GFP signal, consistent with a relative defect in α-tubulin polymerization into microtubules (Fig 3F). Although there may be differences in this polymerization between the mutant strains, we were unable to resolve it and identify a clear defect in just the *cdc420*Δ strain. Moreover, all strains showed similar growth in response to the microtubule destabilizing compound nocodazole (S2D Fig). These data demonstrate that there is an association between kinetochore function and both the Cdc420/Cdc42 proteins, and they also suggest that there is not a specific defect in microtubule assembly in the *cdc420*Δ compared with the *cdc42*Δ mutant strain. However, it is possible that there is a more specific defect in kinetochore function that is not captured in this phenotypic assay. It is also possible that the large set of uncharacterized genes (Fig 3D) that are co-expressed with *CDC42* and *CDC420* have roles in kinetochore or microtubule assembly, but the lack of current annotation prevents us from making those connections. This demonstrates that co-expression can help identify hypotheses to test, but may not have the resolution to identify a specific phenotype for a given mutant strain.

Virulence factor retrospective cluster analysis

To understand how the co-expression network can broadly identify new genes for a given function, we first retrospectively explored the network localized to the genes involved in well-studied functions.

The major virulence factor of *C. neoformans* is the polysaccharide capsule [54]. To seed the network we used a set of known capsule biosynthetic genes [54], identified all the first neighbors in the network with a co-expression score higher than 0.8, and then selected those with at least 5 co-expression edges to other genes in the set for visualization using Cytoscape [55] (Fig 4A and S6 Table). Interestingly, this analysis revealed that although Cap59, Cap60, Cap64 and Cas35 are highly interconnected and co-expressed, there are other capsule biosynthetic genes, such as Cas33, Cas34, and Cap6 that are not directly co-expressed with any other capsule gene above the 0.8 score threshold. Given the importance of condition-dependent adaptations in other cell surface features to promote capsule attachment, we also saw co-enrichment for cell wall and membrane biosynthesis genes in this capsule gene cluster. The other major signature

Fig 4. CryptoCEN can recapitulate core biological processes in ***C. neoformans***. A) A co-expression network for capsule was generated by starting with genes known to be involved in capsule biosynthetic genes. All co-expressed partners with a score > 0.8 and at least 5 co-expression edges with known capsule genes were visualized in Cytoscape. Specific functional classes are highlighted with different colors. Edge width corresponds to degree of co-expression. B) Identification of genes involved in capsule. The indicated mutants were incubated in RPMI at 37ï¿½C with 5% CO_2 for three days before staining with India ink and imaging using brightfield microscopy at 20X magnification. Increased or decreased capsule was determined by comparison with the wild type or *cap64*Δ control strains. C) A co-expression network for ergosterol biosynthesis was started with the known ergosterol biosynthetic genes and all co-expression partners that showed >0.8 co-expression score and interaction with >3 ergosterol biosynthetic genes. Specific functional classes are highlighted with different colors. Edge width corresponds to degree of co-expression. D) Identification of genes involved in fluconazole susceptibility. The indicated strains were grown overnight at 30ï¿½C liquid YPD medium, and the serially diluted cells were spotted onto YPD agar with or without 4 μg/mL fluconazole. The plates were incubated at 30ï¿½C and imaged after 2 days.

https://doi.org/10.1371/journal.pgen.1011158.g004

in the remaining genes was for proteins involved in cell cycle and DNA maintenance. This connection is consistent with previous literature on cell cycle and capsule in *C. neoformans* [16,56]. Notably, the only transcriptional regulators that appeared to show co-expression with capsule genes were orthologues to Ctk2 and the uncharacterized transcription factor Fcz4, despite the known importance of many other specific transcription factors in regulating capsule [15,57–59]. This complements the study from Kim et al., which focused on the capsule regulatory genes in CryptoNet and observed connections between the regulatory cascades but not the capsule biosynthetic genes [20]. Potentially, this is due to the regulation of signaling cascades at the post-translational level, rather than the transcript level. To test the prospective accuracy of CryptoCEN for capsule-related genes, we tested 12 mutant strains for genes in the capsule network for their ability to generate capsule. We identified six mutant strains with a defect in capsule, and two with an increased capsule. Notably, none of these capsule-deficient mutant strains exhibited dry colony morphology, suggesting that the defect may be at the level of capsule maintenance at the cell surface, as the strains are still able to secrete enough capsule to generate a mucoid colony.

As a second test case, we examined the ergosterol biosynthetic cascade, as ergosterol and its biosynthesis are important antifungal drug targets [60]. We seeded the network with 23 orthologs of the known ergosterol biosynthetic machinery and filtered for the first neighbors with at least three co-expression edges with ergosterol biosynthesis proteins and a score of at least 0.8 (Fig 4C and S6 Table). This resulted in a densely connected network, but with a somewhat surprising structure—the ergosterol genes were not organized in biosynthetic order. For example, Erg1, Erg3, Erg5, and Erg25 were highly interconnected despite operating in different parts of the ergosterol biosynthesis pathway. We also observed many proteins involved in translation with strong co-expression with the ergosterol genes, especially Erg20 and Erg6. To test the connection between these co-expressed genes with ergosterol, we turned to hypersensitivity to the antifungal agent fluconazole, which acts by targeting ergosterol biosynthesis and impacting membrane fluidity [60]. When testing available mutant strains for changes in fluconazole sensitivity, we observed that deletion of *CAP64*, *CKB1*, *BIM1*, and *CNAG_06753* resulted in increased susceptibility to fluconazole, whereas deletion of *CNAG_02755* and *ARP4* decreased fluconazole susceptibility (Fig 4D).

Overall, we were able to consistently replicate known networks through CryptoCEN, and potentially identify new signatures associated with core biological processes, and identify new genes involved in capsule and fluconazole susceptibility. This demonstrates that there is utility in using co-expression to explore gene function in *C. neoformans*.

Identification of additional genes involved in DNA repair, including novel uncharacterized proteins

DNA repair is an essential and highly conserved function in cells that ensures genome stability. From a biomedical perspective, SNPs in mismatch repair genes have been linked with hypermutator phenotypes in *C. neoformans*, allowing for increased drug resistance and virulence [61–63]. From an evolutionary perspective, comparative genomic analysis has revealed gene presence/absence variation among canonical DNA repair genes in other microeukaryotes [64,65], but the discovery of novel DNA repair genes has been lacking. In *C. neoformans*, the full set of DNA repair genes is unknown. We hypothesized that genes coexpressed with those known to contribute to DNA repair—RCC, MLH, MSH and RAD proteins identified by Ashton and colleagues [66]—would allow us to identify additional uncharacterized proteins involved in this core biological process.

Fig 5. Identification of new proteins involved in DNA damage responses. A) A co-expression network for DNA damage was started with 34 known genes involved in DNA repair, and all co-expression partners that showed > 0.8 co-expression score and interaction with at least 5 of the known DNA repair genes. Specific functional classes are highlighted with different colors. Edge width corresponds to co-expression score, and node size represents number of connections to other genes in the network. B) Identification of novel genes involved in DNA damage responses. The indicated strains were grown overnight at 30ï¿½C in liquid YPD medium, and the 10-fold serially diluted cells were spotted onto YPD agar. For UV damage, the plates were immediately subjected to 200 μJ UV. For EMS, the cells were incubated in 100 μM EMS for 1 hr before serial dilution and plating. The plates were incubated at 30ï¿½C and imaged after 2 days.

https://doi.org/10.1371/journal.pgen.1011158.g005

To do so, we filtered for co-expression scores > 0.8 and at least four co-expression edges with known DNA repair genes, and then used this set to generate a network that we visualized in Cytoscape (Fig 5A). This revealed a dense network of mismatch repair genes centered around *MSH2* and *MSH6*, and a sparser network with more of the ERCC excision-repair

genes. The initial seed set of DNA damage response genes (purple) was surrounded by genes with known functions in DNA damage and repair, including the *C. neoformans*-specific radiation resistance transcription factor, Bdr1 [67]. The centrality of some of the cell cycle and DNA replication genes in this network, including Bud14, Kel1, Pol3 and Rig2/Dbp11, highlights the strong association between DNA repair and cell cycle processes.

