How host-like signals drive gene expression and gene expression drives capsule expansion in *Cryptococcus neoformans*

Jeffery Jung[1,2,3], Yu Sung Kang[1,2,3], Holly Brown[3], Chase Mateusiak[3], Tamara L. Doering[5], and Michael R. Brent[2,3,4,6]

1. These authors contributed equally.

2. Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, MO 63110

3. Department of Computer Science and Engineering, Washington University, St. Louis, MO 63108

4. Department of Genetics, Washington University School of Medicine, St. Louis, MO 63110

5. Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, MO 63110

6. Corresponding author

Michael Brent

brent@wustl.edu

http://mblab.wustl.edu

314-660-2205

Campus Box 8510

Washington University

Saint Louis, MO 63130

ABSTRACT

Cryptococcus neoformans is an opportunistic fungal pathogen with a polysaccharide capsule that becomes greatly enlarged in the mammalian host and during *in vitro* growth in response to host-like conditions. To understand how individual host-like signals affect capsule size and gene expression, we grew cells with and without all combinations of 5 signals suspected of affecting capsule size and systematically measured cell and capsule sizes of 47,458 cells. We also collected samples for RNA-Seq at 30, 90, 180, and 1440 minutes and carried out RNA-Seq in quadruplicate, yielding 881 RNA-Seq samples. This massive, uniformly collected dataset will be a significant resource for the research community. Analysis revealed that capsule induction requires both tissue culture medium and either CO_2 or exogenous cyclic AMP, a second messenger. Rich medium (YPD) blocks capsule growth completely, DMEM permits it, and RPMI yields the largest capsules. Medium has the biggest impact on overall gene expression, followed by CO_2, mammalian body temperature (37° compared to 30°), and then cAMP. Surprisingly, adding CO_2 or cAMP pushes overall gene expression in the opposite direction from tissue culture media, even though both tissue culture medium and CO_2 or cAMP are required for capsule development. By modeling the relationship between gene expression and capsule size, we identified novel genes whose deletion affects capsule size.

Late in the 19th century, several scientific articles described a budding yeast with a distinctive capsule, now called *Cryptococcus neoformans*. Today, we know *C. neoformans* as the causative agent of deadly meningitis that kills roughly 200,000 people each year worldwide.[1] We also know much about the capsule that surrounds this pathogen, including the chemical structure of the polysaccharides that compose it, its key role in disease, and the fact that it is exquisitely sensitive to environmental conditions.[2] Upon entry to a mammalian host, the capsule dramatically increases in thickness, from a barely perceptible structure to a distinctive shell whose thickness can exceed the cell's diameter.[3] Enlarged capsules inhibit phagocytosis of the yeast by host immune cells and shed capsule polysaccharides inhibit host defenses.[4] The importance of this material in cryptococcal virulence is amply supported by the reduced virulence of strains in which capsule is altered or dysregulated.[2]

Multiple *in vitro* conditions induce the growth of cryptococcal capsule.[5-10] Conditions that reflect aspects of the mammalian host environment are of particular interest, such as those that include tissue culture medium (TCM) or mammalian serum[9], human body temperature (37°), host-like CO_2 concentrations (~5%)[6], host-like pH (7.35-7.45),[6,9-12] and nutrient limitations typical of the host environment.[7] We have previously shown that transcription is required for capsule growth.[13] In this paper, our goal is to understand how environmental signals are transmitted via specific gene expression programs to ultimately yield enlarged capsules. To do this we have isolated key features of capsule inducing conditions, which we call *signals*, and assessed their effects on cell size, capsule size, and gene expression over time during the capsule induction process.

Cyclic AMP (cAMP) signaling is required for capsule growth and virulence.[3,14] The signaling pathways that lead from glucose sensing to production of cAMP, activation of Protein Kinase A (PKA), and some of their effects on transcription factor activity have been well worked out in *Saccharomyces cerevisiae.*[15,16] Much of this machinery can also be found in *C. neoformans*.[14] When activated, the protein Cac1 (adenylyl cyclase) converts ATP to cAMP, which causes the repressive subunit of the PKA complex, Pkr1, to separate from the catalytic subunit, Pka1. This activates Pka1, allowing it to phosphorylate transcription factors including Nrg1[17] and Rim101[18]. Pde1 and Pde2 (phosphodiesterases) convert cAMP to AMP, reducing cAMP signaling. Cells lacking Cac1[3,19] or Pka1[18,20] fail to make capsule in conditions that induce capsule in wild-type (WT) cells, while Pkr1-deficient cells make larger capsules than WT cells in capsule-inducing conditions[3,21] Previous evidence suggests that the capsules of *pkr1*Δ mutants in rich yeast medium (YPD) are the same size as those of WT cells[22] (despite reports to the contrary in a review article citing ref[22]). Thus, while cAMP signaling is necessary for capsule growth, it may not be sufficient.

Here, we systematically explore the effects of all possible combinations of 4 capsule-inducing signals on gene expression, cell size, and capsule size. These signals are tissue-culture medium (DMEM or RPMI), temperature (37°C), CO_2 (5%), and the addition of HEPES buffer. For each combination of conditions, we grew cultures with or without exogenous cAMP, and sampled cells for RNA-Seq at 0, 30, 90, 180, 1440 minutes and for

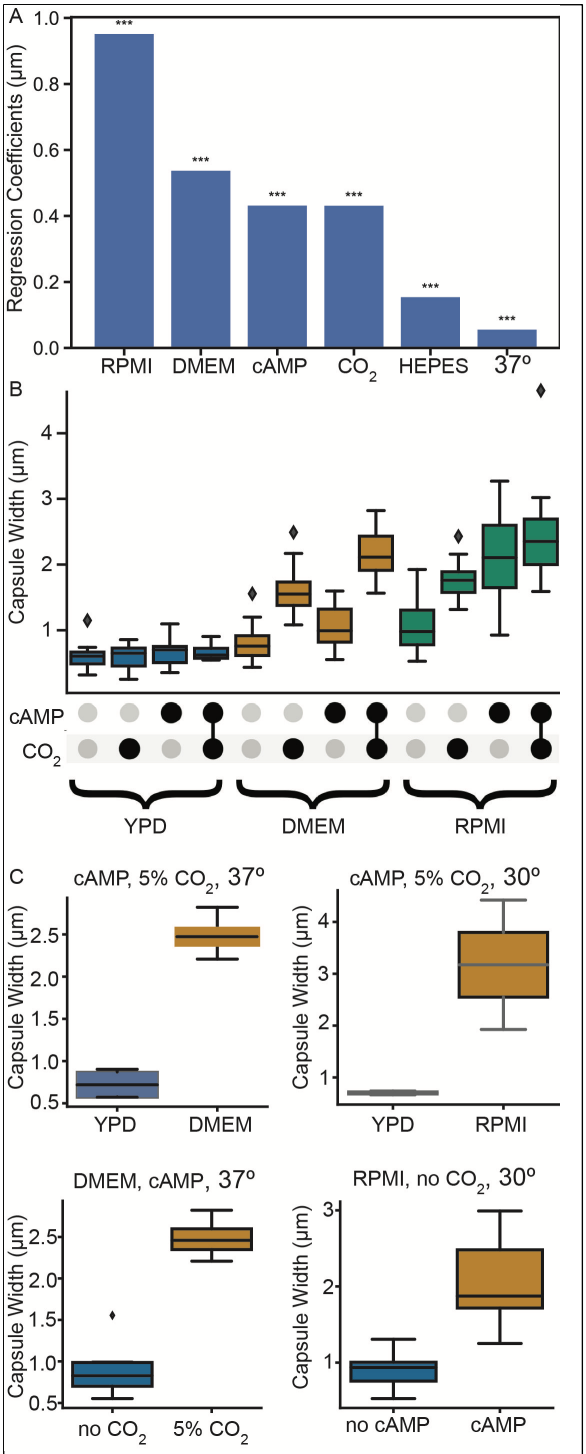

Figure 1 (A) Average effects of environmental signals on capsule size in a linear model without interaction terms. *** P<0.001 (B) Capsule sizes for all combinations of media, cAMP, and CO2 signals. (C) Conditions in which individual signals had their biggest effects on capsule size. None of these conditions includes buffer.

measurement of capsule thickness at 1440 min (24 h). We explore the effects of exogenous cAMP at a range of concentrations and the effects of deletions of cAMP pathway genes *PKR1* and *PDE1*. In total, we have analyzed 881 RNA-Seq samples and 5,175 India Ink micrographs. This large, uniformly collected dataset will be a significant resource for the Cryptococcus community.

Our initial analysis of this dataset has given us an unprecedented ability to trace capsule-inducing signals through their effects on gene expression over 24 hours to their ultimate effects on capsule size. Insights from our initial analyses of these datasets include:

1. Growth media (YPD, DMEM, or RPMI) have bigger effects on gene expression than the other signals studied, but they have little effect on capsule size in the absence of other signals.
2. In the presence of 5% CO_2 or 20 mM cAMP, YPD completely blocks capsule growth, DMEM permits capsule growth, and RPMI yields the largest capsules.
3. CO_2 and cAMP have similar, but distinguishable, effects on gene expression and capsule size.
4. The overall effects of CO_2 or cAMP on capsule gene expression are strongly negatively correlated with those of DMEM or RPMI.

This dataset has enabled us to home in on the changes in gene expression that are most consistently associated with capsule growth across a wide range of 'inducing' treatments and genetic perturbations and to discover three new genes that influence capsule thickness.

RESULTS

<u>Growth medium has a big effect on capsule size, temperature and pH have small effects, and CO_2 and cAMP have intermediate effects.</u>

We investigated the effects on capsule size of all possible combinations of the following growth condition variables: medium (YPD, DMEM, RPMI), CO_2 (room, 5%), temperature (30°, 37°), HEPES buffer pH 7.0 (none, 25 mM), and cAMP (none, 20 mM). For these studies we grew at least four replicate cultures in each combination of conditions for 24 hours, negatively stained the cells with India ink, imaged them at 63X magnification, and manually annotated the images using a custom annotation interface. Fifteen fields of view were annotated for each replicate of each combination, yielding an average of 108 annotated cells per replicate.

