# Tur1 regulates alternative TSS usage in *Cryptococcus*

Thi Tuong Vi DANG[1], Corinne MAUFRAIS[2], Jessie COLIN[1,3], Frédérique MOYRAND[1], Isabelle MOUYNA[1], Jean-Yves COPPEE[1], Chinaemerem U. ONYISHI[5], Joanna LIPECKA[4], Ida Chiara GUERRERA[4], Robin C. MAY[5] and Guilhem JANBON[1*]

1. Université Paris Cité, Institut Pasteur, Unité Biologie des ARN des Pathogènes Fongiques, Département de Mycologie, F-75015, Paris, France

2. Université Paris Cité, Institut Pasteur, HUB Bioinformatique et Biostatistique, C3BI, USR 3756 IP CNRS, F-75015, Paris, France

3. Ecole Pratique des Hautes Etudes, PSL Research University, F-75014, Paris, France

4. Université Paris Cité, SFR Necker INSERM US24/CNRS UAR3633, Proteomics Platform, F-75015, Paris, France

5. Institute of Microbiology and Infection and School of Biosciences, University of Birmingham, Birmingham B15 2TT, United Kingdom

* To whom correspondence should be addressed. Email: Guilhem.Janbon@pasteur.fr

## Abstract

Alternative transcription start site (TSS) usage regulation has been identified as a major means of gene expression regulation in metazoans. However, in fungi, its impact remains elusive as its study has thus far been restricted to model yeasts. Here, we first re-analysed TSS-seq data to define genuine TSS clusters in two species of pathogenic *Cryptococcu*s. We identified two types of TSS clusters associated with specific DNA sequence motifs. Our analysis also revealed that alternative TSS usage regulation in response to environmental cues is widespread in *Cryptococcus*, altering gene expression and protein targeting. Importantly, we performed a forward genetic screen to identify a unique transcription factor (TF) named Tur1, which regulates alternative TSS (altTSS) usage genome-wide when cells switch from exponential phase to stationary phase. Tur1 has been previously shown to be essential for virulence in *C. neoformans*. We demonstrated here that a *tur1Δ* mutant strain is more sensitive to superoxide stress and phagocytosed more efficiently by macrophages than the wild-type (WT) strain.

In eukaryotes, gene transcription begins at the core promoter by the assembly of a pre-initiation complex comprised of general transcription factors that recruit the DNA-dependent RNA polymerase II and melt DNA to create the transcription bubble (1). In recent years, different RNA sequencing strategies have been used in metazoans to define two types of core promoters characterized by specific chromatin structure, DNA sequence and distribution of TSS (2). These analyses have shown that TSS are by nature heterogeneous and organized in clusters. These TSS clusters can be sharp or broad, sharp ones being associated with TATA-box genes (1,3). These studies also revealed that transcription can be initiated from multiple TSS clusters in most genes with a very dynamic pattern of usage (4-8) which can be specific to different cell types and development stages (6,9-11).

In fungi, although TSS sequencing data have been produced in several species (12-14), the analysis of TSS structure and usage is limited to the two model yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* (15-18). Nevertheless, these studies also revealed a significant number of alternative TSS (altTSS) usage patterns, with some being specific to growth conditions or meiotic stages (12,19) and others revealed by mutation of genes encoding chromatin modifiers or remodelers (20-22).

*Cryptococcus neoformans* is a pathogenic basidiomycete yeast which is responsible for 180,000 deaths every year worldwide (23). In addition to its ability to grow at 37°C, its main virulence factors are a polysaccharide capsule, the production of melanin and its ability to replicate in macrophages (24,25). In recent years, we produced several sets of RNA-seq data which were used to produce detailed coding gene annotation of the genomes of several pathogenic *Cryptococcus* species (26-28). Our analysis also revealed that nearly all the genes contain several short introns, which are essential for gene expression (29,30). As expected, alternative splicing is prominent in *Cryptococcus*, although its impact on proteome structure is limited primarily to regulating gene expression (27). More recently, we produced TSS-seq and