Beyond the known proteins in this DNA damage response network, we identified many highly co-expressed hypothetical proteins that we hypothesized may play a role in DNA repair, including proteins of unknown function as well as those with known roles in other aspects of cell biology. Therefore, we tested 13 available deletion mutants of these strains for their responses to either ethyl methanesulfonate (EMS) or UV radiation, as a measure of two different DNA damage response pathwaysï¿½ DNAalkylation and DNA lesions (S7 Table). Notably, all 13 of the mutants tested had a phenotype on EMS, including 9 resistant and 4 sensitive strains (Fig 5B). Of the four EMS-sensitive strains, two were also sensitive to UV damage.

To determine whether the predictive accuracy was above baseline, we tested a set of random mutants for their phenotypes on DNA damaging agents. A potential complication is that DNA replication and damage responses are amongst the most highly co-expressed genes in CryptoCEN. Therefore, we chose 12 genes that showed a similar rank of co-expression as our matched control set for determining the baseline rate of phenotypes in response to DNA damaging agents. Here, only 4 of the mutant strains showed a phenotype (S2 Fig), with two strains showing sensitivity to EMS and two showing slight resistance. Therefore, CryptoCEN has high prospective predictive accuracy for gene function annotation.

One of the UV-sensitive mutants was *MRC1*/*CNAG_03023*, a hypothetical protein with an MRC1 (mediator of replication checkpoint)-like domain with 13 co-expressed partners in the network. When we performed the reciprocal co-expression analysis using *MRC1* as the seed, in addition to the 13 previously identified partners, we identified an additional 13 genes with roles in chromatin binding, DNA damage responses, or cell cycle (S7 Table). The presence of the MRC1 domain in the CNAG_03023 sequence suggested a putative function for CNAG_03023 as a checkpoint protein that would be required for the response to multiple types of DNA damage.

The other UV-sensitive strain was the deletion for *CNAG_06573*, a basidiomycete-specific hypothetical protein with 5 co-expressed partners. This protein did not have any conserved domains, and although reciprocal co-expression analysis identified a further 18 proteins involved in DNA damage, cell cycle, and DNA binding, there was not a clear signature of function beyond this larger category (S7 Table). Therefore, we turned to structural homology, using AlphaFold Monomer v2.0 2022-11-01 [68] to give a predicted structure to use as an input to FoldSeek [69]. The well-structured N-terminal domain of CNAG_06573 (residues ~700ï¿½ 1200) had homology with the CATH Superfamily [70] Leucine-rich Repeat Variant (1.25.10.10), which contains DNA repair protein rad26 (rad26) from *S. pombe* (sequence identify of 11.4 and E-value of 2.5e-4), and the ATR-interacting protein (ATRIP) from *H. sapiens* (sequence identity of 11.2 and E-value of 2.1e-2) [71,72]. ATRIP domains recognize the Replication Protein A complex associated with single stranded DNA to facilitate DNA damage response [71,72]. Therefore, although CNAG_06573 does not share sequence similarity with these known DNA-damage response proteins, the structural homology suggests that CNAG_06573 may act as an ATRIP in *C. neoformans*.

The *CNAG_05141*Δ mutant had minor sensitivity to EMS compared to the wild-type control, and was not sensitive to UV damage. Based on AlphaFold and FoldSeek analysis, this protein contained a predicted EamA-like transporter domain [73] and had structural homology to the Nipal2 and Nipa1 transporter proteins. Potentially, loss of this transporter may lead to a

higher intracellular concentration of EMS resulting in increased cell death. Notably, this gene was also co-expressed with multiple capsule biosynthetic genes.

The *CNAG_00571*Δ mutant also displayed an EMS-resistant phenotype. This gene encodes a hypothetical protein with a predicted karyogamy protein 9 (KAR9) domain. Kar9 facilitates nuclear congression during karyogamy in *S. cerevisiae* [74]. Previous work on the karyogamy machinery in *C. neoformans*, however, had indicated that an ortholog to *S. cerevisiae KAR9* (*ScKAR9*) was not present in *C. neoformans* [75]. These searches were based on sequence homology and synteny, and CNAG_00571 is not syntenic to *ScKAR9* and does not share major sequence homology, despite the predicted Kar9 domain. However, yeast are known to evolve rapidly, resulting in low syntenic conservation [76], and remote homology may be difficult to detect [77]. Using a structural homology-based approach, we found that the predicted structure of CNAG_00571 shared with homology with the CATH superfamily 1.20.58.70, which is enriched in syntaxin proteins which are known to facilitate membrane-membrane fusion events within the cell. However, the lack of sequence homology with *ScKAR9* and the placement of CNAG_00571 in a Tremellales-specific ortholog group (OG6-532064) indicates that this gene does not share an evolutionary history with *ScKAR9*. Despite this, we hypothesize that CNAG_00571 is indeed functioning as a Kar9 protein during karyogamy by facilitating nuclear-nuclear fusion in *C. neoformans*, and thus we propose naming CNAG_00571 as *KAR9*.

We attempted a similar structural-based approach for CNAG_06984, which showed specific hypersensitivity to EMS. This protein may be involved in double-stranded break repair, based on the specific hypersensitivity phenotype [78]. However, neither structure nor sequence-based approaches yielded any related proteins with high confidence. Furthermore, three genes specific to the Tremella lineage (CNAG_02866, CNAG_07605, CNAG_02930, whose mutants were all resistant to EMS) had low-quality structural predictions, likely due to a lack of the required training data.

Beyond the uncharacterized proteins, we also identified two genes that had previously been implicated in different cellular pathways, but still showed an EMS-resistant phenotype. Ccz1 (*CNAG_04456*) is a guanidine exchange factor (GEF) that forms an active complex with Mon1 (*CNAG_00971*) [79], although these two genes are not co-expressed in *C. neoformans*. In *S. cerevisiae*, this Ccz1-Mon1 complex activates Rab7, which is involved in intracellular trafficking to the lysosome, including trafficking of the autophagosome to the lysosome [80–82]. Autophagy is known to be induced by DNA damage, where the inhibition of autophagy can sensitize cancer cells to DNA damage [83–85]. Therefore, we hypothesize that the loss of *CCZ1* in *C. neoformans* leads to a higher overall level of autophagy, which may provide a protective effect against DNA damage. GO term enrichment of *CCZ1* co-expression partners also showed enrichment for macroautophagy (p = 1.87e-4).

The other previously annotated gene implicated in DNA repair using CryptoCEN is *CMK1*, which is more resistant to EMS. This gene encodes a calmodulin-dependent kinase (CaMK) and serves as an effector of the calcium-calcineurin signaling pathway, an important component of fungal stress responses [86–88]. Previous studies have shown that this pathway can play a role in cell cycle control in *S. cerevisiae* and other organisms [89], with a function at the G_2/M checkpoint [90–92]. In *C. neoformans*, loss of *CMK1* may inhibit the mutant from the appropriate stress-induced cell cycle arrest, leading to increased growth in the presence of the cellular stress. GO term enrichment of the *CMK1* co-expression partners shows enrichment for DNA recombination, metal ion transport, and negative regulation of exit from mitosis (S7 Table). Together, this analysis of the DNA damage co-expression network identified multiple new proteins involved in DNA damage response in *C. neoformans*, including proteins without sequence or structural homology to known DNA damage response proteins.

Discussion

Cryptococcus neoformans, despite being a deadly human fungal pathogen and the cause of mortality for nearly 200,000 people per year [93], has a genome that is vastly under-annotated. This lack of functional annotation information makes it difficult to interpret or identify genetic signatures or shed light on genotype-phenotype associations. Co-expression analysis provides a platform for predicting gene function, thus potentially decreasing the experiments needed to define the function of a gene. Recently, we generated a co-expression network for the model fungal pathogen *Candida albicans*, based on available RNAseq data from the reference strain SC5314, which allowed us to predict the function of gene as a cell cycle chaperone protein [7]. Co-expression across diverse clinical isolates of *C*. *albicans* was able to identify regulators of morphogenesis and virulence [94]. These, and others, demonstrated the utility of co-expression in understanding fungal pathogen biology [95,96]. However, these studies were performed in ascomycetes, which are closer to the model yeast *S*. *cerevisiae*, and thus there is much that can be inferred from orthology in these organisms. For the basidiomycete *C*. *neoformans*, there are more genes without clear orthologs and gene annotation information, making the need for computational predictions more urgent. Here, we leveraged publicly available RNAseq data from the *C*. *neoformans* research community to build a robust co-expression network for *C*. *neoformans*. We demonstrate that co-expression can predict gene function, both retrospectively as in the case of capsule and ergosterol biosynthesis, and prospectively, as in the case of DNA damage response proteins. Through co-expression, we were able to identify functions for 11 previously uncharacterized genes, including 6 that are specific to the Tremellales family. Overall, this demonstrates the utility of co-expression for predicting gene function in *C*. *neoformans*.