We first examined how each individual variable affects capsule size, adjusting for the independent effects of all other variables. To do this, we built a linear regression model with the environmental variables as predictors of capsule size. The resulting regression coefficients showed that the factor with the biggest impact on capsule size was RPMI medium, which increased capsule width an average of 0.86 μm, followed by DMEM (0.49 μm), CO_2 (0.43 μm), and cAMP (0.33 μm) (Fig. 1A). All signals increased capsule size and were statistically significant ($P< 0.001$), but the effects of increasing temperature from 30° to 37° or adding HEPES buffer were small. As result, we combined replicates, regardless of temperature or added buffer, and plotted capsule size for all possible combinations of medium, CO_2, and cAMP (Fig. 1B). Strikingly, the results showed that there was no capsule induction in any condition with YPD, regardless of CO_2 or cAMP addition. Capsule could be induced in either DMEM or RPMI, but capsules were bigger in RPMI in every combination of CO_2 and cAMP. DMEM alone yielded negligible induction, but the combination of DMEM with CO_2 or cAMP increased capsule sizes. CO_2 had a bigger effect than cAMP and the combination of both yielded the largest capsules. The results in RPMI were similar, except that RPMI alone produced larger capsules than DMEM alone. cAMP had a bigger effect in RPMI than in DMEM, such that cAMP and CO_2 were almost equally effective in RPMI. Next, we compared combinations of conditions that differed by just a single factor (Fig. 1C). DMEM and RPMI each had their biggest effects in the presence of both CO_2 and cAMP. cAMP had its biggest effect in RPMI without CO_2 at 30°, while CO_2 had its biggest effect in DMEM with cAMP at 37°.

In summary, no combination of temperature, CO_2, cAMP, or buffer induced capsule at all in YPD, suggesting that it has a repressive effect on capsule. RPMI generally led to greater capsule width than DMEM, holding all other variables constant. In RPMI or DMEM, CO_2 or cAMP each increased capsule size, and the combination of the two yielded the largest capsules.

Effects of signals on gene expression mirror their effects on capsule size

During the capsule induction studies described above, we took samples for RNA-Seq before starting the induction and at 30, 90, 180, and 1440 min. After quality control (see Methods), we were left with RNA-Seq data on 720 samples. At least 4 replicates for each combination of environmental signals had been grown and libraries prepared on separate

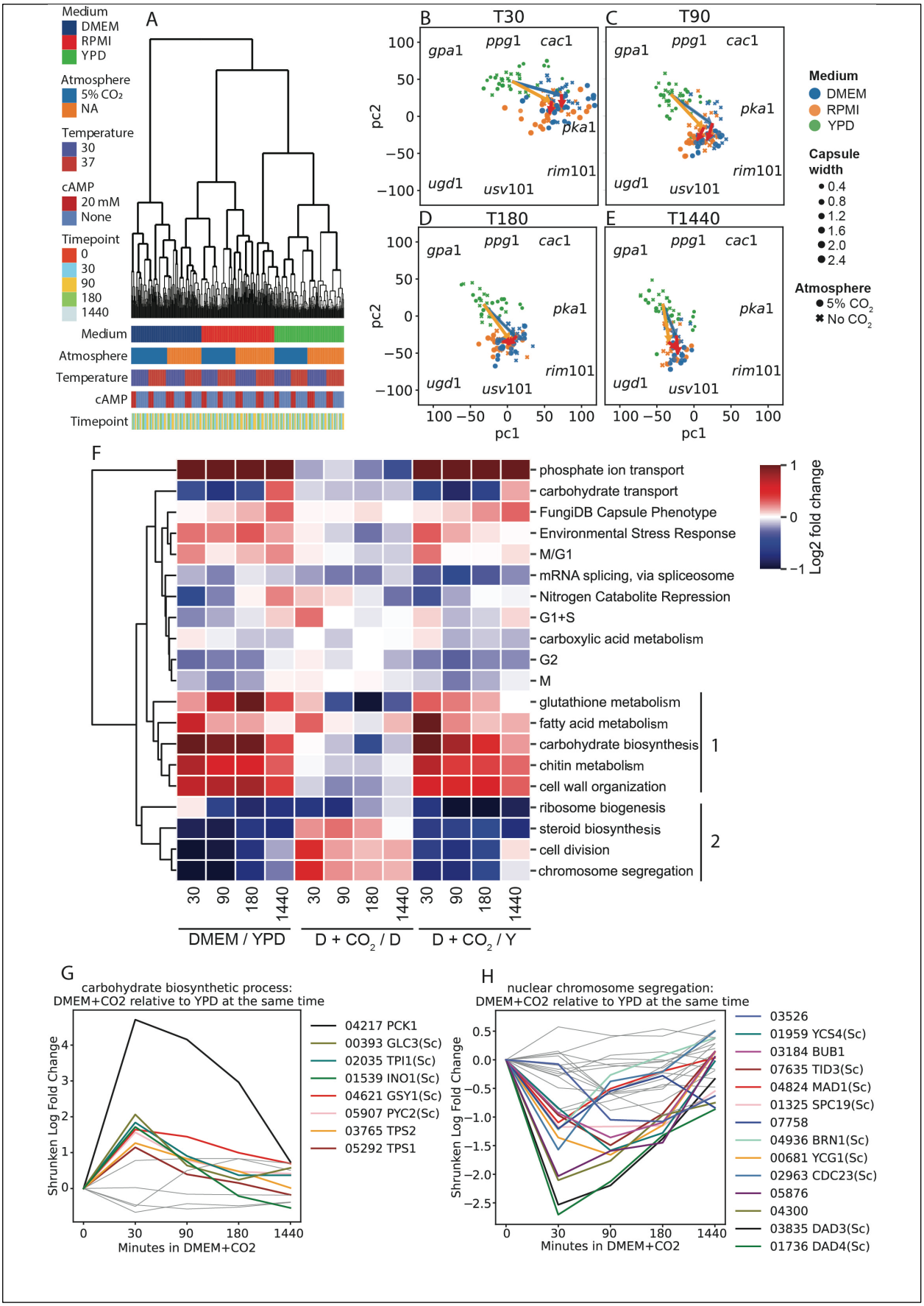
A
Medium
DMEM
RPMI
YPD
Atmosphere
5% CO2
NA
Temperature
30
37
cAMP
20 mM
None
Timepoint
0
30
90
180
1440
Medium
Atmosphere
Temperature
cAMP
Timepoint
B T30
C T90
D T180
E T1440
gpa1
ppg1
cac1
pka1
rim101
ugd1
usv101
pc1
pc2
Medium
DMEM
RPMI
YPD
Capsule width
0.4
0.8
1.2
1.6
2.0
2.4
Atmosphere
5% CO2
No CO2
F
phosphate ion transport
carbohydrate transport
FungiDB Capsule Phenotype
Environmental Stress Response
M/G1
mRNA splicing, via spliceosome
Nitrogen Catabolite Repression
G1+S
carboxylic acid metabolism
G2
M
glutathione metabolism
fatty acid metabolism
carbohydrate biosynthesis
chitin metabolism
cell wall organization
ribosome biogenesis
steroid biosynthesis
cell division
chromosome segregation
1
2
Log2 fold change
DMEM / YPD
D + CO2 / D
D + CO2 / Y
G
carbohydrate biosynthetic process:
DMEM+CO2 relative to YPD at the same time
Shrunken Log Fold Change
Minutes in DMEM+CO2
04217 PCK1
00393 GLC3(Sc)
02035 TPI1(Sc)
01539 INO1(Sc)
04621 GSY1(Sc)
05907 PYC2(Sc)
03765 TPS2
05292 TPS1
H
nuclear chromosome segregation:
DMEM+CO2 relative to YPD at the same time
Shrunken Log Fold Change
Minutes in DMEM+CO2
03526
01959 YCS4(Sc)
03184 BUB1
07635 TID3(Sc)
04824 MAD1(Sc)
01325 SPC19(Sc)
07758
04936 BRN1(Sc)
00681 YCG1(Sc)
02963 CDC23(Sc)
05876
04300
03835 DAD3(Sc)
01736 DAD4(Sc)

Figure 2 (A) Hierarchical clustering of all RNA-Seq samples by Euclidean distance. (B-E) RNA-Seq samples at each time point plotted on the top two principal components of all samples (PCs are the same for all plots). Arrows: From the centroid of YPD samples to the centroid of DMEM samples without CO_2 (green) or RPMI samples without CO_2 (orange). Red arrows: From centroids of DMEM or RPMI samples without CO_2 to centroids of corresponding media with CO_2. (F) Log2 fold changes (LFC) for all genes in the indicated functional categories in DMEM relative to YPD at the same time point (DMEM / YPD), in DMEM with CO_2 relative to DMEM without CO_2 (D + CO_2 / D), or in DMEM with CO_2 relative to YPD (D + CO_2 / Y). Colors indicate the average LFC for all genes in the category, including those that are not significantly differentially expressed. (G-H) LFCs of individual genes (denoted by the numeric part of CKF44 gene IDs) that are annotated with carbohydrate synthesis (G) or chromosome segregation (H). Colored lines indicate genes with absolute LFC > 1 at some time point.

dates to account for day-to-day variability in both processes. Each growth batch and library preparation batch also included a control culture in standard laboratory conditions: YPD, 30°, no CO_2, no cAMP, no added buffer. This design allowed us to regress out the batch (date) effects. We then computed the Euclidean distances between logged, normalized counts for all pairs of samples and carried out hierarchical clustering (Fig. 2A). Notably, all replicates with the same medium, atmosphere, temperature, and cAMP status clustered together, indicating highly effective batch correction. Consistent with their effects on capsule size, media had the biggest effect on gene expression state, followed by CO_2. Notably, temperature had a greater effect on gene expression than cAMP, even though cAMP had a greater effect on capsule size. Time was the least differentiating variable, suggesting that the bulk of the response to condition changes was established by 30 minutes.