3UTR-seq data from *C. neoformans* and its sibling species *Cryptococcus deneoformans* grown under four conditions (*i.e.* exponential phase and stationary phase at 30 and 37°C) (14). We used this dataset to re-annotate the transcript leader (TL) and 3'-UTR sequences in these species. Our analysis revealed that, in contrast to *S. cerevisiae*, *Cryptococcus* TL sequences frequently contain upstream Open Reading Frames (uORFs). These yeasts use the strength of a Kozac-like consensus to determine translation start codon usage, thus regulating both gene expression and protein localization (14). This analysis also revealed thousands of additional TSS clusters associated with coding genes, suggesting that alternative TSS (altTSS) usage might be widespread in *Cryptococcus* (14). This hypothesis was supported by several recent studies reporting gene specific examples of altTSS usage regulation in these yeasts. For instance, Pum1, an RNA-binding protein, is known to positively regulate the expression of *ZNF2*, a master regulator of filamentation and virulence in *C. neoformans* (31). Under filamentation-inducing conditions, the TF Znf2 favors *PUM1* transcription from a downstream TSS, which excludes from the transcript the Pum1-binding site that is normally found in the TL. This shorter form being immune to the negative autoregulation, this leads to the accumulation of the Pum1 protein which in turn activates *ZNF2* expression (31). Moreover, when exposed to UV light, *C. neoformans* switches off TSS at the *UVE1* gene and uses an upstream altTSS. This promotes the transcription of a longer mRNA coding for the mitochondrial isoform of the DNA damage repair endonuclease Uve1, thus protecting the mitochondrial genome from potentially lethal UV-induced DNA damage (32). Finally, a recent study reports that, under copper-limited conditions, *C. neoformans* cells promote the usage of a downstream altTSS at both *SOD1* and *SOD2* superoxide dismutase genes, thus regulating transcript stability and protein subcellular localization of these proteins, respectively (33).

In the present study, we re-analyzed these *Cryptococcus* TSS-seq data to first characterize the structure of a genuine TSS cluster in these yeasts. We then described alternative

TSS usage and demonstrated that it represents a major means to regulate transcriptome and proteome structure in these fungi. More importantly, we screened a transcription factor mutant collection to identify genes regulating altTSS usage in *C. neoformans*. We showed that the transcription factor Tur1 is necessary for genome wide altTSS usage regulation during the exponential to stationary phase transition. Finally, we showed that Tur1 regulates superoxide stress resistance and interaction with macrophage linking altTSS usage and virulence in this major fungal pathogen.

# Material and Methods

## DNA and RNA sequencing

Each *Cryptococcus* cell preparation was spiked in with one tenth (OD/OD) of *S. cerevisiae* strain FY834 (34) and cells were grown in YPD at 30°C in exponential phase. Cells were washed, snap frozen and used to prepare RNA and total DNA samples as previously described (35,36). For RNA-seq, strand-specific, paired-end cDNA libraries were prepared from 10 µg of total RNA by polyA selection using the TruSeq Stranded mRNA kit (Illumina) according to manufacturer's instructions. cDNA fragments of ~400 bp were purified from each library and confirmed for quality by Bioanalyzer (Agilent). DNA-seq libraries were prepared using the kit TruSeq DNA PCR-Free (Illumina). Then, 100 bases were sequenced from both ends using an Illumina HiSeq2500 instrument according to the manufacturer's instructions (Illumina). Trimmed reads were mapped to the *Cryptococcus* and *S. cerevisiae* genomes using Bowtie2 (37) and Tophat2 (38) as previously described (14). RNA-seq reads counts were then normalized considering the ratios of DNA and RNA aligned to each genomes using the strategy previously described by Malabat and colleagues (39). Differential gene expression analysis was performed using the DESeq2 package (40) using default settings but ommiting the associated read count normalization step.

TSS-seq libraries preparations were performed as previously described (14,39) replacing the TAP enzyme by the Cap-Clip™ Acid Pyrophosphatase (Cellscript). Sequencing was performed on Illumina HiSeq2500 to obtain fifty base single end reads. Data processing including reads trimming, clipping, and mapping were performed as previously described (14). TSS reads were first clustered using arbitrarily chosen maximum intra-cluster distance (d = 50 nt) between sites. We then looked for the optimal "d" value to define clusters, reasoning that an excessively small value would separate reads belonging to a single cluster into several ones, whereas an

overestimated “d” would merge reads from different clusters into a single one. We thus tested increasing values of “d” (from d =1 to d =50) reasoning that the size of the clusters thus defined will also increase until their actual boundaries would be reached. We here considered only the genes with only one TSS cluster in their TL sequence using the exponential 30°C dataset and d =50 defined. We then plotted the “d” values with the percentage “P” of clusters reaching their maximum cluster size (Supplementary Figure S1A. We then used the nls function in R to modelized this relationship in the following equation where P = 100/ (1+exp(-0.1655*d)^4.2957, (Supplementary Figure S1B). With the model, the correlation of d and P reached its peak at d=17 (Supplementary Figure S1B). We thus choose this value for our analysis.