Co-expression network generation has been performed using multiple methods. We chose to use spearman rank correlation because it is relatively robust and interpretable, due its simplicity, and is among the top-performing co-expression methods in a recent benchmark [97]. We normalized expression scores based on Counts adjustment with TMM Factors (CTF) and Counts adjustment with Upper quartile Factors (CUF) normalizations. As an alternate to our approach, there is SNAIL, a method based on smooth quantile normalization aimed at reducing spurious associations [97]. Future refinements of the CryptoCEN network could use these alternate methods.

Importantly, CryptoCEN complements the current gene function annotation pipelines for *C*. *neoformans* by adding information not already captured from these databases. CryptoNet is based on a Bayesian integration of large-scale genomic and proteomic datasets, and this approach was effective at identifying genes involved in core virulence and drug resistance phenotypes [20]. Combining CryptoCEN with CryptoNET increased the retrospective predictive accuracy, and for a specific capsule network, we observed complementary connections.

A persistent limitation of CryptoCEN, however, is that without initial annotations, we cannot propagate information across the network. In *C*. *neoformans*, due to the high number of genes without annotation, there are many instances in which there is no signature or annotation in the entire network. This is especially the case for *Cryptococcus* or Basidiomycete-specific genes. For example, CNAG_00080 is a hypothetical protein that is *Cryptococcus*-specific. However, this gene is lowly expressed, and the top co-expression partner has a score of only 0.78, which already limits our ability to identify partners. Of the top 50 co-expressed partners, only 5 have any annotation at all, and the others are all hypothetical or unspecified product. As another example, CNAG_00465 is highly expressed basidiomycete-specific gene with many co-expressed partners; however, all the partners are only annotated as hypothetical proteins and there is no GO term enrichment. In these cases, the co-expression network has no

information that can be used to predict gene function for these unknown proteins. In the future, functional genomic screens of mutant libraries will be critical for building the baseline information needed to predict gene function. For this reason, we focused on extending our information about known processes in *C. neoformans*; the DNA damage and response pathway provided a clear opportunity to identify novel proteins involved in a core biological process. Moreover, DNA damage and response processes in pathogens are thought to be potential targets of therapeutic potential [21]. Future work will include mechanistic studies to determine how these specific proteins contribute to EMS resistance or hypersensitivity.

Methods

Collating *C. neoformans* RNAseq studies

Publicly available RNAseq studies were downloaded from SRA based on searches for *Cryptococcus neoformans* and filtered based on H99 or KN99 strains and mutant derivatives. For each study we collected the study accession, taxon, number of runs, study title and depositor. Using this information, where possible, we linked each study to a published journal article, and collected the experimental design and culture conditions for each strain.

Mapping RNAseq reads

RNA expression was estimated by aligning reads to *Cryptococcus neoformans* H99 coding transcripts using release 49 from FungiDB downloaded on 8/05/2021 from https://fungidb.org/common/downloads/release-49/CneoformansH99/fasta/data/FungiDB-49_CneoformansH99_AnnotatedTranscripts.fasta. This release contains 9,189 ORFs defined by distinct unique cnag_id identifiers. Of these, 4 are labeled as pseudo genes and 838 are labeled as alternative splicing variants. Uncharacterized proteins were defined as non-pseudo gene, and the description does not contain ï¿½hypotheticalï¿½ protein ï¿½ unspecified product ï¿½. SRA files were downloaded using prefetch and converted to FASTQ format using fastq-dump from the NCBI SRA-ToolKit package and then aligned using the RSEM package v1.2.31 [28] with bowtie2 [29] using the default settings. To assess sample quality, we measured the percent of genes with non-zero expression and number of reads that map uniquely to the C. *neoformans* genome. Of the 1,524 runs, the average percent genes with non-zero expression was 89%, with a minimum of 61%. The mean number of reads that mapped uniquely to the *C. neoformans* genome was 7.1M, and the minimum was 62k (S4 Fig). For each study, the co-expression rank was estimated by the Spearman rank correlation coefficients of the FPKM values across all runs using in the study the EGAD R package (24) and then averaged across all studies to give the final network.

Complementary networks

To build the sequence homology-based BlastP network, we used Protein-Protein BLAST 2.12.0+(BlastP) to compute sequence similarity between all C. *neoformans* open reading frames. We then ranked the bit-scores, and scores having identical values were given the average rank, then converted into a network using the build_weighted_network() function from the EGAD package.

Data for S. *cerevisiae* S288C was gathered from the Saccharomyces Genome Database (SGD) including the reference genome, all translated open reading frames, and annotations to the slim subset of the gene ontology using release 64-3-1 from 4/21/2021. S. *cerevisiae* genetic and physical interactor networks were built from data collected from BioGRID release 4.4.216, filtering for interactions with experimental system type of `genetic' and `physical' respectively,

mapping to *C. neoformans* open reading frames by BlastP orthology, and building into networks using the build_binary_network() from EGAD package. The sparse binary networks were then extended defining edge weights as the inverse path length using the extend_network () function from the EGAD package.

We downloaded the CryptoNetV1 from http://www.inetbio.org/cryptonet/ across all 14 data type specific networks, which contains 156,506 edges across 5,649 genes.

Embedding of the Co-Exp network

Using the R monocle3 package, we pre-processed the gene by study co-expression matrix by PCA and then used UMAP to non-linearly reduce to 2 dimensions with parameters a = 30, b = 0.8, and default parameters otherwise. To cluster, we used Louvain clustering with parameters k = 30, num_iter = 15, resolution = 0.001. We then plotted the embedding coloring by cluster using ggplot2. To assign functional annotations to each cluster we used GO term enrichment through FungiDB, focusing on the biological process information.

Retrospective function AUROC calculations

13,920 Functional annotations for C. *neoformans* were downloaded from FungiDB release 49 and mapped to GO ontology terms using the GO.db R package gathered on 8/19/2021. Annotations with the NOT qualifier were excluded, and the remaining annotations were propagated along `isa' and `part of' relationships in the GO ontology, yielding 23,863 annotations. Then, to facilitate guilt-by-association gene function prediction, terms with more than 1000 or less than 20 annotations were excluded, yielding 14,215 annotations across 3,421 open reading frames for 145 terms with an average of 4.6 annotations per open reading frame and 98 annotations per go_id.

Co-expression of *S. cerevisiae* Complexes analysis

616 *S. cerevisiae* complexes were downloaded from EBI Complex Portal (v2021-10-13) [98]. Members with identifiers that began with `URS-' (RNA from rnacentral.org id), `EBI-' (mRNA), `CHEBI-' (small molecule), `CPX-' (other complex) or `P12294' (mitochondrially encoded), yielding 1,948 proteins from S. *cerevisiae* strain S288C (vR64-2-1) that were mapped to genes in the Saccharomyces Genome Database [99] using UniProtKB [100]. Of these genes, 1,093 mapped 1-to-1 to *C. neoformans* genes via uniport accession to gene identifier and OrthoMCL (v6.7) [101] based orthology with an additional 211 genes via BlastP based orthology, yielding 1,304 *C. neoformans* genes predicted to participate in 558 complexes (S3 Table). Among the *C. neoformans* orthologs in these complexes, there are 13,950 distinct co-complex associations. 51% C. *neoformans* co-complex associations are within 17 complexes with ï¿½ ribosomalï¿½ï¿½ ribonucleoproteinï¿½ï¿½ their name.

Orthology with S. *cerevisiae*

Pairwise sequence-based orthology using BlastP was defined by reciprocal best hits and having E-values less than 1e-5 in both directions yielding 2,248 associations. Ortholog based on sequence-based clustering using OrthoMCL release 6.7 was defined by orthogroups containing exactly one member from *C. neoformans* and one from S. *cerevisiae*, yielding 2,274 associations.

Co-evolution coefficient determination

To identify gene pairs with signatures of co-evolution, which refers to the covariation of the relative evolutionary rates in two genes across speciation events [102], we used the Covarying Evolutionary Rates (CovER) function in PhyKIT, v1.11.10 [103]. This analysis requires phylogenetic trees of single-copy orthologs from a panel of species. To generate this data, 15 genomes from *Cryptococcus* and the sister genus *Kwoniella* were obtained from NCBI, which spans all publicly available annotated genomes at the time of downloading (07/2022). Orthology was inferred using OrthoFinder, v2.3.8 [104], an algorithm that conducts graph-based clustering of global sequence similarity values calculated using DIAMOND, v2.0.13.151 [105]. Orthology was inferred among protein sequences using an inflation value of 1.8 resulting in 3,828 single-copy orthologs.