Next, we carried out principal components analysis (PCA) on the log normalized expression levels of all genes in all samples and plotted samples for each time point separately (Fig. 2B-E). Comparing the four plots to consider the data over time, there is an overall migration of all points toward the lower left, but the blue and orange points representing cells in TCM move more than the YPD points (green). Arrows show the effects of DMEM (blue) or RPMI (orange) relative to YPD, in the absence of CO_2, and the effects of CO_2 in those TCM (red lines). At 30 and 90 minutes, the effects of CO_2 are largely orthogonal to those of the TCM, indicating distinct effects on gene expression. By 24 hours, however, the differences among all cultures are smaller and the effects of CO_2 are more aligned with those of TCM. While DMEM and RPMI move cells to slightly different areas, the further addition of CO_2 causes them to move to essentially the same gene expression state.

To gain further insight into the effects of TCM and CO_2 on gene expression, we ran DESeq2[23] with a linear model that predicts log normalized gene expression levels from signals. We used the model to calculate shrunken log fold changes in response to each signal (see Methods). Next, we carried out Gene Ontology (GO) enrichment analysis on the genes that were most responsive to DMEM or RPMI, CO_2, and the combination of both, at each time point. From among the enriched GO biological process terms, we selected those that were most interesting and interpretable and created a heatmap showing the average expression levels of all genes annotated with each term (not just the significantly differentially expressed genes; Fig. 2F). For each time point, we show the effects

of DMEM relative to YPD, the effects of DMEM+CO_2 relative to DMEM alone, and the combined effects of DMEM+CO_2 relative to YPD alone.

Genes involved in phosphate transport are strongly induced by DMEM, consistent with recent reports that they are induced by another capsule-inducing growth condition, 10% Sabouraud's dextrose medium.[24] Other genes that are strongly upregulated in DMEM (Fig. 2F, group 1) include those involved in cell wall synthesis, adaptation to low glucose, and stress response. For example, induced genes involved in carbohydrate biosynthesis (Fig. 2G) included those involved in gluconeogenesis, glycogen synthesis, and trehalose synthesis, all of which are typical of adaptation to lower glucose. Trehalose synthases *TPS1*[25] and *TPS2*[26] are also required for virulence in *C. neoformans*[25] and in the related species *Cryptococcus gattii*[27]. Genes annotated with a capsule phenotype in FungiDB are, on average, upregulated by DMEM and barely affected by CO_2. Homologs of yeast genes expressed specifically during the M/G phases of the cell cycle are upregulated by DMEM and slightly down regulated by CO_2 while those expressed at other phases are generally downregulated by DMEM, relative to YPD.

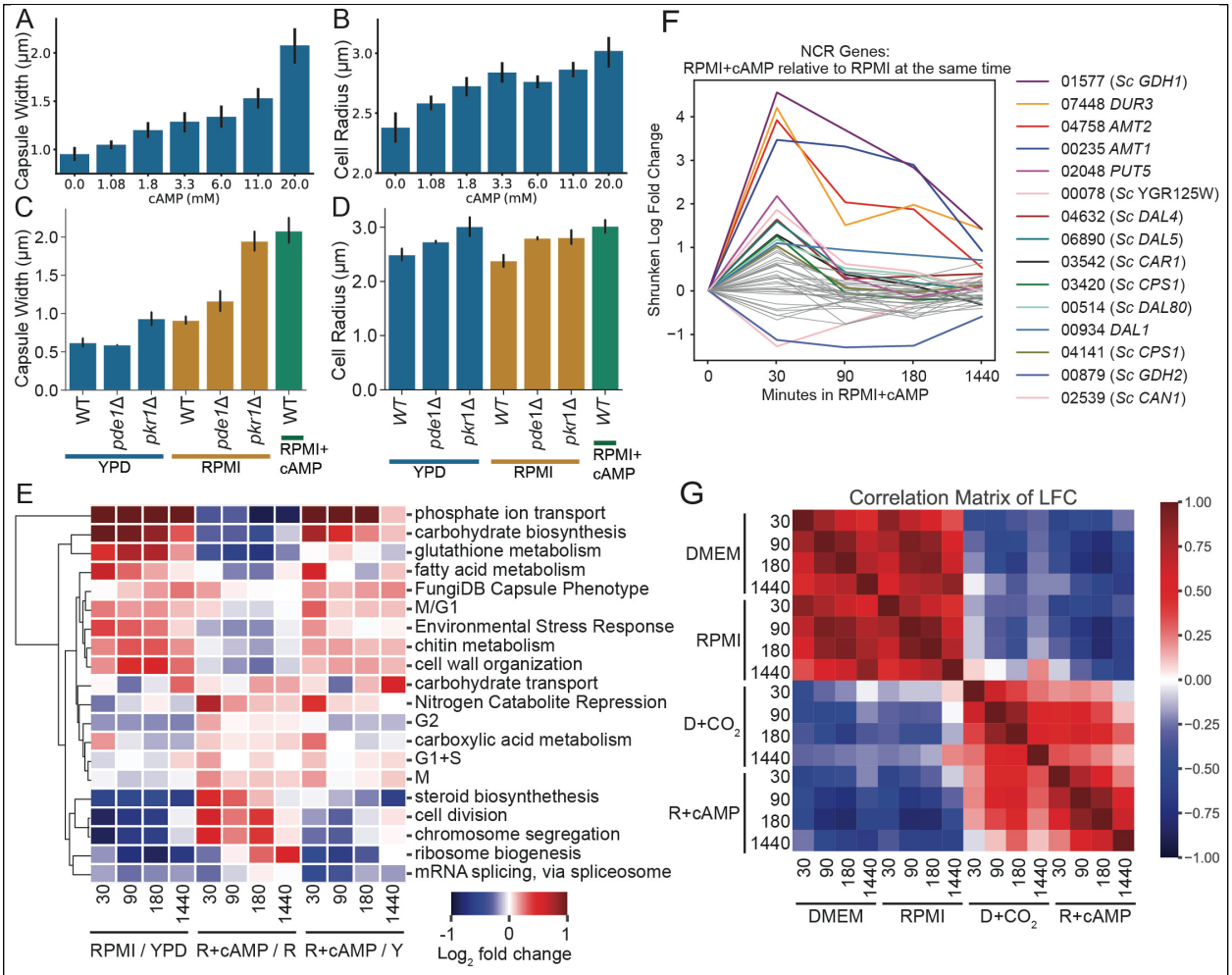

Figure 3 (A) Capsule width or (B) cell body radius as a function of added cAMP. (C) Capsule width or (D) cell body radius in strains lacking cAMP-pathway gene *PDE1* or *PKR1*. The effect of cAMP can be explained almost entirely by its inactivating effect on Pkr1. (E) Log2 fold changes (LFC) for all genes in the indicated functional categories in RPMI (no cAMP) relative to YPD at the same time point (RPMI / YPD), in RPMI with 20 mM cAMP relative to RPMI without cAMP at the same time point (R + cAMP / R), or in RPMI with 20 mM cAMP relative to YPD at the same time point (R + cAMP / Y). Colors indicate the average LFC for all genes in the category, including those that are not significantly differentially expressed. (F) LFCs of individual genes (denoted as in Fig. 2) that are repressed by the Nitrogen Catabolite Repression (NCR) pathway. Colored lines indicate genes with absolute LFC > 1 at some time point. (G) Correlations between effects of tissue culture media, CO_2, and cAMP, on expression of all genes. Broadly, the effects of the two tissue culture media without CO_2 or cAMP are positively correlated at all time points. The effects of CO_2 in DMEM and of cAMP in RPMI, relative to each medium alone, are also positively correlated at all time points. However, the effects of CO_2 or cAMP, when added to tissue culture media, are negatively correlated with the effects of the media alone.

The strongly down regulated gene sets (Fig. 2F, group 2) are mainly involved in growth, cell division, and carbohydrate transport. For example, many genes involved in chromosome segregation are down regulated (Fig. 2H). Among the down-regulated carbohydrate transport genes are *LPI8*,[28] whose product affects phagocytosis, and *GMT2*, which encodes a GDP-mannose transporter.[29] Interestingly, for most gene sets in the heat map, the effect of CO_2, when added to DMEM, was in the opposite direction from

the effect of DMEM relative to YPD. However, the CO_2 response was weaker so the overall effect of DMEM+CO_2 is in the same direction as DMEM alone. Exceptions to that rule include ribosome biogenesis and mRNA splicing, both of which are consistently down.

Effects of cyclic AMP on cell size, capsule size, and gene expression

Cyclic AMP signaling is essential for capsule growth.[3,14,19] To investigate the effects of exogenous cAMP on capsule size and gene expression, we carried out experiments in conditions in which capsule is normally small, but addition of CO_2 or cAMP generates much larger capsules: RPMI, 30°, room air, no HEPES. We found that adding cAMP at 1.1, 1.8, 3.3, 11, and 20 mM (increasing by a factor of ~1.8 at each step) showed a consistent dose response in capsule width (Fig. 3A), with the logarithm of exogenous cAMP concentration a highly significant predictor of average capsule size ($P<10^{-10}$). On average, increasing cAMP by a factor of 1.8 increased the capsule size by 0.2 μm, although the final step from 11 mM to 20 mM had a much bigger effect. Exogenous cAMP also increased the average cell body radius, from center to cell wall ($P < 3*10^{-5}$), with a 1.8 fold increase in cAMP causing an average increase of 0.11 μm (Fig. 3B). Unlike capsule width, the effect on cell size was greater at lower concentrations – the difference between 0 mM and 3.3 mM was 0.46 μm while the difference from there to 20 mM was only 0.18 μm. The effect of cAMP on cell size does not explain its effect on capsule width, since the capsule index (capsule width as a fraction of total radius) increased with cAMP concentration (Fig. S1).