**Calculating thermodynamic entropy-related shape index (SI) of TSS cluster**

The SI index was calculated as previously demonstrated by Hoskin and colleagues (41) as below:

$$SI = 2 + \sum_{i}^{L} p_i \log_2 p_i \,,$$

where p is the probability of observing a TSS at base position i within the cluster, and L is the set of positions that possess mapped reads in the cluster.

**TSS cluster subpopulation detection and cluster shape classification**

We considered only genes with only one TSS cluster in their TL and kept only clusters with both Size and SI between 2.5% and 97.5% quantiles, as evaluated using the data obtained in exponential phase 30°C (719 genes in exponential phase 30°C, 691 in stationary phase 30°C, 635 genes in exponential 37°C, 586 genes in stationary phase 37°C). To detect if there is more than one population of TSS clusters, the first principal component (PC1) of Size and SI was used as input for the function “emtest.norm” of R package “‘MixtureInf” (42) with the default value of the finite mixture model under null hypothesis m0=1. The output of function

"emtest.norm" includes the percentage of each sub-population together with the detected models with theirs mean and standard deviation. This information was then supplied to the argument "theta" of the function "plotmix.norm" to plot histogram of the observations and the fitted density of a mixture of models. TSS clusters were then classified into two groups based on the mean value of PC1 plus and minus one standard deviation of each group as returned by the function "emtest.norm".

**Motif enrichment discovery in core promoter region**

To find motifs enriched in promoter regions, the genomic sequences from -100 to +30 of the major TSS peaks were extracted and used as input in MEME Suite software 5.4.1 (43) (*de novo* pattern discovery mode). Default settings were applied except these following parameters: number of motifs: 6, minimum width: 6, maximum width: 10, and search given strand only. The six motifs found were then scanned in each group of Sharp and Broad clusters in each condition with Fimo mode (threshold $p$-value $\leq 0.0005$) to examine their number and positions in each dataset.

**Definition of annotated TSS cluster and alternative TSS cluster**

We defined a TSS cluster as annotated TSS clusters if it satisfies all of these three criteria: it is the most upstream cluster within a gene, the distance between the 5' boundary of the cluster and the annotated 5' extremity of the gene does not exceed 50 bp, and the 3' border of the cluster is at least 20 bp upstream of the annotated ATG. Every other TSS clusters were considered as alternative TSS clusters.

**altTSS usage analysis**

The "expression level" of each TSS cluster was calculated for each replicate and normalized by the total number of reads associated with gene containing the cluster. We did not consider the clusters that never contribute more than 10% total reads associated with a given gene in any

replicate. We then performed four comparisons: EXPO 30 vs STAT 30, EXPO 37 vs STAT 37, EXPO 30 vs EXPO 37, and STAT 30 vs STAT 37.

**N- terminomic experiments**

*Lysis and protein extraction*

Cells were lysed by sonication in lysis buffer containing 0.5% SDS, 100 mM HEPES pH 8, 10 mM DTT, 10 mM EDTA, supplemented with protease inhibitors (cOmplete$^{TM}$ Mini EDTA-free, Roche). After sonication, the lysate was incubated for 20 min at RT with 1 µL of benzoase and centrifuged for 10 min at 12,000 *g* at 4°C. The supernatant was reduced and alkylated using 20 mM TCEP and 10 mM NEM (N-ethylmaleimide) respectively, for 30 min at RT in the dark. Protein concentration was determined with DC assay (BioRad) and 200 µg of each sample were further processed.

*TAILS protocol (Terminal amine isotopic labeling of substrates)*

TAILS experiment was based on reference protocol (44) with some modifications. Prior to labeling, the samples were first precipitated using acetone-methanol method to remove all primary amines. Briefly, 8 sample volumes of ice-cold acetone and 1 volume of ice-cold methanol were added, samples were let to precipitate overnight at -80°C, centrifuged for 30 min at 14,000 *g* at 4°C and the pellet washed twice with ice-cold methanol. The dried pellet was resuspended in 1% SDS, 100 mM TEAB pH 8 and incubated for 5 min at 95°C. pH was adjusted to 7.5. Dimethylation labeling was carried out by adding 40 mM formaldehyde and 20 mM sodium cyanoborohydride ($NaBH_3CN$) and incubating overnight at 37°C. The reaction was quenched by addition of 500 mM of Tris-HCl pH 6.8 and incubation for 1h at 37°C.