To obtain additional groups of orthologous genes for coevolutionary coefficient determination, the OrthoSNAP pipeline, which identifies single-copy orthologs nested within larger gene families [106], was used. To do so, phylogenetic trees were inferred from protein sequences of multi-copy orthogroups with at least eight taxa using IQ-TREE, v2.0.6 [107], with 1,000 ultrafast bootstrap approximations [108]; multi-copy orthogroups were first aligned using MAFFT, v7.402 [109], with the auto parameter, and trimmed using ClipKIT, v1.3.0 [110], with default parameters. The resulting phylogenies were used as input into OrthoSNAP, v1.0.0, which resulted in 1,630 additional single-copy orthologs (5,458 total single-copy orthologs).

The resulting single-copy orthologs were concatenated into a supermatrix using the ï¿½create_concatï¿¼nction in PhyKIT, v1.11.10 [103] and used for species tree estimation using IQ-TREE 2 (best fitting substitution model: JTT+F+I+G4). For each single-copy ortholog, branch lengths were inferred along the species tree using IQ-TREE 2. For every pairwise combination, the coevolutionary coefficient was calculated using the "coverï¿¼nction in PhyKIT, v1.11.10 [103]. In brief, PhyKIT identifies pairs of coevolving genes by first accounting for confounding variables like time since divergence and mutation rate by correcting single-copy ortholog branch lengths with the corresponding branch length in the species tree; the resulting values are Z-transformed and used for Pearson correlation analysis, representing a quantitative measure of gene-gene coevolution.

Paralog identification

Orthology information from the previous section was used to determine pairs of paralogs. 6,641 ortholog groups were mapped over 6,975 C. *neoformans* genes. Among these groups, we identified 1,056 paralog associations across 550 genes.

Strain construction

The following *C*. *neoformans* strains were generated by mating crosses which consisted of co-culturing strains of opposite mating type on MS medium for 7-days [111]. Recombinant progeny were isolated from the mating mixture by random spore dissection and analyzed for genotype, phenotype, and fluorescence. A mating cross between MATa *cdc42*Δ::*nat* (ERB010) and a MATα wild type strain expressing GFP-tagged alpha tubulin (*TUB1a*; CBN242) generated *cdc42*Δ::*nat* + *GFP-TUB1a* (CBN587). Similarly, a mating cross between MATa *cdc420*Δ::*neo* (ERB007) and CBN242 generated *cdc420*Δ::*neo* + *GFP-TUB1a* (CBN589). Mating crosses of strains ERB010 and ERB007 with a MATα wild type strain expressing a GFP-tagged Histone H4 (CNV108) generated *cdc42*Δ::*nat* + *GFP-H4* (CBN594) and *cdc420*Δ::*neo* + *GFP-H4* (CBN593).

Nocodazole sensitivity assay

To test sensitivity, each strain was 2-fold serially diluted and 5 μL spots of each dilution were plated onto YPD with or without the indicated concentration of nocodazole. Plates were incubated for 3 days before imaging.

Histone / tubulin localization assay

The following strains were incubated to mid-logarithmic phase with shaking (200 rpm) in YPD medium at either 30ï¿½C 37ï¿½C with or without the addition of 0.125 μM nocodazole: strains CNV108 (WT + GFP-Histone H4), CBN242 (WT + GFP- Tub1), CBN593(*cdc420*Δ + GFP-Histone H4), and CBN589 (*cdc420*Δ + GFP- Tub1). Nuclear size and division (noted by GFP-H4) and tubulin filaments (GFP-Tub1) were assessed using a Zeiss Axio Imager A1 microscope equipped with an Axio-Cam MRmdigital camera. Cells were imaged by DIC and with eGFP filter. Identical exposure times were used to image all cells. Fiji software [112] was used to process images.

Capsule assay

Capsule assays were performed as previously described [54]. Briefly, cells were inoculated into DMEM (Gibco 11995) in 12-well plates and incubated at 37ï¿½C with 5% CO_2 for 3 days before staining with India ink (Fisher Scientific) and imaging on a Lionheart FX using brightfield microscopy. Images are representative of three biological replicates.

Fluconazole sensitivity assay

To test sensitivity, each strain was serially diluted and 5 μL spots of each dilution were plated onto YPD with or without the indicated concentration of fluconazole. Plates were incubated for 3 days before imaging.

DNA damage response assay

EMS and UV mutagenesis assays were based on the protocol described by Winston [113]. Overnight cultures of *C. neoformans* were normalized to an OD_{600} of 1. Cultures were then centrifuged in 1 mL aliquots and washed and resuspended in 0.1 M sodium phosphate buffer (pH 7.4) to remove excess YPD. For EMS mutagenesis, cells were then transferred to 15 mL conical tubes where either 5% ethyl methanesulfonate or sodium phosphate buffer was added. Dose of EMS was chosen after dose-response assays (S5 Fig). Final volume in each tube was adjusted to 2 mL and cells were incubated at 30ï¿½C for 1 hour while shaking. After incubation, all samples were pelleted and washed with a 5% sodium thiosulfate solution to inactivate the EMS. Following this wash, cells pellets were resuspended in 1 mL of sterile water and diluted 10-fold. 5 μL of each dilution were spotted onto YPD plates and incubated for 48 hours before imaging. For UV mutagenesis, mock-treated samples were spotted onto YPD, but exposed to 200 μJ of UV radiation using a UV-Stratalinker before incubation at 30ï¿½C for 2 days before imaging on a BioRad gel dock.

Supporting information

S1 Fig. Retrospective predictive accuracy of CryptoCEN. A. A re-embedded gene by expression matrix, scanning the key embedding parameters umap_a and umap_b over the ranges [20, 30, 40, 50, 60] and [0.45, 0.5, 0.55], respectively, while keeping the remaining parameters fixed (prereduction of dimension to 500 using PCA, n_neighbors = 30, negative_sample_rate = 50, umap_repulsion_strength = 3, n_epochs = 2000). Rows are umap_a and the columns

are umap_b. The points are clustered using leiden clustering using with a resolution parameter of 1e-3 and points are colored by the cluster index. B. UpSet plot of the retrospective prediction accuracy, as determined by the neighbor voting guilt-by-association (GBA) area under the ROC curve (AUROC). AUROCs values range between 0.5 for random predictor and 1 for a perfect predictor. As data sources are combined, the prediction accuracy increases. Each annotated GO term is colored by ontology biological process (BP), cellular component (CC), or molecular function (MF). C. Enrichment of co-expression in the gene by expression matrix UMAP. For each cluster, we selected inter and intra-cluster associations. We then computed the area under the receiver operator characteristic (AUROC) for the enrichment of the intra-cluster associations over the inter-cluster associations based on the co-expression score. Enrichment within each cluster is indicated by color.
(TIF)

S2 Fig. Evolutionary constraints on co-expression. A) Phylogeny used for co-evolution. Tips indicate each species, scale bar indicates 0.05 substitutions/site. B) Co-expression vs. co-evolution scores over the 5,264 overlapping gene sets does not show a positive correlation (correlation coefficient is -0.001). C) There is no difference in histone localization between the strains. Histones were marked by fusion of H4 with GFP. Cells were incubated in liquid YPD at 30ï¿½C before imaging. Images taken at 40X magnification, scale = 5 microns. D) There is no difference in nocodazole sensitivity between the *cdc42*Δ or *cdc420*Δ mutant strains. The indicated strains were grown overnight at 30ï¿½C in liquid YPD medium and then serially diluted onto YPD or plates containing 0.3 μM nocodazole and incubated for 2 days before imaging.
(TIF)

S3 Fig. Matched controls do not show enrichment for DNA repair. A) Analysis of matched controls for phenotypes on DNA damaging agents. The indicated strains were grown overnight at 30ï¿½C in liquid YPD medium, and the 10-fold serially diluted cells were spotted onto YPD agar. For UV damage, the plates were immediately subjected to 200 μJ UV. For EMS, the cells were incubated in 5% EMS for 1 hr before serial dilution and plating. The plates were incubated at 30ï¿½C and imaged after 2 days.
(TIF)

S4 Fig. Mapping ***C. neoformans*** RNAseq runs. A. A total of 1,523 RNAseq runs from 34 identified studies are scatter-plotted as the number of genes with nonzero expression versus the fraction transcripts that map exactly once. B. The RNAseq runs that have nonzero expression for at least half of the genes which are used to construct the CryptoCEN network.
(TIF)

S5 Fig. Establishing concentrations for EMS treatment. A) H99 wild type cells were incubated with the indicated concentrations of EMS for 1 hr before serial dilution and plating onto YPD. The plates were incubated at 30ï¿½C and imaged after 2 days. CFU/mL was calculated from the serial dilutions.
(TIF)

S1 Table. Studies included in CryptoCEN. Tab 1) RNAseq studies. Tab 2) RNASeq runs.
(XLSX)

S2 Table. Co-Expression scores.
(XLSX)

S3 Table. cluster analysis.
(XLSX)

S4 Table. Saccharomyces complexes. Tab 1) summary. Tab 2) pairwise.
(XLSX)

S5 Table. orthogroups.
(XLSX)

S6 Table. case studies. Tab 1) capsule co-expression values. Tab 2) ergosterol co-expression values.
(XLSX)

S7 Table. DNA damage. Tab 1) co-expression. Tab 2) phenotype summary. Tab 3) reciprocal co-expression.
(XLSX)

Acknowledgments

We thank Vikas Yadav for thoughtful comments on kinetochores.