To further investigate the role of the cAMP, we measured the capsules of cells lacking *PDE1* or *PKR1*, growing in either YPD or RPMI (both at 30° with no CO_2 or cAMP; Fig 3C). *PDE1* encodes a phosphodiesterase capable of reducing the intracellular cAMP concentration. Consistent with previous reports,[30] *PDE1* deletion had a very small effect on capsule size. *PKR1* encodes the repressive moiety of the Protein Kinase A (PKA) complex. cAMP causes Pkr1 to dissociate from the complex, activating the kinase moiety. Thus, the *pkr1* deletion mutant might be expected to have an enlarged capsule and indeed we see that, when grown in RPMI without CO_2 or cAMP, the *pkr1* mutant's capsule is similar to that of WT grown with 20 mM cAMP (Fig. 3C). This suggests that the main effect of cAMP on capsule size is mediated by its effect on Pkr1. We also saw an increase in capsule size of *pkr1* growing in YPD. It has been reported previously that *pkr1* has larger capsule than WT cells when grown in fully capsule-inducing conditions (low iron medium with EDDHA[22] or DMEM 37° 5% CO_2[3]), but we have found no previous evidence that its capsule is enlarged in conditions such as RPMI or YPD at 30° without other inducing factors. No other perturbation, including 20 mM exogenous cAMP, yielded increased capsule size in YPD (Fig. 1B). A difference between deletion of *PKR1* and addition of 20 mM cAMP is that, in RPMI, the deletion does not increase cell body size as much as the addition of cAMP does (Fig. 3D).

To better understand how cAMP signaling increases capsule size, we characterized subsets of cAMP responsive genes at each time point, as in Figure 2F. The effect of adding cAMP or CO_2 tended to moderate the responses to TCM more often than it reinforced them. Overall, the RPMI and cAMP responses were broadly similar to the DMEM

and CO_2 responses, respectively (Fig. 3E). However, there were differences. Both RPMI and DMEM reduced the expression of ribosome biogenesis genes, but adding CO_2 reinforced that effect at all time points, while adding cAMP moderated it at later time points. Genes involved in carboxylic acid metabolic processes (mainly nitrogen assimilation and amino acid metabolism) also responded to RMPI+cAMP and DMEM+CO_2 differently, especially at early time points. We noticed that many of these genes were targets of nitrogen catabolite repression (NCR), a process that represses the expression of certain genes when preferred nitrogen sources such as glutamine (the main nitrogen source in DMEM and RPMI) are available.[31] RPMI engaged NCR, repressing NCR genes. However, adding cAMP released NCR rapidly, most notably at 30 min (Fig. 3F), suggesting that the cells responded to cAMP as though they had been moved to a less preferred nitrogen source (despite the presence of glutamine in the medium). This overcame the RPMI effect, so that the combined effect of RPMI and cAMP was to relieve NCR (consistent with less preferred nitrogen sources), relative to YPD.

Across all genes, the two tissue culture media (DMEM and RPMI) have highly correlated effects on gene expression (Fig. 3G). The effects of CO_2 and cAMP are negatively correlated with those of tissue culture medium, especially at 180 min. However, by 1440 minutes, the effects of adding CO2 to DMEM show a slight positive correlation with the effects of RPMI relative to YPD. This makes sense because both treatments increase capsule size at 1440 min. The effects of CO_2 and cAMP were somewhat positively correlated with each other, but they diverge by 1440 min, at which point they have only a weak positive correlation.

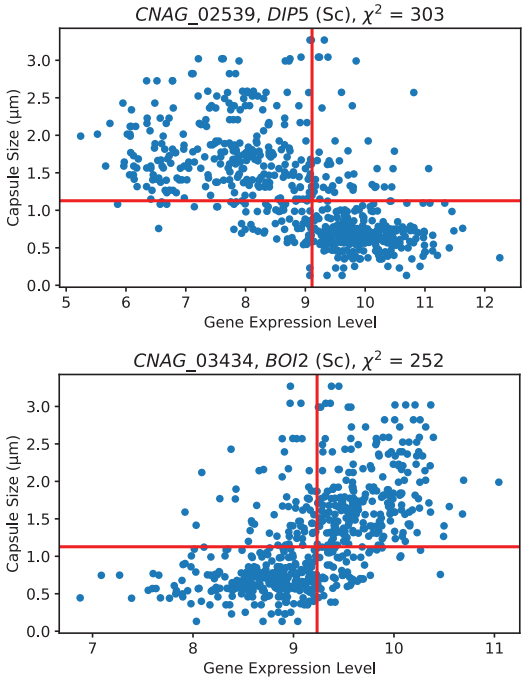

Figure 4 Examples of genes with high χ^2, reflecting the visual relationship between their expression levels and capsule sizes. Cells with capsules larger than the horizontal lines are considered induced while genes with expression levels to the right of the vertical red lines are considered highly expressed. See Methods for details.

Identifying genes whose expression levels are associated with capsule size

Next, we set out to identify genes whose expression is strongly associated with capsule development. The first approach was to dichotomize capsule size into “induced” or “not induced”, dichotomize the expression of each gene into “high” or “low”, and analyze the relationship between expression and capsule induction for each gene separately. Since YPD completely blocks capsule induction in WT cells, regardless of other signals, we set a threshold for “induced” at 1.13 µm, three standard deviations above the mean of all WT cells in YPD. We searched for the most meaningful split between high and low expression for each gene by trying each observed level

as a possible threshold. For each possible threshold, we calculated the χ^2 statistic for association between induction status and gene expression status (Fig. 4A, B). We were not concerned with P-values here, but with using the statistic itself to rank genes by their association with capsule size. The maximum χ^2 value, over all time points, was assigned to the gene and the genes with the highest χ^2 values were identified as possibly involved in capsule induction and slated for testing (Table S1).

In addition to using dichotomized values, we used continuous expression levels and capsule sizes to identify genes whose expression levels are highly predictive of capsule width. Specifically, we trained XGBoost[32], a tree boosting algorithm, to predict the mean capsule size in a sample from features consisting of the expression levels of all genes at all time points. Before training, we iteratively merged features with correlation >= 0.8 across all RNA-Seq samples, replacing them with a single feature, the average of the correlated features. This avoids the tendency of regression models to choose between correlated features arbitrarily or to be overly sensitive to small fluctuations in correlated features. We then calculated a SHAP value for each feature in each prediction.[33] The SHAP value explains a prediction in terms of how much each feature value influences the model to increase or decrease its capsule width prediction for that sample, relative to the average capsule width. To calculate a single influence score for each feature, we summed the absolute SHAP values for that feature over all the samples. The genes with the largest influence scores were then slated for testing, where deletion mutants were available (Table S2).

Testing predicted capsule-associated genes

Mutants lacking the genes listed in Tables S1 & S2 were tested for altered capsule size if the mutant was available to us, either through a deletion collection[34] or in our own collections. Conditions we tested included non-inducing (YPD, 37°, 5% CO_2, with or without cAMP), 'almost-inducing' (RPMI, 37°, room air, no cAMP), or inducing (RPMI, 37°, 5% CO_2, no cAMP). Consistent with previous reports,[8] capsule widths in our studies were highly variable among biological replicates. To overcome this challenge, we annotated at least 100 cells in each of at least three replicates performed on three different days.

We did not identify any mutants that generated enlarged capsule in non-inducing (YPD) conditions. In almost-inducing conditions, no mutants had mean capsule width that differed from WT by 0.3 μm or more. In inducing conditions, however, we identified several mutants that were not known previously to influence capsule width: Deletion of CNAG 00368 reduced mean capsule width by 0.44 μm and deletion of CNAG 00328 (encoding a homolog of the yeast DNA excision repair protein Rad2) reduced mean capsule width by 0.58 μm. Deletion of a DNA excision repair protein could lead to increased secondary mutations, but genome sequencing of this strain did not show such mutations. A strain carrying a deletion of CNAG 05977 had capsule width 0.54 μm smaller than wild-type cells in inducing conditions. This strain was not chosen by the methods described above – we intended to test CNAG 06050, but routine genome sequencing for quality control showed that the strain in fact had a deletion of CNAG 05977, not 06050. In a linear model

with fixed effects for genotype and random intercepts for each sample, the effects of deleting CNAG 00328 (4 replicates) and CNAG 05977 (5 replicates) were statistically significant with Benjamini-Hochberg FDR < 0.05.

METHODS

Additional details for all methods can be found in the online supplement.

Cell growth

To maximize reproducibility of RNA-seq and capsule imaging studies, cell recovery from frozen stocks, initial culture in YPD, inoculation into preconditioned media, and growth followed a strictly controlled protocol. These methods are detailed in the Supplemental Material.

Microscopy and manual image annotation

1-ml samples were collected for imaging from the stock cell suspension prior to inoculation of flasks or from cultures at 1440 min, fixed, resuspended in PBS, and mixed with India ink (5 parts cells:2 parts ink) for brightfield microscopy imaging. Images were manually annotated using a custom annotation interface written in Mathematica / Wolfram Language (available on request). Fifteen fields were annotated for each replicate of each combination, yielding an average of 92 annotated cells per replicate.

Analysis of capsule thickness in gene deletion mutants

Effects of deletion mutants on capsule thickness were evaluated using a Linear Mixed Models framework with individual cells' thicknesses as the datapoints. The model had fixed effects for intercept and genotype; each specific biological sample was used as a grouping variable with a random intercept for each group.

RNA-Seq Computational Pipeline and Analysis

Reads were aligned with Novoalign (version 4.03.02) and quantified with HTSeq (0.9.1) using the FungiDB KN99α genome sequence. Custom scripts were used to verify the strain in each sample by calculating CDS coverage over putatively deleted genes and over marker genes, ensuring that the former were absent and the latter present. RNA-Seq samples that passed strain validation were subjected to two phases of QC. In Phase 1, files were labeled as 'passing' if they contained at least 10^6 reads aligned to protein coding regions and less than 7% of all reads failed to align.

In Phase 2, we evaluated replicate agreement using the Regularized Log Expression (RLE) [35]. First, we used DESeq2 (version 1.34.0) to estimate the effect of the library date (the known batch effect). We removed the batch effect using the DESeq coefficients for the library dates such that the data was standardized to a single date, resulting in adjusted expression levels on a $\log_2$ normalized scale. To compute the RLE value for a given gene, the median expression level of that gene, across all samples in a replicate set, was subtracted from the expression level of the gene in each sample. For each sample, we then calculated the interquartile range of the distribution of these RLE values across genes. If the interquartile range of RLEs of a given sample was greater than 1, indicating that more than half of genes deviated from their respective medians by a factor of 2 in adjusted count, the sample was considered an outlier and failed for replicate agreement. 22 samples (2.75%) failed replicate agreement and were discarded.