<u>Tryptic digestion</u>

S-Trap™ mini spin column (Protifi, Hutington, USA) digestion was performed on dimethylated samples, according to manufacturer's instructions. Briefly, SDS concentration was first adjusted to 5% and aqueous phosphoric acid was then added to a final concentration of 2.5% following by the addition of S-Trap binding buffer (90% aqueous methanol, 100 mM TEAB, pH 7.1). Mixtures were loaded on S-Trap columns by 30 sec centrifugation at 4,000 *g*. Six washes were performed before adding the trypsin (Promega) at 1/20 ratio and incubation for 2 h at 47°C. After elution, 10 µg of peptides were kept for MS/MS analysis and the remaining samples were vacuum dried.

Negative selection by HPG-ALDII polymer

Dried peptides were resuspended in 100 µL of 100 mM HEPES pH 7. The HPG-ALDII polymer was added to samples at a polymer/peptide ratio 5:1 (w/w), followed by addition of 20 mM $NaBH_3CN$. Samples were then incubated overnight at 37°C. 100 mM Tris-HCl pH 7 was added and samples were incubated for 30 min at 37°C before loading by centrifuging for 10 min at 12,000 *g* on 10 kDa Amicon (Millipore) centrifugation device previously washed with 100 mM NaOH and water. The flow-through was collected and additional washing step was performed. The filters were then rinsed with water, polymer discarded and filters spined-down in upside-down position for 3 min at 5,000 *g*. Collected fractions were pooled with flowthroughs and vacuum-dried.

MS/MS analysis

Peptides were resuspended in 10% ACN, 0.1% TFA in HPLC-grade water prior to MS analysis. For each run, 1/10 of peptide solution was injected in a nanoRSLC-Q Exactive PLUS (RSLC Ultimate 3000) (Thermo Scientific,Waltham MA, USA). Peptides were loaded onto a µ-precolumn (Acclaim PepMap 100 C18, cartridge, 300 µm i.d.×5 mm, 5 µm) (Thermo Scientific), and were separated on a 50 cm reversed-phase liquid chromatographic column

(0.075 mm ID, Acclaim PepMap 100, C18, 2 µm) (Thermo Scientific). Chromatography solvents were (A) 0.1% formic acid in water, and (B) 80% acetonitrile, 0.08% formic acid. Peptides were eluted from the column with the following gradient 5% to 40% B (120 minutes), 40% to 80% (1 minute). At 121 minutes, the gradient stayed at 80% for 5 minutes and, at 126 minutes, it returned to 5% to re-equilibrate the column for 20 minutes before the next injection. One blank was run between each replicate to prevent sample carryover. Peptides eluting from the column were analyzed by data dependent MS/MS, using top-10 acquisition method. Peptides were fragmented using higher-energy collisional dissociation (HCD). Briefly, the instrument settings were as follows: resolution was set to 70,000 for MS scans and 17,500 for the data dependent MS/MS scans to increase speed. The MS AGC target was set to 3.106 counts with maximum injection time set to 200 ms, while MS/MS AGC target was set to 1.105 with maximum injection time set to 120 ms. The MS scan range was from 400 to 2,000 m/z. Dynamic exclusion was set to 30 seconds duration.

Data analysis

The MS files were processed with the MaxQuant software version 1.5.8.3 and searched with Andromeda search engine against the annotated *C. neofomans* database (14) but also against a database containing all alternative proteins potentially resulting for alternative TSS usage. To search parent mass and fragment ions, we set a mass deviation of 3 ppm and 20 ppm respectively. The minimum peptide length was set to 7 amino acids and strict specificity for trypsin cleavage was required, allowing up to two missed cleavage sites. NEM alkylation (Cys) was set as fixed modification, whereas oxidation (Met) and N-term acetylation, N-term Dimethylation (any N-term), Dimethylation (Lys) were set as variable modifications. The false discovery rates (FDRs) at the protein and peptide level were set to 1%. Scores were calculated in MaxQuant as described previously (45). The reverse and common contaminants hits were

removed from MaxQuant output. The output files "peptides" and "modificationSpecificPeptides" were combined and further analyzed on excel.

**Transcription factor mutant library screening**

Total RNA (5 μg) extracted from each of the *C. neoformans* 150 TF mutant strains (46) was subjected to DNAse I treatment (Roche) to eliminate contaminating genomic DNA. A total of 1 μg of the DNase I-treated RNA was then used for reverse-transcription (RT) with Maxima First Strand cDNA Synthesis Kit (Thermo Scientific™). qPCR experiments using CFX96 Touch Real-Time PCR Detection System (Bio-Rad) with SsoAdvanced Universal SYBR Green Supermix (Bio-Rad) according to manufacturer's instructions. The primer pairs used were specific of either long (in the region located between TSS1 and TSS2) or both isoforms (downstream of TSS2) of *PKP1* or *ACT1* (Supplementary Table S1). The number of short isoforms was calculated as the number of both *PKP1* RNA isoforms minus the number of long isoforms. The levels of the *PKP1* long and short isoforms were compared between WT and mutant strains. The level of expression of *ACT1* was used as a control.