Author Contributions

Conceptualization: Matthew J. O'Meara, Connie B. Nichols, R. Blake Billmyre, Jacob L Steenwyk, J. Andrew Alspaugh, Teresa R. O'Meara.

Data curation: Matthew J. O'Meara, Jackson R. Rapala, Connie B. Nichols, A. Christina Alexandre, Jacob L Steenwyk, Teresa R. O'Meara.

Formal analysis: Matthew J. O'Meara, Jackson R. Rapala, R. Blake Billmyre, Teresa R. O'Meara.

Funding acquisition: J. Andrew Alspaugh, Teresa R. O'Meara.

Investigation: Matthew J. O'Meara, Jackson R. Rapala, Connie B. Nichols, A. Christina Alexandre, Jacob L Steenwyk, Teresa R. O'Meara.

Methodology: Matthew J. O'Meara, Jackson R. Rapala, Connie B. Nichols, Jacob L Steenwyk, Teresa R. O'Meara.

Project administration: Matthew J. O'Meara, J. Andrew Alspaugh, Teresa R. O'Meara.

Resources: Matthew J. O'Meara, Teresa R. O'Meara.

Software: Matthew J. O'Meara.

Supervision: Matthew J. O'Meara, Teresa R. O'Meara.

Validation: Teresa R. O'Meara.

Visualization: Matthew J. O'Meara, Jackson R. Rapala, Connie B. Nichols, Teresa R. O'Meara.

Writing ï¿½ original draft: Matthew J. O'Meara, Jackson R. Rapala, Connie B. Nichols, Jacob L Steenwyk, J. Andrew Alspaugh, Teresa R. O'Meara.

Writing ï¿½ review & editing: Matthew J. O'Meara, Jackson R. Rapala, Connie B. Nichols, A. Christina Alexandre, R. Blake Billmyre, Jacob L Steenwyk, J. Andrew Alspaugh, Teresa R. O'Meara.

References

1. Cozzetto D, Jones DT. Computational Methods for Annotation Transfers from Sequence. Methods Mol Biol. 2017; 1446: 55ï¿½67. https://doi.org/10.1007/978-1-4939-3743-1_5 PMID: 27812935
2. Törönen P, Holm L. PANNZER-A practical tool for protein function prediction. Protein Sci. 2022; 31: 118ï¿½132. https://doi.org/10.1002/pro.4193 PMID: 34562305
3. Gligorijevic V, Renfrew PD, Kosciolek T, Leman JK, Berenberg D, Vatanen T, et al. Structure-based protein function prediction using graph convolutional networks. bioRxiv. bioRxiv; 2019. p. 786236. https://doi.org/10.1101/786236
4. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019; 47: D607ï¿½613. https://doi.org/10.1093/nar/gky1131 PMID: 30476243
5. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A. 1998; 95: 14863ï¿½14868. https://doi.org/10.1073/pnas.95.25.14863 PMID: 9843981
6. Ballouz S, Verleyen W, Gillis J. Guidance for RNA-seq co-expression network construction and analysis: safety in numbers. Bioinformatics. 2015; 31: 2123ï¿½2130. https://doi.org/10.1093/bioinformatics/btv118 PMID: 25717192
7. O'Meara TR, O'Meara MJ. DeORFanizing Candida albicans Genes using Coexpression. mSphere. 2021; 6. https://doi.org/10.1128/mSphere.01245-20 PMID: 33472984
8. Wisecaver JH, Borowsky AT, Tzin V, Jander G, Kliebenstein DJ, Rokas A. A global coexpression network approach for connecting genes to specialized metabolic pathways in plants. Plant Cell. 2017; 29: 944ï¿½965. https://doi.org/10.1105/tpc.17.00009 PMID: 28408660
9. Zhou N, Jiang Y, Bergquist TR, Lee AJ, Kacsoh BZ, Crocker AW, et al. The CAFA challenge reports improved protein function prediction and new functional annotations for hundreds of genes through experimental screens. Genome Biol. 2019; 20: 244. https://doi.org/10.1186/s13059-019-1835-8 PMID: 31744546
10. Fisher MC, Denning DW. The WHO fungal priority pathogens list as a game-changer. Nat Rev Microbiol. 2023; 21: 211ï¿½212. https://doi.org/10.1038/s41579-023-00861-x PMID: 36747091
11. Taylor JW, Berbee ML. Dating divergences in the Fungal Tree of Life: review and new analyses. Mycologia. 2007; 98: 838ï¿½849. https://doi.org/10.3852/mycologia.98.6.838 PMID: 17486961
12. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000; 25: 25ï¿½29. https://doi.org/10.1038/75556 PMID: 10802651
13. Gene Ontology Consortium, Aleksander SA, Balhoff, Carbon S, Cherry JM, Drabkin HJ, et al. The Gene Ontology knowledgebase in 2023. Genetics. 2023; 224. https://doi.org/10.1093/genetics/iyad031 PMID: 36866529
14. D'Souza CA, Alspaugh JA, Yue C, Harashima T, Cox GM, Perfect JR, et al. Cyclic AMP-dependent protein kinase controls virulence of the fungal pathogen Cryptococcus neoformans. Mol Cell Biol. 2001; 21: 3179ï¿½3191. https://doi.org/10.1128/MCB.21.9.3179-3191.2001 PMID: 11287622
15. O'Meara TR, Norton D, Price MS, Hay C, Clements MF, Nichols CB, et al. Interaction of Cryptococcus neoformans Rim101 and protein kinase A regulates capsule. PLoS Pathog. 2010; 6: e1000776. https://doi.org/10.1371/journal.ppat.1000776 PMID: 20174553
16. Kelliher CM, Haase SB. Connecting virulence pathways to cell-cycle progression in the fungal pathogen Cryptococcus neoformans. Curr Genet. 2017; 63: 803ï¿½811. https://doi.org/10.1007/s00294-017-0688-5 PMID: 28265742
17. Yu C-H, Sephton-Clark P, Tenor JL, Toffaletti DL, Giamberardino C, Haverkamp M, et al. Gene expression of diverse Cryptococcus isolates during infection of the human central nervous system. MBio. 2021; 12: e0231321. https://doi.org/10.1128/mBio.02313-21 PMID: 34724829
18. Maier EJ, Haynes BC, Gish SR, Wang ZA, Skowyra ML, Marulli AL, et al. Model-driven mapping of transcriptional networks reveals the circuitry and dynamics of virulence regulation. Genome Res. 2015; 25: 690ï¿½700. https://doi.org/10.1101/gr.184101.114 PMID: 25644834
19. Brown HE, Telzrow CL, Saelens JW, Fernandes L, Alspaugh JA. Sterol-response pathways mediate alkaline survival in diverse fungi. MBio. 2020; 11. https://doi.org/10.1128/mBio.00719-20 PMID: 32546619
20. Kim H, Jung K-W, Maeng S, Chen Y-L, Shin J, Shim JE, et al. Network-assisted genetic dissection of pathogenicity and drug resistance in the opportunistic human pathogenic fungus Cryptococcus neoformans. Sci Rep. 2015; 5: 8767. https://doi.org/10.1038/srep08767 PMID: 25739925

21. Shor E, Garcia-Rubio R, DeGregorio L, Perlin DS. A noncanonical DNA damage checkpoint response in a major fungal pathogen. MBio. 2020; 11. https://doi.org/10.1128/mbio.03044-20 PMID: 33323516