Selection of potential capsule-associated genes: Chi-squared method

The following procedure was carried out separately for samples at each time point. To calculate the χ^2 statistic for each gene, we first classified all samples with average capsule width greater than 1.13 μm (3 S.D. above the mean of all YPD conditions) as induced. Other samples were classified as uninduced. Samples were also classified by high or low expression of each gene, yielding a 2x2 contingency table. For each gene, we tried all possible thresholds of high versus low expression, constructed the contingency table, and calculated the χ^2 statistic. Finally, we chose the threshold that yielded largest χ^2 statistic. These statistics were only used to characterize the degree of association between the expression of a gene in a sample and that sample's induction status. They were not used for hypothesis testing.

Genes were ranked for likelihood of the corresponding deletion strain exhibiting a capsule phenotype by the maximum of their χ^2 values, across all time points. All genes with max χ^2 statistics above 200 were selected for testing. For genes that were positively correlated with capsule size, the gene deletion mutant was tested in inducing conditions (RPMI, 37°C, 5% CO_2) with the prediction that deletion would reduce capsule size compared to WT grown in the same conditions. For negatively correlated genes, the gene deletion mutant was tested in almost-inducing conditions (RPMI, 37°, room air, no cAMP, no buffer) and non-inducing conditions (YPD, 37°, 5% CO_2, with or without cAMP, no buffer).

Forming metagenes as features for machine learning

From the original gene expression data matrix, we sought to decrease the number of features by filtering and combining genes. We first removed low variance genes by filtering out genes whose expression in >= 95% samples was within one $\log_2$ of their mean, indicating that less than 5% of samples showed substantial changes. We next combined highly correlated genes into metagenes (correlation threshold > 0.8). The "expression level" of a metagene was then set to the mean of the expression levels of its constituent genes. We repeated this process to combine genes into metagenes until we had combined all highly correlated genes into metagenes.

Selection of genes to test for capsule phenotypes: XGBoost method

Using the metagene expression matrix, we trained a model to predict capsule size at each time point. We then calculated the SHAP value for each metagene in each sample. The SHAP value of a given gene in a given sample is an estimate of how much that gene's expression level affects the model's prediction for that sample.[33] When a metagene was created from more than one gene's expression data, the metagene's SHAP value was attributed to all the contributing genes. We selected the maximum SHAP value across timepoints to associate with each gene and selected genes with absolute mean SHAP value > 0.05 in those conditions for testing. Testing of the corresponding deletion mutant was then performed as described for the χ^2 method. In both analyses, gene deletion mutants that were not in our collection were omitted from our experiments.

If two features are correlated with each other, ML algorithms may select one and discard the other, even if their predictive values are very similar. Therefore, we removed

the selected genes (or their metagenes) from the feature set, retrained XGBoost, and repeated the selection process. These genes are denoted with a 2 in the "Round" column of Table 2.

DISCUSSION

When we undertook this project, it was known that capsule growth requires both new transcription[6,36] and an intact cAMP/PKA pathway.[3,14,19] It was also known that, in tissue culture medium (TCM), capsule can be induced by increasing the concentration of dissolved CO_2/HCO_3^-, achieved by growth in a high CO_2 atmosphere, addition of $NaHCO_3$, or both.[6] We set out to determine how various combinations of capsule inducing signals, with or without exogenous cAMP, would affect gene expression and capsule size. This led to two positive outcomes. (1) Numerous interesting observations about how cells react to capsule-inducing signals. (2) A massive, freely available dataset consisting of RNA-Seq in biological quadruplicate in 42 combinations of potentially capsule-inducing signals, as well as cAMP titrations and mutants in the cAMP signaling pathway (881 total RNA-Seq samples). Each RNA-Seq experiment comes with matched India ink images and capsule width measurements. The dataset also includes images and capsule-width measurements for gene deletion mutants, yielding a total of 47,458 annotated cells in 5,175 images of 392 biological samples.

We found that either a 5% CO_2 atmosphere or cAMP could induce capsule in TCM (DMEM or RPMI), but YPD, a rich medium containing yeast extract and peptone, completely blocks capsule growth. Deleting *PKR1* increases capsule sizes of cells growing in YPD (Fig. 3C), so the repressive effect of YPD may be upstream of PKA activation, such that direct activation by *PKR1* deletion partially overcomes it. However, adding 20 mM exogenous cAMP does not enlarge the capsules of cells growing in YPD (Fig. 1B). Since cAMP is thought to act directly on the Pka1/Pkr1 complex, it may be that cells do not maintain a high enough internal concentration for long enough to fully inactivate Pkr1. RPMI alone caused capsules to enlarge slightly, to a degree that is clearly distinguishable from both uninduced and fully induced, whereas DMEM alone produces a barely detectable enlargement (Fig. 1B). RPMI contains less of most nutrients than DMEM, making it a generally more stressful condition. One good candidate explanation for the greater capsule size in RPMI is that DMEM contains 10^{-4} g/L ferric nitrate whereas RPMI contains no added iron. These results are reminiscent of observations concerning titan cells, extremely large cells relevant to pathogenesis that form during infection and specific *in vitro* conditions.[37] Interestingly, titan cell formation can be induced by low-density inoculation into RPMI at 37 with 5% CO_2, but substitution of DMEM for RPMI prevents this. The critical differences in this case were identified as the presence of iron in DMEM and the presence of para-aminobenzoic acid in RPMI.[38]

In addition to the conditions we studied, capsule growth is known to be stimulated by specific stressors, some of which resemble stresses encountered in a mammalian host. For example, growth in mammalian serum without added nutrients induces capsule growth at 30° or 37°, with or without CO_2, but this is blocked by addition of Sabouraud dextrose broth, a rich medium like YPD.[9] Sabouraud medium alone, diluted to 10% normal concentration and buffered to pH 7.4, is reported to induce capsule (CAP medium).[8]

Iron deprivation in a medium consisting of 5 g/L glucose, 5 g/L asparagine, minerals, and 55 mM EDDA (a chelator) at pH 7.4 can also induce capsule (LIM medium).[7] Thus, rich media consistently block capsule induction while several forms of nutrient deprivation can induce it. Nutrient deprivation is likely the normal state of Cryptococcus, whether in a mammalian host or in the environment, so it makes more sense to think of rich media as blocking capsule growth than to think of TCM as inducing it. Low pH (6.1) can also block capsule induction.[6] High osmolarity appears to reduce capsule thickness, although there has been speculation about whether this is a cellular response or simply physical compression of capsule under high osmolarity.[12] A cellular response is consistent with the observation that deletion of *HOG1*, a key component of the high osmolarity response, increases capsule size in DMEM+CO_2 induction.[3,39] Studying the effects of these stimuli on gene expression may provide additional insights about which changes in gene expression are essential for capsule growth.

The effects of the inducing signals on gene expression largely mirror their effects on capsule size, with medium being the largest influence, followed by CO_2 (Fig. 2A). However, temperature (30° vs. 37°) had a larger effect on overall gene expression than cAMP, while cAMP had the larger effect on capsule width. This suggests that most of the transcriptional changes that result from growth at 37° are not related to capsule expansion. The differences between timepoints in the same condition were much less than the differences between conditions, suggesting that the bulk of the impact of all signals was established by 30 minutes and, more generally, that cryptococcal cells rapidly equilibrate to new environments.

Tissue culture media (TCM) have several interesting effects on the fungal cells. First, they decrease the expression of genes associated with growth and increase that of genes associated with stress. Further, gene expression in TCM suggests an accumulation of cells in the M/G1, post-mitotic phase of the cell cycle. Finally, TCM tended to engage nitrogen catabolite repression, probably because of its glutamine content.

Surprisingly, both CO_2 and cAMP tend to moderate or reduce most of the effects of TCM. The fact that either of these components can act as the necessary partner with TCM for capsule induction, but their overall effects on gene expression are opposite that of TCM, poses intriguing questions: When CO_2 is added to TCM, which of the resulting gene expression changes are critical for capsule induction and do those changes work by dialing down the effects of TCM on key genes, by reinforcing the effects of media on the few genes that respond in the same direction, or by acting on genes that do not respond to TCM? Answering these questions is an exciting direction for future research.

Both nitrogen and phosphate availability have been suspected as regulators of capsule size. Both nutrients activate sensors that stimulate the PKA pathway, including the methionine sensor Gpr4,[40] and, in *S. cerevisiae,* the general amino acid permease Gap1, ammonium sensor Mep2[41] and phosphate sensor Pho84.[42] It has also been reported that expression of phosphate acquisition genes is strongly induced in CAP medium and subsequent addition of KH_2PO_4 to CAP reduces capsule width.[10] However, this may not be a specific effect on capsule, since KH_2PO_4 also greatly reduces cell size. We found that phosphate transporters were also strongly induced by TCM, although this effect is much reduced by addition of CO_2 or cAMP. In other connections to phosphate utilization,

we also found that *NPP1*, which encodes a type 1 nucleotide pyrophosphatase, is the gene whose expression is most predictive of capsule size in our XGBoost/SHAP analysis and *IPP1*, which encodes an organic phosphatase, is also highly predictive (Table S2).

It has also been reported that YNB medium (trace nutrients plus a nitrogen source) with bicarbonate and 5% CO_2 induces capsule, but only if $(NH_4)_2SO$ is replaced by arginine.[6] Both are considered preferred nitrogen sources for *S. cerevisiae*[31], so the cells may be responding to amino acids, rather than quality of nitrogen source. We found that genes previously identified as being repressed in the presence of preferred nitrogen sources are also repressed by TCM, consistent with the presence of glutamine as the main nitrogen source. Furthermore, one of the genes that is most predictive of capsule size is CNAG_01119, the ortholog of *S. cerevisiae PTR2*, which encodes a POT family proton-dependent oligopeptide transporter. Another highly predictive gene is CNAG_02539, ortholog of *DIP5*, which encodes an AAT family amino acid transporter (Tables S1 and S2). However, deletion of *DIP5* did not affect capsule width.