**Construction of strains deleted and complemented for *TUR1* gene in *C. neoformans***

The *TUR1* gene was deleted in *C. neoformans* KN99α reference strain (26) (Supplementary Table S2) by inserting a $NEO^R$ cassette in/in place of the coding sequence by electroporation using a transient CRISPR-Cas9 expression (TRACE) system (47). Plasmids and primers to amplify Cas9 and gRNA against *TUR1* are listed in Supplementary Tables S3 and S1, respectively. For the complemented strain, we PCR amplified a DNA fragment spanning a genomic DNA region form 1kb upstream of annotated ATG to 1kb downstream of stop codon of the gene *TUR1*. This fragment was then cloned into pSDMA57 (Supplementary Table S2)

allowing the integration at the genomic "safe haven 1" locus and complementation of *TUR1* in *tur1Δ::NAT* mutant (48). For galactose-inducible complementation, the *TUR1* sequence from annotated ATG to 1kb downstream of stop codon was amplified, and then was fused at the 5' end with 2xFLAG-CBP (49) amplified from plasmid pCM189-NTAP (50). This construct was positioned under the control of the *GAL7* promoter in the plasmid pSDMA57-GAL7p-TUR1. The construct was then PCR-amplified and co-transformed into *tur1Δ* mutant strain together with a PCR-amplified DNA fragments containing the sequence of Cas9 and a PCR-amplified DNA fragment coding the gRNA targeting Safe Haven 1 as previously described (47).

**Construction of mNeon tagged *MAE102* genes, fused ATG and mutated aATG mutant strains**

For *MAE102* gene tagging, the region covering mNeonGreen-CBP-2xFlag-terminator together with NAT1 cassette was PCR amplified from pBHM2404 and fused in 5' with 1 kb upstream the STOP codon of *MAE102* (without including it) and in 3' with 1kb downstream of the STOP codon to allow recombination in the genome (see Supplementary Table S1 for primers). This cassette was co-transformed with a PCR fragment allowing production of a gRNA directed against the 3' of *MAE102* in the strain CM2049 in which the Cas9 gene is integrated at the Safe Heaven 2 (51,52) and constitutively expressed. The *mae102-MTSΔ* mutant in which the region containing a putative MTS has been deleted was constructed using a similar strategy. Here a cassette in which 1kb upstream of the annotated ATG was fused with 1kb starting at the second ATG was first constructed before being co-transformed with the NEO marker (pSDMA57 cut by PacI), these two constructs being targeted by appropriate gRNAs at the aATG of *MAE102* and at the Safe Haven 1 locus, respectively. These mutations were done on the mNeonGreen-tagged version of *MAE102*. Finally, the *mae102-M1R* mutant was constructed with the same strategy using a cassette in which the annotated ATG was replaced by a CGT. Primers, strains,

and plasmids used to construct these mutants are listed in the supplementary tables S1, S2 and S3, respectively.

**Western Blot**

The *C. neoformans tur1Δ* mutant strain expressing the fusion protein 2xFlag-CBP-Tur1 under the control of the *GAL7* promoter (53) was cultured at 30°C in media containing glucose (YPD) or galactose (YPGAL) until exponential phase or stationary phase. Total proteins were extracted from 10 $OD_{600}$ units of cells and were dissolved in 250 µL of 1X SDS-PAGE sample buffer. 10 µL of protein extract were loaded on a denaturing polyacrylamide gel before being transferred to a nylon membrane and probed with an anti-flag M2 antibody (Sigma Aldrich F1804). An anti-β-actin antibody (Sigma Aldrich A2066) was used as control.

**Northern Blot**

RNA was extracted with TRIZOL Reagent (Thermo Fisher Scientific) following the manufacturer's instructions. Total RNA (10 µg) was separated by denaturing agarose gel electrophoresis and transferred onto Hybon-N+ membrane (Sigma-Aldrich) and probed with [$^{32}$P]dCTP-radiolabelled DNA fragments. The banding pattern was quantified with a Typhoon 9200 imager and Image Quantifier 5.2 sofware (Molecular dynamics).

**Phenotypic assays**

For spot assays, cells grown in liquid YPD medium at 30 °C for 16 h were counted, serially diluted and spotted on solid YPD medium supplemented with chemical compounds purchased from Sigma Aldrich : 0.2 mM cumen hydroperoxide, 2 mM $NaNO_2$, 100 mM sodium

dithiocarbamate , 0.4 mM rotenone , 10 mM dimethyl malonate, and 0.2 µg/mL oligomycin. Cells were incubated at 30 °C for 3 days and the plates were photographed.