22. Lee D, Jang E-H, Lee M, Kim S-W, Lee Y, Lee K-T, et al. Unraveling Melanin Biosynthesis and Signaling Networks in Cryptococcus neoformans. MBio. 2019; 10. https://doi.org/10.1128/mBio.02267-19 PMID: 31575776

23. Summers DK, Perry DS, Rao B, Madhani HD. Coordinate genomic association of transcription factors controlled by an imported quorum sensing peptide in Cryptococcus neoformans. PLoS Genet. 2020; 16: e1008744. https://doi.org/10.1371/journal.pgen.1008744 PMID: 32956370

24. Do E, Cho Y-J, Kim D, Kronstad JW, Jung WH. A transcriptional regulatory map of iron homeostasis reveals a new control circuit for capsule formation in Cryptococcus neoformans. Genetics. 2020; 215: 1171ï¿½1189. https://doi.org/10.1534/genetics.120.303270 PMID: 32580959

25. Janbon G, Ormerod KL, Paulet D, Byrnes EJ 3rd, Yadav V, Chatterjee G, et al. Analysis of the genome and transcriptome of Cryptococcus neoformans var. grubii reveals complex RNA expression and microevolution leading to virulence attenuation. PLoS Genet. 2014; 10: e1004261. https://doi.org/10.1371/journal.pgen.1004261 PMID: 24743168

26. Wallace EWJ, Maufrais C, Sales-Lee J, Tuck LR, de Oliveira L, Feuerbach F, et al. Quantitative global studies reveal differential translational control by start codon context across the fungal kingdom. Nucleic Acids Res. 2020; 48: 2312ï¿½2333. https://doi.org/10.1093/nar/gkaa060 PMID: 32020195

27. Telzrow CL, Zwack PJ, Esher Righi S, Dietrich FS, Chan C, Owzar K, et al. Comparative analysis of RNA enrichment methods for preparation of Cryptococcus neoformans RNA sequencing libraries. G3 (Bethesda). 2021; 11. https://doi.org/10.1093/g3journal/jkab301 PMID: 34518880

28. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011; 12: 323. https://doi.org/10.1186/1471-2105-12-323 PMID: 21816040

29. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012; 9: 357ï¿½359. https://doi.org/10.1038/nmeth.1923 PMID: 22388286

30. Ballouz S, Weber M, Pavlidis P, Gillis J. EGAD: ultra-fast functional analysis of gene networks. Bioinformatics. 2016; 33: 612ï¿½614. 1

31. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv [stat.ML]. 2018. http://arxiv.org/abs/1802.03426

32. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J. Stat. Mech. (2008) P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

33. Basenko EY, Pulman JA, Shanmugasundram A, Harb OS, Crouch K, Starns D, et al. FungiDB: An Integrated Bioinformatic Resource for Fungi and Oomycetes. J Fungi (Basel). 2018; 4. https://doi.org/10.3390/jof4010039 PMID: 30152809

34. Gillis J, Pavlidis P. The impact of multifunctional genes on ïï¿½ guilt-by-association ïï¿½ analysis. PLoS One. 2011; 6: e17258. https://doi.org/10.1371/journal.pone.0017258 PMID: 21364756

35. Chen Y, Müller JD. Determining the stoichiometry of protein heterocomplexes in living cells with fluorescence fluctuation spectroscopy. Proc Natl Acad Sci U S A. 2007; 104: 3147ï¿½3152. https://doi.org/10.1073/pnas.0606557104 PMID: 17307882

36. Brennan CM, Vaites LP, Wells JN, Santaguida S, Paulo JA, Storchova Z, et al. Protein aggregation mediates stoichiometry of protein complexes in aneuploid cells. Genes Dev. 2019; 33: 1031ï¿½1047. https://doi.org/10.1101/gad.327494.119 PMID: 31196865

37. Li L, Stoeckert CJ Jr, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 2003; 13: 2178ï¿½2189. https://doi.org/10.1101/gr.1224503 PMID: 12952885

38. Steenwyk JL, Phillips MA, Yang F, Date SS, Graham TR, Berman J, et al. An orthologous gene coevolution network provides insight into eukaryotic cellular and genomic structure and function. Sci Adv. 2022; 8: eabn0105. https://doi.org/10.1126/sciadv.abn0105 PMID: 35507651

39. Hickman MA, Rusche LN. Transcriptional silencing functions of the yeast protein Orc1/Sir3 subfunctionalized after gene duplication. Proc Natl Acad Sci U S A. 2010; 107: 19384ï¿½19389. https://doi.org/10.1073/pnas.1006436107 PMID: 20974972

40. Malik Ghulam M, Catala M, Reulet G, Scott MS, Abou Elela S. Duplicated ribosomal protein paralogs promote alternative translation and drug resistance. Nat Commun. 2022; 13: 4938. https://doi.org/10.1038/s41467-022-32717-y PMID: 35999447

41. Coate JE, Farmer AD, Schiefelbein JW, Doyle JJ. Expression partitioning of duplicate genes at single cell resolution in Arabidopsis roots. Front Genet. 2020; 11: 596150. https://doi.org/10.3389/fgene.2020.596150 PMID: 33240334

42. Loehlin DW, Carroll SB. Expression of tandem gene duplicates is often greater than twofold. Proc Natl Acad Sci U S A. 2016; 113: 5988ï¿½5992 https://doi.org/10.1073/pnas.1605886113 PMID: 27162370

43. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. Preservation of duplicate genes by complementary, degenerative mutations. Genetics. 1999; 151: 1531ï¿½1545 https://doi.org/10.1093/genetics/151.4.1531 PMID: 10101175

44. Ballou ER, Selvig K, Narloch JL, Nichols CB, Alspaugh JA. Two Rac paralogs regulate polarized growth in the human fungal pathogen Cryptococcus neoformans. Fungal Genet Biol. 2013; 57: 58ï¿½75. https://doi.org/10.1016/j.fgb.2013.05.006 PMID: 23748012

45. Ballou ER, Kozubowski L, Nichols CB, Alspaugh JA. Ras1 acts through duplicated Cdc42 and Rac proteins to regulate morphogenesis and pathogenesis in the human fungal pathogen Cryptococcus neoformans. PLoS Genet. 2013; 9: e1003687. https://doi.org/10.1371/journal.pgen.1003687 PMID: 23950731

46. Waugh MS, Nichols CB, DeCesare CM, Cox GM, Heitman J, Alspaugh JA. Ras1 and Ras2 contribute shared and unique roles in physiology and virulence of Cryptococcus neoformans. Microbiology. 2002; 148: 191ï¿½201 https://doi.org/10.1099/00221287-148-1-191 PMID: 11782511

47. Alspaugh JA, Cavallo LM, Perfect JR, Heitman J. RAS1 regulates filamentation, mating and growth at high temperature of Cryptococcus neoformans. Mol Microbiol. 2000; 36: 352ï¿½365 https://doi.org/10.1046/j.1365-2958.2000.01852.x PMID: 10792722

48. Ballou ER, Nichols CB, Miglia KJ, Kozubowski L, Alspaugh JA. Two CDC42 paralogues modulate Cryptococcus neoformans thermotolerance and morphogenesis under host physiological conditions. Mol Microbiol. 2010; 75: 763ï¿½780 https://doi.org/10.1111/j.1365-2958.2009.07019.x PMID: 20025659

49. Kozubowski L, Yadav V, Chatterjee G, Sridhar S, Yamaguchi M, Kawamoto S, et al. Ordered kinetochore assembly in the human-pathogenic basidiomycetous yeast Cryptococcus neoformans. MBio. 2013; 4: e00614ï¿½13 https://doi.org/10.1128/mBio.00614-13 PMID: 24085781

50. Yadav V, Sanyal K. Sad1 spatiotemporally regulates kinetochore clustering to ensure high-fidelity chromosome segregation in the human fungal pathogen Cryptococcus neoformans. mSphere. 2018;3. https://doi.org/10.1128/mSphere.00190-18 PMID: 29976642

51. Varshney N, Som S, Chatterjee S, Sridhar S, Bhattacharyya D, Paul R, et al. Spatio-temporal regulation of nuclear division by Aurora B kinase Ipl1 in Cryptococcus neoformans. PLoS Genet. 2019; 15: e1007959. https://doi.org/10.1371/journal.pgen.1007959 PMID: 30763303