For many of the genes whose expression is most predictive of capsule thickness there is no deletion mutant in the Madhani collection, including *NPP1*, *IPP1*, *PTR2* (CNAG_01119), *GCS2*, and *PAM18* (CNAG_00233) (see Tables S1 and S2 for a complete list). Our own attempts to delete these genes (as well as CNAG_01359, CNAG_02520, CNAG_03985, CNAG_04450, and CNAG_5100) also failed, suggesting that they may be essential. Studying their effects on capsule thickness, therefore, may require construction of over-expression strains or development of more robust tunable promoter systems than are currently available for *C. neoformans*. However, we did identify three genes that were not previously known to influence capsule whose deletion affects capsule size in inducing conditions. Deletion of CNAG 00368 reduced mean capsule width by 0.44 μm. This gene encodes an ortholog of *S. cerevisiae* Vps53, which is involved in recycling of proteins from endosomes to the late Golgi. Vps53 not been studied in Cryptococcus, but another protein that is part of a different complex involved in recycling proteins from endosomes, Vps23, is known to be required for capsule elaboration.[43] Reinforcing the theme of protein and peptide recycling, deletion of CNAG_05977 (proteasome activator subunit 4) reduced capsule size by 0.54 μm in inducing conditions. Finally, deletion of CNAG_00328 (homolog of the *RAD2*) reduced capsule width by 0.58 μm. In addition to its role in DNA excision repair, *RAD2* is required for efficient transcription in yeast.[44]

This paper introduced a dataset of matched RNA-Seq time courses and capsule thickness measurements across 42 combination of growth conditions, by far the largest gene expression study in *Cryptococcus*. So far, this data set has yielded insights into the environmental signals and gene expression changes that affect capsule size. However, this is only the beginning. We have made this remarkable dataset available in the hope it will be analyzed by other researchers. We expect that future analyses will uncover many more insights into the cascade by which environmental signals produce gene expression changes that ultimately lead to the growth of the polysaccharide capsule, a critical virulence factor for this human pathogen.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AVAILABILITY OF DATA AND MATERIALS

All RNA-Seq data are available from NCBI GEO under accession numbers GSE226255, GSE226637, or GSE226651. All capsule size measurements are available in Supplemental File S1. Corresponding India ink images and mutant strains are available on request.

FUNDING

This work was supported by NIH award AI087794.

1. Rajasingham, R. et al. Global burden of disease of HIV-associated cryptococcal meningitis: an updated analysis. *Lancet Infect Dis* **17**, 873-881 (2017).
2. Kwon-Chung, K.J. et al. *Cryptococcus neoformans* and *Cryptococcus gattii*, the Etiologic Agents of Cryptococcosis. *Cold Spring Harb Perspect Med* **4** (2014).
3. Maier, E.J. et al. Model-driven mapping of transcriptional networks reveals the circuitry and dynamics of virulence regulation. *Genome Res* **25**, 690-700 (2015).
4. O'Meara, T.R. & Alspaugh, J.A. The Cryptococcus neoformans capsule: a sword and a shield. *Clinical microbiology reviews* **25**, 387-408 (2012).
5. Littman, M.L. Capsule synthesis by Cryptococcus neoformans. *Trans N Y Acad Sci* **20**, 623-48 (1958).
6. Granger, D.L., Perfect, J.R. & Durack, D.T. Virulence of Cryptococcus neoformans. Regulation of capsule synthesis by carbon dioxide. *The Journal of clinical investigation* **76**, 508-16 (1985).
7. Vartivarian, S.E. et al. Regulation of cryptococcal capsular polysaccharide by iron. *J Infect Dis* **167**, 186-90 (1993).
8. Zaragoza, O. & Casadevall, A. Experimental modulation of capsule size in Cryptococcus neoformans. *Biol Proced Online* **6**, 10-15 (2004).
9. Zaragoza, O., Fries, B.C. & Casadevall, A. Induction of capsule growth in Cryptococcus neoformans by mammalian serum and CO(2). *Infect Immun* **71**, 6155-64 (2003).
10. Denham, S.T. et al. A dissemination-prone morphotype enhances extrapulmonary organ entry by the fungus Cryptococcus neoformans. *bioRxiv*, 2020.08.27.270843 (2022).
11. Farhi, F., Bulmer, G.S. & Tacker, J.R. Cryptococcus neoformans IV. The Not-So-Encapsulated Yeast. *Infect Immun* **1**, 526-31 (1970).
12. Dykstra, M.A., Friedman, L. & Murphy, J.W. Capsule size of Cryptococcus neoformans: control and relationship to virulence. *Infect Immun* **16**, 129-35 (1977).
13. Haynes, B.C. et al. Toward an Integrated Model of Capsule Regulation in Cryptococcus neoformans. *PLoS Pathogens* **7**, e1002411 (2011).
14. Caza, M. & Kronstad, J.W. The cAMP/Protein Kinase a Pathway Regulates Virulence and Adaptation to Host Conditions in Cryptococcus neoformans. *Front Cell Infect Microbiol* **9**, 212 (2019).
15. Zaman, S., Lippman, S.I., Zhao, X. & Broach, J.R. How Saccharomyces responds to nutrients. *Annu Rev Genet* **42**, 27-81 (2008).
16. Conrad, M. et al. Nutrient sensing and signaling in the yeast Saccharomyces cerevisiae. *FEMS microbiology reviews* **38**, 254-99 (2014).
17. Cramer, K.L., Gerrald, Q.D., Nichols, C.B., Price, M.S. & Alspaugh, J.A. Transcription factor Nrg1 mediates capsule formation, stress response, and pathogenesis in Cryptococcus neoformans. *Eukaryot Cell* **5**, 1147-56 (2006).

18. O'Meara, T.R. et al. Interaction of Cryptococcus neoformans Rim101 and protein kinase A regulates capsule. *PLoS Pathogens* **6**, e1000776 (2010).
19. Alspaugh, J.A. et al. Adenylyl cyclase functions downstream of the Galpha protein Gpa1 and controls mating and pathogenicity of Cryptococcus neoformans. *Eukaryot Cell* **1**, 75-84 (2002).
20. Hu, G. et al. Transcriptional regulation by protein kinase A in Cryptococcus neoformans. *PLoS Pathog* **3**, e42 (2007).
21. D'Souza, C.A., Hagen, F., Boekhout, T., Cox, G.M. & Heitman, J. Investigation of the basis of virulence in serotype A strains of Cryptococcus neoformans from apparently immunocompetent individuals. *Curr Genet* **46**, 92-102 (2004).
22. D'Souza, C.A. et al. Cyclic AMP-dependent protein kinase controls virulence of the fungal pathogen Cryptococcus neoformans. *Molecular and cellular biology* **21**, 3179-91 (2001).
23. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
24. Denham, S.T. et al. Regulated Release of Cryptococcal Polysaccharide Drives Virulence and Suppresses Immune Cell Infiltration into the Central Nervous System. *Infect Immun* **86** (2018).
25. Petzold, E.W. et al. Characterization and regulation of the trehalose synthesis pathway and its importance in the pathogenicity of Cryptococcus neoformans. *Infect Immun* **74**, 5877-87 (2006).
26. Jin, J.H. et al. Genome-wide functional analysis of phosphatases in the pathogenic fungus Cryptococcus neoformans. *Nat Commun* **11**, 4212 (2020).
27. Ngamskulrungroj, P. et al. The trehalose synthesis pathway is an integral part of the virulence composite for Cryptococcus gattii. *Infect Immun* **77**, 4584-96 (2009).
28. Santiago-Tirado, F.H., Peng, T., Yang, M., Hang, H.C. & Doering, T.L. A Single Protein S-acyl Transferase Acts through Diverse Substrates to Determine Cryptococcal Morphology, Stress Tolerance, and Pathogenic Outcome. *PLoS Pathog* **11**, e1004908 (2015).
29. Cottrell, T.R., Griffith, C.L., Liu, H., Nenninger, A.A. & Doering, T.L. The pathogenic fungus Cryptococcus neoformans expresses two functional GDP-mannose transporters with distinct expression patterns and roles in capsule synthesis. *Eukaryot Cell* **6**, 776-85 (2007).
30. Hicks, J.K., Bahn, Y.S. & Heitman, J. Pde1 phosphodiesterase modulates cyclic AMP levels through a protein kinase A-mediated negative feedback loop in Cryptococcus neoformans. *Eukaryot Cell* **4**, 1971-81 (2005).
31. Ljungdahl, P.O. & Daignan-Fornier, B. Regulation of amino acid, nucleotide, and phosphate metabolism in Saccharomyces cerevisiae. *Genetics* **190**, 885-929 (2012).
32. Chen, T. & Guestrin, C. in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 785–794 (Association for Computing Machinery, San Francisco, California, USA, 2016).
33. Lundberg, S.M., Erion, G.G. & Lee, S.-I. Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:1802.03888* (2018).
34. Liu, O.W. et al. Systematic genetic analysis of virulence in the human fungal pathogen Cryptococcus neoformans. *Cell* **135**, 174-88 (2008).
35. Gandolfo, L.C. & Speed, T.P. RLE plots: Visualizing unwanted variation in high dimensional data. *PLoS One* **13**, e0191629 (2018).
36. Gish, S.R. et al. Computational Analysis Reveals a Key Regulator of Cryptococcal Virulence and Determinant of Host Response. *MBio* **7** (2016).
37. Zaragoza, O. & Nielsen, K. Titan cells in Cryptococcus neoformans: cells with a giant impact. *Curr Opin Microbiol* **16**, 409-13 (2013).
38. Saidykhan, L. et al. An in vitro method for inducing titan cells reveals novel features of yeast-to-titan switching in the human fungal pathogen Cryptococcus gattii. *PLoS Pathog* **18**, e1010321 (2022).