For disc inhibition assays, overnight cultures of indicated *C. neoformans* strains were grown overnight. The following day cells were diluted to a final $OD_{600}$ of 0.1 and 100 µL was then spread on YPD plates. Filter discs containing 20 µL of 10 mM menadione or 5 µg/mL antimycin A were positioned at the center of the Petri dish and plates were incubated at 30 °C for 3 days and photographed.

**Phagocytosis assay**

The murine macrophage J774A.1 cell line (54) was cultured in T-75 flasks (Fisher Scientific) in Dulbecco's Modified Eagle medium, low glucose (DMEM) (Sigma-Aldrich), containing 10 % live fetal bovine serum (FBS) (Sigma-Aldrich), 2 mM L-glutamine (Sigma-Aldrich), and 1 % Penicillin and Streptomycin solution (Sigma-Aldrich) at 37°C and 5 % $CO_2$. Cells were passaged when 70 % to 90 % confluent by scraping and resuspending in fresh complete DMEM at a ratio of 1:3 to 1:6.

Phagocytosis assays were performed to measure the capacity of J774 macrophages to engulf *C. neoformans* WT, *tur1Δ*, and *tur1Δ+TUR1* strains. Twenty-four hours before the start of the phagocytosis assay, $1x10^5$ J774 macrophages were seeded onto wells of a 24-well plate (Greiner Bio-One). The cells were then incubated overnight at 37°C and 5 % $CO_2$. At the same time, a colony of each strain was picked from the stock plate and resuspended in 3 mL liquid YPD broth. The culture was incubated at 25°C overnight under constant rotation (20 rpm). On the day of the assay, macrophages were activated using 150 ng/mL phorbol 12-myristate 13-acetate (PMA) (Sigma-Aldrich) for 1 hour at 37°C. PMA stimulation was performed in serum-free media to eliminate the contribution of complement proteins during phagocytosis. To prepare *C. neoformans* for infection, an overnight *C. neoformans* culture was washed two times in 1X PBS

and centrifuged at 6,500 rpm for 2.5 min. To infect macrophages with unopsonised *C. neoformans*, after the final wash, the *C. neoformans* pellet was resuspended in 1 mL PBS, counted using a hematocytometer, and fungi incubated with macrophages at a multiplicity of infection (MOI) of 10:1. The infection was allowed to take place for 2 hours at 37°C and 5% $CO_2$. Afterwards, extracellular *Cryptococcus* was washed off and wells were treated with 10 µg/mL calcofluor white (CFW) (Sigma-Aldrich) for 10 min at 37°C to distinguish between phagocytosed and extracellular *C. neoformans*. Images were acquired using the Nikon Eclipse Ti inverted microscope (Nikon) fitted with the QICAM Fast 1394 camera (Hamamatsu) under 20X magnification and using the phase contrast objective and CFW channel. Images were analyzed using the Fiji image processing software (ImageJ). Two hundred macrophages were counted and phagocytic index (PI) was quantified as the percentage of macrophages that phagocytosed at least one *Cryptococcus* cell (PI= ((number of infected macrophage/total number of macrophages counted) * 100%)).

# Results

## TSS-cluster characterisation in *Cryptococcus*

We previously used TSS-seq data obtained from cells grown under four conditions in triplicate to construct twelve GFF files defining the coordinates of several thousand TSS clusters associated with coding genes in two *Cryptococcus* species (14). To evaluate alternative TSS usage, we aimed to construct a TSS cluster reference GFF file. We first thought to merge the twelve condition-specific GFF files originally constructed. However, this simplistic strategy resulted in a very poor result. Firstly, most of the resulting merged TSS were doubtful since they were only 1 bp-wide. Secondly, at several loci the TSS cluster size and number varied between conditions. This merging operation also resulted in the definition of very large TSS (more than 100 bp wide) with probably poor biological relevance (see Supplementary Figure S2). This result suggested that the strategy previously used to define the TSS clusters (14) was not optimal and resulted in the identification of a number of doubtful TSS clusters. To better define the characteristics of genuine TSS clusters in *C. neoformans*, we reasoned that when only one TSS cluster is found upstream of the annotated ATG (aATG) of a well-expressed gene, this cluster should be genuine. We thus first considered only the 50% most expressed genes in exponential phase 30°C and for which a single TSS cluster has been originally identified upstream of the aATG (n = 761). We then performed sub-clusterization analysis (see material and methods) and characterized the structure of the resulting clusters using two parameters: size and thermodynamic entropy-related shape index (SI), which is a measure of the heterogeneity of the TSS cluster (41). Finally, we kept only the TSS clusters with both sizes and SI values comprised between the 2.5% and 97.5% quantiles for further analysis (n = 719). Plotting the size and SI of this set of newly defined TSS clusters with the MixtureInf (42) based statistical test revealed a bimodal distribution defining two categories: the Sharp clusters characterized by

a lower size and high SI values and the Broad TSS clusters having the converse characteristics (Figures 1A and 1B).