52. Leontiou I, Davies T, Clark I, Aktar K, Suresh AP, Abad MA, et al. Bub1 kinase acts as a signalling hub for the entire Cryptococcus neoformans spindle assembly checkpoint pathway. bioRxiv. 2022. https://doi.org/10.1101/2022.09.21.508923

53. Sridhar S, Hori T, Nakagawa R, Fukagawa T, Sanyal K. Bridgin connects the outer kinetochore to centromeric chromatin. Nat Commun. 2021; 12: 146. https://doi.org/10.1038/s41467-020-20161-9 PMID: 33420015

54. O'Meara TR, Alspaugh JA. The Cryptococcus neoformans capsule: a sword and a shield. Clin Microbiol Rev. 2012; 25: 387ï¿½408 https://doi.org/10.1128/CMR.00001-12 PMID: 22763631

55. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003; 13: 2498ï¿½2504 https://doi.org/10.1101/gr.1239303 PMID: 14597658

56. Garcï¿½a-Rodas R, Cordero RJB, Trevijano-Contador N, Janbon G, Moyrand F, Casadevall A, et al. Capsule growth in Cryptococcus neoformans is coordinated with cell cycle progression. MBio. 2014; 5: e00945ï¿½14 https://doi.org/10.1128/mBio.00945-14 PMID: 24939886

57. Gish SR, Maier EJ, Haynes BC, Santiago-Tirado FH, Srikanta DL, Ma CZ, et al. Computational Analysis Reveals a Key Regulator of Cryptococcal Virulence and Determinant of Host Response. MBio. 2016; 7: e00313ï¿½16 https://doi.org/10.1128/mBio.00313-16 PMID: 27094327

58. Cramer KL, Gerrald QD, Nichols CB, Price MS, Alspaugh JA. Transcription Factor Nrg1 Mediates Capsule Formation, Stress Response, and Pathogenesis in Cryptococcus neoformans. Eukaryot Cell. 2006; 5: 1147ï¿½1156 https://doi.org/10.1128/EC.00145-06 PMID: 16835458

59. Liu OW, Chun CD, Chow ED, Chen C, Madhani HD, Noble SM. Systematic genetic analysis of virulence in the human fungal pathogen Cryptococcus neoformans. Cell. 2008; 135: 174ï¿½188 https://doi.org/10.1016/j.cell.2008.07.046 PMID: 18854164

60. Revie NM, Iyer KR, Robbins N, Cowen LE. Antifungal drug resistance: evolution, mechanisms and impact. Curr Opin Microbiol. 2018; 45: 70ï¿½76 https://doi.org/10.1016/j.mib.2018.02.005 PMID: 29547801

61. Billmyre RB, Clancey SA, Heitman J. Natural mismatch repair mutations mediate phenotypic diversity and drug resistance in Cryptococcus deuterogattii. Elife. 2017; 6. https://doi.org/10.7554/eLife.28802 PMID: 28948913

62. Boyce KJ, Wang Y, Verma S, Shakya VPS, Xue C, Idnurm A. Mismatch repair of DNA replication errors contributes to microevolution in the pathogenic fungus Cryptococcus neoformans. MBio. 2017; 8. https://doi.org/10.1128/mBio.00595-17 PMID: 28559486

63. Albehaijani SHI, Macreadie I, Morrissey CO, Boyce KJ. Molecular mechanisms underlying the emergence of polygenetic antifungal drug resistance in msh2 mismatch repair mutants of Cryptococcus. JAC Antimicrob Resist. 2022; 4: dlac033. https://doi.org/10.1093/jacamr/dlac033 PMID: 35402912

64. Steenwyk JL, Opulente DA, Kominek J, Shen X-X, Zhou X, Labella AL, et al. Extensive loss of cell-cycle and DNA repair genes in an ancient lineage of bipolar budding yeasts. PLoS Biol. 2019; 17: e3000255. https://doi.org/10.1371/journal.pbio.3000255 PMID: 31112549

65. Salas-Leiva DE, Tromer EC, Curtis BA, Jerlström-Hultqvist J, Kolisko M, Yi Z, et al. Genomic analysis finds no evidence of canonical eukaryotic DNA processing complexes in a free-living protist. Nat Commun. 2021; 12: 6003. https://doi.org/10.1038/s41467-021-26077-2 PMID: 34650064

66. Ashton PM, Thanh LT, Trieu PH, Van Anh D, Trinh NM, Beardsley J, et al. Three phylogenetic groups have driven the recent population expansion of Cryptococcus neoformans. Nat Commun. 2019; 10: 2035. https://doi.org/10.1038/s41467-019-10092-5 PMID: 31048698

67. Jung K-W, Yang D-H, Kim M-K, Seo HS, Lim S, Bahn Y-S. Unraveling Fungal Radiation Resistance Regulatory Networks through the Genome-Wide Transcriptome and Genetic Analyses of Cryptococcus neoformans. MBio. 2016; 7. https://doi.org/10.1128/mBio.01483-16 PMID: 27899501

68. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021. https://doi.org/10.1038/s41586-021-03819-2 PMID: 34265844

69. van Kempen M, Kim S, Tumescheit C, Mirdita M, Gilchrist CLM, Soeding J, et al. Foldseek: fast and accurate protein structure search. bioRxiv. 2022. https://doi.org/10.1101/2022.02.07.479398

70. Knudsen M, Wiuf C. The CATH database. Hum Genomics. 2010; 4: 207–212. https://doi.org/10.1186/1479-7364-4-3-207 PMID: 20368142

71. Zou L, Elledge SJ. Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. Science. 2003; 300: 1542–1548. https://doi.org/10.1126/science.1083430 PMID: 12791985

72. Cortez D, Guntuku S, Qin J, Elledge SJ. ATR and ATRIP: partners in checkpoint signaling. Science. 2001; 294: 1713–1716. https://doi.org/10.1126/science.1065521 PMID: 11721054

73. Li Z, Kim KS. RELATe enables genome-scale engineering in fungal genomics. Sci Adv. 2020; 6. https://doi.org/10.1126/sciadv.abb8783 PMID: 32948588

74. Fu C, Heitman J. PRM1 and KAR5 function in cell-cell fusion and karyogamy to drive distinct bisexual and unisexual cycles in the Cryptococcus pathogenic species complex. PLoS Genet. 2017; 13: e1007113. https://doi.org/10.1371/journal.pgen.1007113 PMID: 29176784

75. Chan Lee Soo, Joseph Heitman. Function of Cryptococcus neoformans KAR7 (SEC66) in Karyogamy during Unisexual and Opposite-Sex Mating. Eukaryot Cell. 2012; 11: 783–794. https://doi.org/10.1128/EC.00066-12 PMID: 22544906

76. Li Y, Liu H, Steenwyk JL, LaBella AL, Harrison M-C, Groenewald M, et al. Contrasting modes of macro and microsynteny evolution in a eukaryotic subphylum. Curr Biol. 2022; 32: 5335–5343.e4. https://doi.org/10.1016/j.cub.2022.10.025 PMID: 36334587

77. Weisman CM, Murray AW, Eddy SR. Many, but not all, lineage-specific genes can be explained by homology detection failure. PLoS Biol. 2020; 18: e3000862. https://doi.org/10.1371/journal.pbio.3000862 PMID: 33137085

78. Legrand M, Chan CL, Jauert PA, Kirkpatrick DT. Role of DNA mismatch repair and double-strand break repair in genome stability and antifungal drug resistance in Candida albicans. Eukaryot Cell. 2007; 6: 2194–2205. https://doi.org/10.1128/EC.00299-07 PMID: 17965250

79. Nordmann M, Cabrera M, Perz A, Bröcker C, Ostrowicz C, Engelbrecht-Vandré S, et al. The Mon1-Ccz1 complex is the GEF of the late endosomal Rab7 homolog Ypt7. Curr Biol. 2010; 20: 1654–1659. https://doi.org/10.1016/j.cub.2010.08.002 PMID: 20797862

80. Stenmark H. Rab GTPases as coordinators of vesicle traffic. Nat Rev Mol Cell Biol. 2009; 10: 513–525. https://doi.org/10.1038/nrm2728 PMID: 19603039

81. Gao J, Langemeyer L, Kümmel D, Reggiori F, Ungermann C. Molecular mechanism to target the endosomal Mon1-Ccz1 GEF complex to the pre-autophagosomal structure. Elife. 2018;7. https://doi.org/10.7554/eLife.31145 PMID: 29446751