39. Bahn, Y.S., Kojima, K., Cox, G.M. & Heitman, J. Specialization of the HOG pathway and its impact on differentiation and virulence of Cryptococcus neoformans. *Mol Biol Cell* **16**, 2285-300 (2005).
40. Xue, C., Bahn, Y.S., Cox, G.M. & Heitman, J. G protein-coupled receptor Gpr4 senses amino acids and activates the cAMP-PKA pathway in Cryptococcus neoformans. *Molecular biology of the cell* **17**, 667-79 (2006).
41. Steyfkens, F., Zhang, Z., Van Zeebroeck, G. & Thevelein, J.M. Multiple Transceptors for Macro- and Micro-Nutrients Control Diverse Cellular Properties Through the PKA Pathway in Yeast: A Paradigm for the Rapidly Expanding World of Eukaryotic Nutrient Transceptors Up to Those in Human Cells. *Front Pharmacol* **9**, 191 (2018).
42. Samyn, D.R. et al. Mutational analysis of putative phosphate- and proton-binding sites in the Saccharomyces cerevisiae Pho84 phosphate:H(+) transceptor and its effect on signalling to the PKA and PHO pathways. *Biochem J* **445**, 413-22 (2012).
43. Hu, G. et al. Cryptococcus neoformans requires the ESCRT protein Vps23 for iron acquisition from heme, for capsule formation, and for virulence. *Infect Immun* **81**, 292-302 (2013).
44. Lee, S.K., Yu, S.L., Prakash, L. & Prakash, S. Requirement of yeast RAD2, a homolog of human XPG gene, for efficient RNA polymerase II transcription. implications for Cockayne syndrome. *Cell* **109**, 823-34 (2002).

Data availability and access

All RNA-Seq data (reads and the count tables we used for analysis) are available from NCBI GEO under accession numbers GSE226255, GSE226637, or GSE226651.

Supplemental Data files

After uncompressing “File S1 Capsule Size Data.zip”, the resulting folder contains four files. Each file contains capsule and cell size measurements for individual annotated cells and descriptions of genotypes and growth conditions. The four files are:

- 42 Combinations of Growth Conditions.csv
- cAMP Titration WT.csv
- PKR1 PDE1 deletions.csv
- Deletion Mutants from Tables 1 and 2.csv

Supplemental Figures

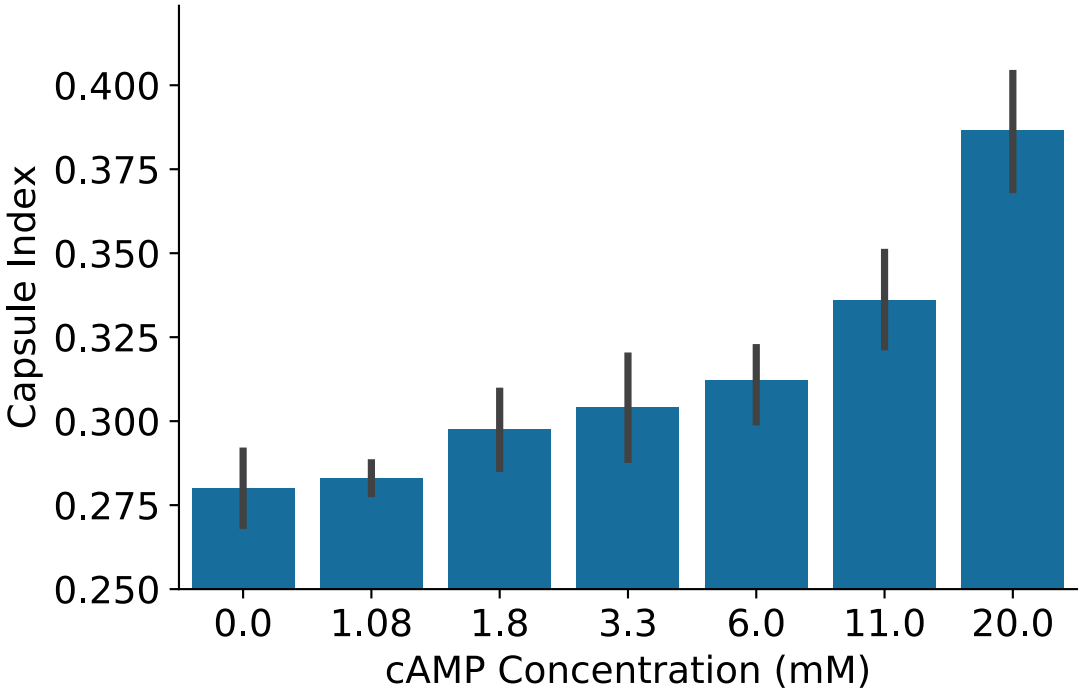

Fig. S1. Same as main text Figure 3, except that capsule index is shown instead of capsule width. Capsule index is capsule width divided by (cell radius + capsule width), i.e. capsule width as a fraction of total radius.

CNAG	Name	Ortholog	Description	$\chi 2$	Pred.	Result Inducing	Result: Almost
00012		YMR315W	hypothetical	211	Hypo	No effect	
00045		SKI2	hypothetical	202	Hypo	No effect	
00146	PBY101	PBY1	5'/3'-nucleotidase SurE	222	Hypo	No effect	No effect
00233		PAM18	chaperone protein	216	Hyper	No mutant	No mutant
00328		RAD2	DNA excision repair protein ERCC-5	212	Hypo	-0.58 um	
00576		YGR031W	mitochondrial protein	214	Hyper		No effect
00586		VPS41	hypothetical	230	Hyper		No effect
01119		PTR2	POT family proton-dependent oligopeptide transporter	215	Hyper	No mutant	No mutant
01230	MP98	CDA2	chitin deacetylase 2	268	Hyper		No effect
01359		RPB7	DNA-directed RNA polymerase II subunit RPB7	218	Hyper	No mutant	No mutant
01430		RPC53	hypothetical	209	Hypo	No mutant	No mutant
01768		MRP13	hypothetical	223	Hyper		No effect
02190		CMC2	hypothetical	219	Hyper	No effect	No effect
02455		HNM1	choline transporter	211	Hyper	No effect	Hyper
02539		DIP5	AAT family amino acid transporter	303	Hyper		No effect
02540		TNA1	hypothetical	257	Hyper		No effect
03151	CSN4	NUT1	COP9 signalosome complex subunit 4	211	Hypo	No effect	No effect
03194		PYK2	Saccharopine dehydrogenase	216	Hyper		No effect
03434		BOI2	solute carrier family 45, member 1/2/4	252	Hypo	No effect	
03985	GRX5	GRX5	Grx4 family monothiol glutaredoxin	213	Hyper	No mutant	No mutant
04280		PPX1	hypothetical	203	Hypo	No effect	
04749		COX15	hypothetical	205	Hyper		No effect
04887		TGL4	hypothetical	207	Hyper		No effect
05009		AUR1	Integral membrane protein	211	Hyper		No effect
05100		ENT3	ENTH domain-containing protein c	221	Hypo	No mutant	No mutant
05423		SET2	SET domain-containing protein	250	Hyper		No effect
05517		CHD1	solute carrier family 45, member 1/2/4	201	Hypo	No mutant	No mutant
05812		KHA1	potassium:hydrogen antiporter	200	Hypo	No effect	
06000		MYO5	glycoprotein	208	Hypo	No effect	
06265		TNA1	high-affinity nicotinic acid transporter	201	Hypo	No effect	
06969	HAD7	YER134C	magnesium-dependent phosphatase-1	209	Hyper		No effect
07314		GAS1	hypothetical	204	Hypo	No mutant	No mutant
07975			hypothetical	224	Hyper	No mutant	No mutant

Table S1 Genes the chi-squared method predicts to be involved in capsule development by. Blue: Mutant's average capsule width differs from wild type by < 0.4 μm. Teal: We did not have access to the appropriate mutant strain. Yellow: Mutant's average capsule width differs from wild type by >= 0.4 μm.

CNAG	Name	Ortholog	Description	Round	SHAP in I	SHAP in AI	Pred.	Result Inducing	Result: Almost
03078	NPP1	NPP1	type I phosphodiesterase/nucleotide	1	-0.3		Hypo	No mutant	No mutant
01119		PTR2	POT family proton-dependent oligopeptide	1		0.18	Hyper	No mutant	No mutant
02005		YBR190W	hypothetical	1	-0.18		Hypo	No effect	
03045		MVP 1	hypothetical	1		0.18	Hyper		No effect
04877		TCO89	hypothetical	1	-0.18		Hypo	No effect	
02545	IPP1	IPP1	Inorganic pyrophosphatase	2	-0.18		Hypo	No mutant	No mutant
05050	GCS2	GCS1	ADP-ribosylation factor GTPase-activating	2	-0.18		Hypo	No mutant	No mutant
01088		BNI1	hypothetical	1	-0.14		Hypo	No effect	
01993		HAA1	hypothetical	1	-0.14		Hypo	No effect	
06241	CFO1	FET5	acidic laccase	1	-0.14		Hypo	No effect	No effect
06242	CFT1	FTR1	Cryptococcus ferric transporter 1	1	-0.14		Hypo	No effect	
07870		MGM1	hypothetical	2		0.12	Hyper		No effect
03493		PMA1	1-alkyl-2-acetylglycerophosphocholine	1	-0.11		Hypo	No effect	
06050	UGE2	GAL10	UDP-glucose 4-epimerase	2		0.10	Hyper	No mutant	No mutant
02539		DIP5	AAT family amino acid transporter	1		0.09	Hyper	No effect	No effect
02540		TNA1	hypothetical	1		0.09	Hyper		No effect
00368		VPS53	vacuolar-sorting protein 53 long isoform	1	-0.08		Hypo	-0.44 um	
06509		SMC6	hypothetical	2	-0.08		Hypo	No effect	
06567		QRI7	hypothetical	2	-0.08		Hypo	No effect	
04794		TPO3	spermine transporter	2		0.07	Hyper		No effect
04887		TGL4	hypothetical	2		0.07	Hyper		No effect
04103		ICE2	hypothetical	2		0.05	Hyper		No effect
02478		GCY1	Glycerol dehydrogenase	2	-0.05		Hypo	No effect	
00871	CLR3	HAC1	hypothetical (with bZIP DNA binding	2	-0.05		Hypo	No effect	No effect
01230	MP98	CDA2	chitin deacetylase 2	1		0.05	Hyper		No effect
03007		ANT1	hypothetical	2		0.05	Hyper		No effect

Table S2 Genes predicted to be involved in capsule development by the SHAP method. Teal: We did not have access to the required mutant strain. Blue: Mutant's average capsule width differs from wild-type by < 0.4 um. Yellow: Mutant's average capsule width differs from wild-type by >= 0.4 um.