MEME Suite software 5.4.1 (43) *de novo* pattern discovery mode was used to examine the region going from -100 to +30 relative to the major position within each TSS cluster to identify six enriched motifs (Figure 1C). Notably, our analysis identified a TATA motif positioned -40 to -35 bp upstream of the TSS. We noticed that genes with a sharp TSS cluster are enriched within the TATA-box containing genes, as previously observed in some other model organisms (1) (Figure 1C). Overall, 23.5 % of the Sharp TSS clusters are associated with a TATA-box containing genes compared to only 11.7 % of the Broad TSS clusters and 14.5 % of both groups of TSS clusters ($\chi^2$ test, $p$-value <0.05). The distance between the TATA-box and the TSS in *Cryptococcus* is comparable with that reported in metazoans (28 to 33 bp) and in *S. pombe* (25 to 32 bp) (2,3) This short distance suggests that *C. neoformans* uses a classic model of transcription initiation in which the preinitiation complex (PIC) assembles on TATA-box and loads the RNAPII on a TSS located in close proximity. In contrast, *S. cerevisiae* as well as most hemiascomycetes use a scanning model of transcription initiation in which the PIC scans the sequence downstream to find TSSs which results in longer and more variable distances between the TATA-box and the TSS (40 to 120 bp) (3).

We performed the same analysis using the TSS-seq data obtained from cells grown in the three other conditions (*i.e.* Exponential phase 37°C; Stationary phase 30°C and 37°C). Even though the gene set considered was different each time, similar bimodal distribution of sizes and SI as well as motif enrichments and positions were observed, suggesting general features in *C. neoformans* (Figure S2 and Figure S3). Previous studies in other organisms indicated that regulated genes are associated with the presence of a TATA-box whereas constitutive ones tend to be devoid of TATA-box (2,5,8,41,55). To determine if this phenomenon is conserved in *C.*

*neoformans*, we used or produced spiked in RNA-seq data from *C. neoformans* cells grown in diverse conditions (*i.e.* exponential phase at 30°C in the presence of either fluconazole (15 µg/mL) or SDS (0.01% w/v), exponential phase at 37°C or stationary phase at 30°C) (14) to perform differential gene expression analysis using RNA-seq data produced from cells growing in exponential phase at 30°C as the reference. We then used a variability score as a proxy to evaluate the degree of gene expression variability. A gene not regulated in any of these four alternative conditions will have a score of 0 whereas a gene regulated by all these modifications of culture conditions will have a score of 4. Our results revealed that genes harboring a TATA-box in their promoter are more prone to be regulated than the genes of the control group (devoid of TATA-box) (Wilcoxon rank sum test, $p$-value < 0.0001) (Figure 1D). TATA-box containing genes also display a higher absolute log fold-change between conditions than genes devoid of TATA-box (Wilcoxon rank sum test, p-value < 0.0001) (Figure 1E).