82. Bhargava HK, Tabata K, Byck JM, Hamasaki M, Farrell DP, Anishchenko I, et al. Structural basis for autophagy inhibition by the human Rubicon-Rab7 complex. Proc Natl Acad Sci U S A. 2020; 117: 17003–17010. https://doi.org/10.1073/pnas.2008030117 PMID: 32632011

83. Vanzo R, Bartkova J, Merchut-Maya JM, Hall A, Bouchal J, Dyrskjøt L, et al. Autophagy role(s) in response to oncogenes and DNA replication stress. Cell Death Differ. 2020; 27: 1134ï¿½1153. https://doi.org/10.1038/s41418-019-0403-9 PMID: 31409894

84. Chabes A, Georgieva B, Domkin V, Zhao X, Rothstein R, Thelander L. Survival of DNA damage in yeast directly depends on increased dNTP levels allowed by relaxed feedback inhibition of ribonucleotide reductase. Cell. 2003. pp. 391ï¿½401. https://doi.org/10.1016/s0092-8674(03)00075-8 PMID: 12581528

85. Eapen VV, Waterman DP, Bernard A, Schiffmann N, Sayas E, Kamber R, et al. A pathway of targeted autophagy is induced by DNA damage in budding yeast. Proc Natl Acad Sci U S A. 2017; 114: E1158ï¿½E1167. https://doi.org/10.1073/pnas.1614364114 PMID: 28154131

86. Lee K-T, Hong J, Lee D-G, Lee M, Cha S, Lim Y-G, et al. Fungal kinases and transcription factors regulating brain infection in Cryptococcus neoformans. Nat Commun. 2020; 11: 1521. https://doi.org/10.1038/s41467-020-15329-2 PMID: 32251295

87. Bonilla M, Nastase KK, Cunningham KW. Essential role of calcineurin in response to endoplasmic reticulum stress. EMBO J. 2002; 21: 2343ï¿½2353. https://doi.org/10.1093/emboj/21.10.2343 PMID: 12006487

88. Liu S, Hou Y, Liu W, Lu C, Wang W, Sun S. Components of the calcium-calcineurin signaling pathway in fungal cells and their potential as antifungal targets. Eukaryot Cell. 2015; 14: 324ï¿½334. https://doi.org/10.1128/EC.00271-14 PMID: 25636321

89. Miyakawa T, Mizunuma M. Physiological roles of calcineurin in Saccharomyces cerevisiae with special emphasis on its roles in G2/M cell-cycle regulation. Biosci Biotechnol Biochem. 2007; 71: 633ï¿½645. https://doi.org/10.1271/bbb.60495 PMID: 17341827

90. Dayton JS, Sumi M, Nanthakumar NN, Means AR. Expression of a Constitutively Active Ca2+/Calmodulin-dependent Kinase in Aspergillus nidulans Spores Prevents Germination and Entry into the Cell Cycle*. J Biol Chem. 1997; 272: 3223ï¿½3229. https://doi.org/10.1074/jbc.272.6.3223 PMID: 9013558

91. Planas-Silva MD, Means AR. Expression of a constitutive form of calcium/calmodulin dependent protein kinase II leads to arrest of the cell cycle in G2. EMBO J. 1992; 11: 507ï¿½517. https://doi.org/10.1002/j.1460-2075.1992.tb05081.x PMID: 1371461

92. Li G, Liu S, Wu L, Wang X, Cuan R, Zheng Y, et al. Characterization and Functional Analysis of a New Calcium/Calmodulin-Dependent Protein Kinase (CaMK1) in the Citrus Pathogenic Fungus Penicillium italicum. J Fungi (Basel). 2022; 8. https://doi.org/10.3390/jof8070667 PMID: 35887424

93. Rajasingham R, Govender NP, Jordan A, Loyse A, Shroufi A, Denning DW, et al. The global burden of HIV-associated cryptococcal infection in adults in 2020: a modelling analysis. Lancet Infect Dis. 2022; 22: 1748ï¿½1755. https://doi.org/10.1016/S1473-3099(22)00499-6 PMID: 36049486

94. Wang JM, Woodruff AL, Dunn MJ, Fillinger RJ, Bennett RJ, Anderson MZ. Intraspecies Transcriptional Profiling Reveals Key Regulators of Candida albicans Pathogenic Traits. MBio. 2021; 12. https://doi.org/10.1128/mBio.00586-21 PMID: 33879584

95. Ihmels J, Bergmann S, Berman J, Barkai N. Comparative gene expression analysis by differential clustering approach: application to the Candida albicans transcription program. PLoS Genet. 2005; 1: e39. https://doi.org/10.1371/journal.pgen.0010039 PMID: 16470937

96. Schäpe P, Kwon MJ, Baumann B, Gutschmann B, Jung S, Lenz S, et al. Updating genome annotation for the microbial cell factory Aspergillus niger using gene co-expression networks. Nucleic Acids Res. 2019; 47: 559ï¿½569. https://doi.org/10.1093/nar/gky1183 PMID: 30496528

97. Hsieh P-H, Lopes-Ramos CM, Zucknick M, Sandve GK, Glass K, Kuijjer ML. Adjustment of spurious correlations in co-expression measurements from RNA-Sequencing data. bioRxiv. bioRxiv; 2021. https://doi.org/10.1101/2021.03.25.436972

98. Meldal BHM, Bye-A-Jee H, Gajdoš L, Hammerová Z, Horáčková A, Melicher F, et al. Complex Portal 2018: extended content and enhanced visualization tools for macromolecular complexes. Nucleic Acids Res. 2019; 47: D550ï¿½D558. https://doi.org/10.1093/nar/gky1001 PMID: 30357405

99. Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, et al. Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Res. 2012; 40: D700ï¿½5. https://doi.org/10.1093/nar/gkr1029 PMID: 22110037

100. Consortium UniProt. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. 2021; 49: D480ï¿½D489. https://doi.org/10.1093/nar/gkaa1100 PMID: 33237286

101. Chen F, Mackey AJ, Stoeckert CJ Jr, Roos DS. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. Nucleic Acids Res. 2006; 34: D363ï¿½8. https://doi.org/10.1093/nar/gkj123 PMID: 16381887

102. Clark NL, Alani E, Aquadro CF. Evolutionary rate covariation reveals shared functionality and coexpression of genes. Genome Res. 2012; 22: 714ï¿½720https://doi.org/10.1101/gr.132647.111 PMID: 22287101

103. Steenwyk JL, Buida TJ III, Labella AL, Li Y, Shen X-X, Rokas A. PhyKIT: a broadly applicable UNIX shell toolkit for processing and analyzing phylogenomic data. Bioinformatics. 2021; 37: 2325ï¿½2331. https://doi.org/10.1093/bioinformatics/btab096 PMID: 33560364

104. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20. https://doi.org/10.1186/s13059-019-1832-y PMID: 31727128

105. Buchfink B, Reuter K, Drost H-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods. 2021; 18: 366ï¿½368https://doi.org/10.1038/s41592-021-01101-x PMID: 33828273

106. Steenwyk JL, Goltz DC, Buida TJ, Li Y, Shen X-X, Rokas A. OrthoSNAP: A tree splitting and pruning algorithm for retrieving single-copy orthologs from gene family trees. PLoS Biol. 2022; 20: e3001827. https://doi.org/10.1371/journal.pbio.3001827 PMID: 36228036

107. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020; 37: 1530ï¿½1534https://doi.org/10.1093/molbev/msaa015 PMID: 32011700

108. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the ultrafast bootstrap approximation. Mol Biol Evol. 2018; 35: 518ï¿½522https://doi.org/10.1093/molbev/msx281 PMID: 29077904

109. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol. 2013; 30: 772ï¿½780https://doi.org/10.1093/molbev/mst010 PMID: 23329690

110. Steenwyk JL, Buida TJ, Li Y, Shen X-X, Rokas A. ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. PLoS Biol. 2020; 18: e3001007. https://doi.org/10.1371/journal.pbio.3001007 PMID: 33264284

111. Sun S, Priest SJ, Heitman J. Cryptococcus neoformans Mating and Genetic Crosses. Curr Protoc Microbiol. 2019; 53: e75. https://doi.org/10.1002/cpmc.75 PMID: 30661293

112. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. Nat Methods. 2012; 9: 676ï¿½682https://doi.org/10.1038/nmeth.2019 PMID: 22743772

113. Winston F. EMS and UV mutagenesis in yeast. Curr Protoc Mol Biol. 2008; Chapter 13: Unit 13.3B. https://doi.org/10.1002/0471142727.mb1303bs82 PMID: 18425760