Supplemental Methods

Cell growth

Strains were removed from -80C to YPD agar, grown at 30 for 2 days, and a single colony was inoculated into 4 mL yeast peptone dextrose (YPD) and grown overnight at 30° shaking at 230 rpm. One day before the experiment, 1 mL of the ON culture was inoculated into 100 mL YPD and grown as above. At the beginning of the experiment, this culture was spun down, washed in 25 mL room temperature (RT) PBS, and resuspended in 10 mL PBS. 2 x 10^8 cells were then inoculated into each of four T75 flasks (MIDSCI TP90076) containing 20 mls preconditioned medium (described below) and incubated under the conditions specified for the experiment.

A single flask was harvested for each time point (30, 90, 180, and 1440 min after inoculation), by transferring 18.5 mL of culture to a tube on ice containing 2 mL RNA Stop solution (5% tris-saturated phenol in ethanol).The cells were then sedimented (3000g, 5 min, 4C), the supernatant fraction decanted, and the pellet frozen in liquid nitrogen and stored at -80C.

For imaging, 1-ml aliquots of cells were collected prior to inoculation of flasks and at 1440 min, sedimented as above, resuspended in PBS, and counted with a Nexcelom BioScience Cellometer Auto M10. 10^6 cells were added to 500 μL PBS containing 3.7% formaldehyde and the cells were fixed by end-over-end rotation at RT for at least 30 minutes. Fixed cells were stored at 4C.

Environmental signals and growth conditions

All flasks were prepared and then preconditioned for 24 hours at the designated temperature and pH concentration before use. Cultures were then inoculated as above and incubated without rotation. DMEM (Sigma D6429) and RPMI (Sigma R8758) were used as purchased except in “buffer” conditions, where sterile 1M HEPES pH 7.0 was added to a final concentration of 25 mM.

Microscopy and manual image annotation

Fixed cells were sedimented as above and resuspended in 250 μL PBS. 10 μL of this suspension were mixed with 4 μL of India ink and imaged by brightfield microscopy at 63X magnification. Images were manually annotated using a custom interface written in Mathematica / Wolfram Language (available on request). Fifteen fields were annotated for each replicate of each combination, yielding an average of 108 annotated cells per replicate (SD 56).

Cell size censoring and capsule size adjustment for background darkness

For WT cells across various growth conditions, 21,530 cells were annotated. 4.5% of them had cell radius (excluding capsule) < 1.5 um and 1.7% had cell radius > 4.0 um. Most of these came from a few growth conditions: For small cells, most were from RPMI 30° no CO_2 no cAMP, YPD at 30° or 37° with no CO_2 no cAMP, or RPMI 30° 5% CO_2 no cAMP; for large cells, most came from conditions with cAMP but not CO_2: RPMI 30, RPMI 37, or YPD 37. Since cell size and capsule size are correlated and our primary interest is

the latter, we removed the cells with radii outside the [1.5 um – 4.0 um] interval before further analysis (however, they have not been removed in the datasets provided as a supplementary file). This had no effect on the conclusions.

Analysis of capsule thickness in gene deletion mutants

Effects of deletion mutants on capsule thickness were evaluated using a Linear Mixed Models framework with individual cells' thicknesses as the datapoints. The model had fixed effects for intercept and genotype; each specific biological sample was used as a grouping variable with a random intercept for each group.

RNA Isolation

Frozen cell pellets were thawed on ice and 875 ml TRIzol (Ambion 15-596-018) was added. Cells were then lysed by beating with 750 μl 0.5mm silica-zirconia beads for 3 min in a Minibeadbeater (Biospec Products) and RNA was isolated following the manufacturer's protocol. Residual DNA was removed with the Turbo DNA-free Kit (Life Technologies AM1907) and polyA+ RNA isolated from ~1 μg of total RNA using the NEBNext Poly(A) mRNA Magnetic Isolation Module.

Library preparation and sequencing

Libraries were constructed using the NEBNext Ultra Directional RNA Library Prep Kit from Illumina and samples were pooled at 10 nM. The pools were sequenced on a NextSeq 500 using the High75v2 kit as a 1x75 with 7 bp Index1 and 6 bp Index2 with 1% PhiX. Fastq files were demultiplexed using the Illumina bcl2fastq2 allowing 1 bp mismatch.

RNA-Seq Computational Pipeline and Analysis

Reads were demultiplexed into fastq files using Illumina bcl2fastq2 allowing 1 bp mismatch. Resulting fastqs were aligned with novoalign (version 4.03.02) and quantified with HTSeq (0.9.1) using the FungiDB KN99α genome sequence. Genome annotations were augmented with noncoding RNA regions lifted over from the current H99 genome. The quantification for downstream analysis was calculated over gene feature exons, but CDS quantification was also performed for QC purposes. Samtools (version 1.12) was used to convert the SAM output from novoalign to BAM and index Bam files. Novoalign and HTSeq logs were collated by MultiQC (version 1.2).

Custom scripts were used to verify the strain represented in each fastq by calculating CDS coverage over putatively deleted genes and marker genes, ensuring that the former were absent and the latter present. We found that CDS deletion frequently resulted in increased transcription of UTRs, so full-transcript counts are not appropriate for validating CDS deletions. RNA-Seq samples that passed strain validation were subjected to two phases of QC. In Phase 1, files were labeled as 'passing' if they contained at least 10^6 reads aligned to protein coding regions and less than 7% of all reads failed to align.

In Phase 2, we evaluated replicate agreement using the metric Regularized Log Expression (RLE) [32]. First, we used DESeq2 (version 1.34.0) to estimate the effect of the library date (the known batch effect). We removed the batch effect using the DESeq coefficients for the library dates such that the data was standardized to a single date, resulting in adjusted expression levels on a $\log_2$ normalized scale. To compute the RLE value for a given gene, the median log expression level of that gene, across all samples

in a replicate set, was subtracted from the log expression level of the gene in each sample. For each sample, we then calculated the interquartile range of the distribution of these RLE values across genes. If the interquartile range of RLEs of a given sample was greater than 1, indicating that more than half of genes deviated from their respective medians by a factor of 2 in adjusted count, the sample was considered an outlier and failed for replicate agreement. 22 samples failed replicate agreement and were discarded.

Selection of potential capsule-associated genes: Chi-squared method

The following procedure was carried out separately for samples at each time point. To calculate the χ^2 statistic for each gene, we first classified all samples from all conditions as either enlarged or unenlarged. To be enlarged, a sample needed a mean capsule width of greater than 1.13 μm, or 3 standard deviations off the mean (across samples) of the means (across cells within a sample) of all samples grown in YPD, regardless of other signals provided. This criterion was chosen because there was a big gap in capsule thickness between all YPD samples and all samples in conditions that we expected to be enlarged. Samples were also classified by high or low expression of each gene, yielding a 2x2 contingency table, in which samples were classified as enlarged or unenlarged and high versus low expressors of the gene. For each contingency table, we calculated the largest χ^2 statistic. For each gene, we tried all possible thresholds of high versus low expression and chose the threshold that yielded the largest χ^2 statistic. These statistics were only used to characterize the degree of association between the expression of a gene in a sample and that sample's induction status. Since the statistics were not used for statistical hypothesis testing, the issue of multiple hypothesis testing does not apply.

Genes were ranked for likelihood of the corresponding deletion strain exhibiting a capsule phenotype by the maximum of their χ^2 values, across all time points. All genes with max χ^2 statistics above 200 were selected for testing. For genes that were positively correlated with capsule size, the gene deletion mutant was tested in inducing conditions (RPMI, 37°C, 5% CO_2, no cAMP) with the prediction that deletion would reduce capsule size compared to WT grown in the same conditions. For negatively correlated genes, the gene deletion mutant was tested in almost-inducing conditions (RPMI, 37°, room air, no cAMP, no buffer) and non-inducing conditions (YPD, 37°, 5% CO_2, with or without cAMP, no buffer).

Forming metagenes as features for machine learning

From the original gene expression data matrix, we sought to decrease the number of features by filtering and combining genes. We first removed low variance genes by filtering out genes whose expression in >= 95% samples was within one $\log_2$ of their mean, indicating that less than 5% of samples showed substantial changes. We next combined highly correlated genes into metagenes (correlation threshold > 0.8). The "expression level" of a metagene was then set to the mean of the expression levels of its constituent genes. We repeated this process to combine genes into metagenes until we had combined all highly correlated genes into metagenes.

Selection of genes to test for capsule phenotypes: XGBoost method

Using the gene expression data matrix of metagenes $(p' \times n)$, we trained a model to predict capsule size at each time point. Features were normalized prior to training. The

final parameters for the model (max_depth=4, n_estimators=50) were selected by using 10-fold cross validation. We investigated the use of the regularization parameter, γ, and did not find significant improvement in the model's predictive performance.

We calculated the SHAP value for each metagene in each sample. When a metagene was created from more than one gene's expression data, the metagene's SHAP value was attributed to all the contributing genes. We selected the maximum SHAP value across timepoints to associate with each gene and selected genes with absolute mean SHAP value > 0.05 in those conditions for testing. Testing of the corresponding deletion mutant, when available, was then performed as described for the χ^2 method.

If two features are correlated with each other, ML algorithms may select one and discard the other, even if their predictive values are very similar. Therefore, we removed the selected genes (or their metagenes) from the feature set, retrained XGBoost, and repeated the selection process. These genes have 2 in the "Round" column of Table 2.