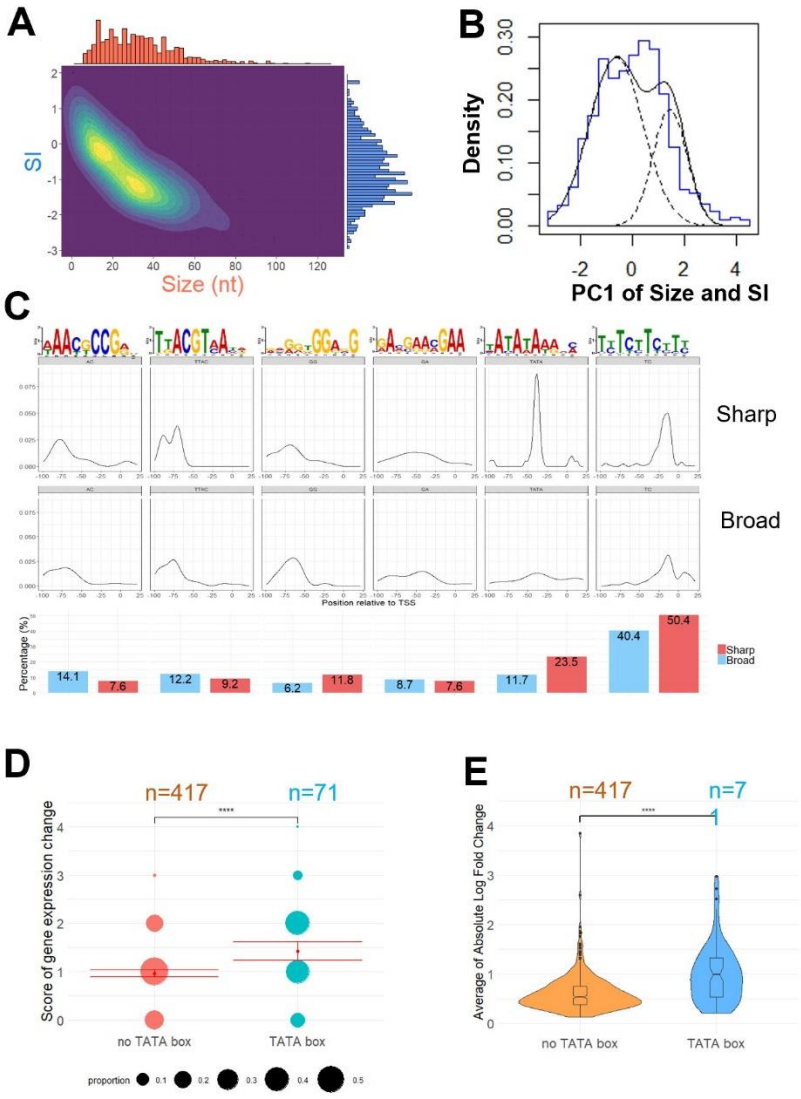



**Figure 1. TSS cluster classification, motif enrichment, and the relationship with gene expression**

(A) The distribution of TSS clusters in terms of cluster width (Size) and shape index (SI) is represented as a 2D density plot in which Size and SI distribution of TSS clusters display a bimodal pattern. Corresponding histogram of Size and SI are projected on the x-axis and y-axis, respectively. (B) Histogram (bold blue line) and density plot (bold black line) of the first principal component (PC1) of Size and SI. Two subpopulations of PC1 are detected by statistical test using "MixtureInf" R package and represented as the theoretical density plot (doted black line). (C) Upper panel: Enrichment and position relative to the major position within the TSS cluster of 6 detected motifs within the promoter sequence. Lower panel: Percentage of presence of each motif in Sharp and Broad clusters associated genes. (D) Score of gene expression change for TATA-box-harboring genes (pink) or devoid of TATA-box

(blue) are represented as dots. Dot size is proportional to the percentage of each score value in y-axis. Sample sizes n, mean (red dots) and confidence interval (red lines) of the mean are indicated for both groups. Significance was computed using Wilcoxon rank sum test, *p*-value < 0.0001. (E) Magnitude of gene expression change for TATA-box-harboring genes and genes devoid of TATA-box is depicted as violin plots and box plots. Average of Absolute Log Fold Change in 4 growth conditions compared to WT (main text) is on y-axis. Sample sizes n are indicated for both groups. Significance was computed using Wilcoxon rank sum test, *p*-value < 0.0001.

Given that Sharp TSS cluster associated genes are enriched in TATA-boxes, it is no surprise that these genes are also more prone to be regulated and with a higher magnitude than Broad cluster associated genes (Wilcoxon rank sum test, p-value < 0.05) (Figure 2A). To figure out which feature, the presence of TATA-box or the TSS cluster shape, is the true determinant for gene expression variability, we conducted stratification analyses. We compared absolute log fold-change between Sharp and Broad cluster associated genes within the TATA-box-harboring genes and within the no-TATA-box genes. In both groups, the difference between Sharp and Broad cluster is no longer detected, suggesting that this higher magnitude of gene expression change in Sharp cluster is most likely due to higher percentage of TATA-box containing genes in this group (Figure 2B). Meanwhile, TATA-box containing genes display higher absolute log fold-change between conditions compared to the genes without TATA-box regardless of the TSS cluster shape (Wilcoxon rank sum test, p-value < 0.0001) (Figure 2C) confirming the presence of TATA-box as the decisive factor to explain gene expression variability in *C. neoformans*. Overall, this re-analysis of TSS-seq data revealed the existence of two types of TSS clusters in *Cryptococcus*. Sharp TSS cluster-containing genes are more often associated with a TATA-box and more prone to be regulated in response to a modification of the environment than the Broad TSS cluster containing genes. These results are in good agreement

with data published on several other organisms (1,12,16,56,57) suggesting that our TSS cluster definition approach is valid.

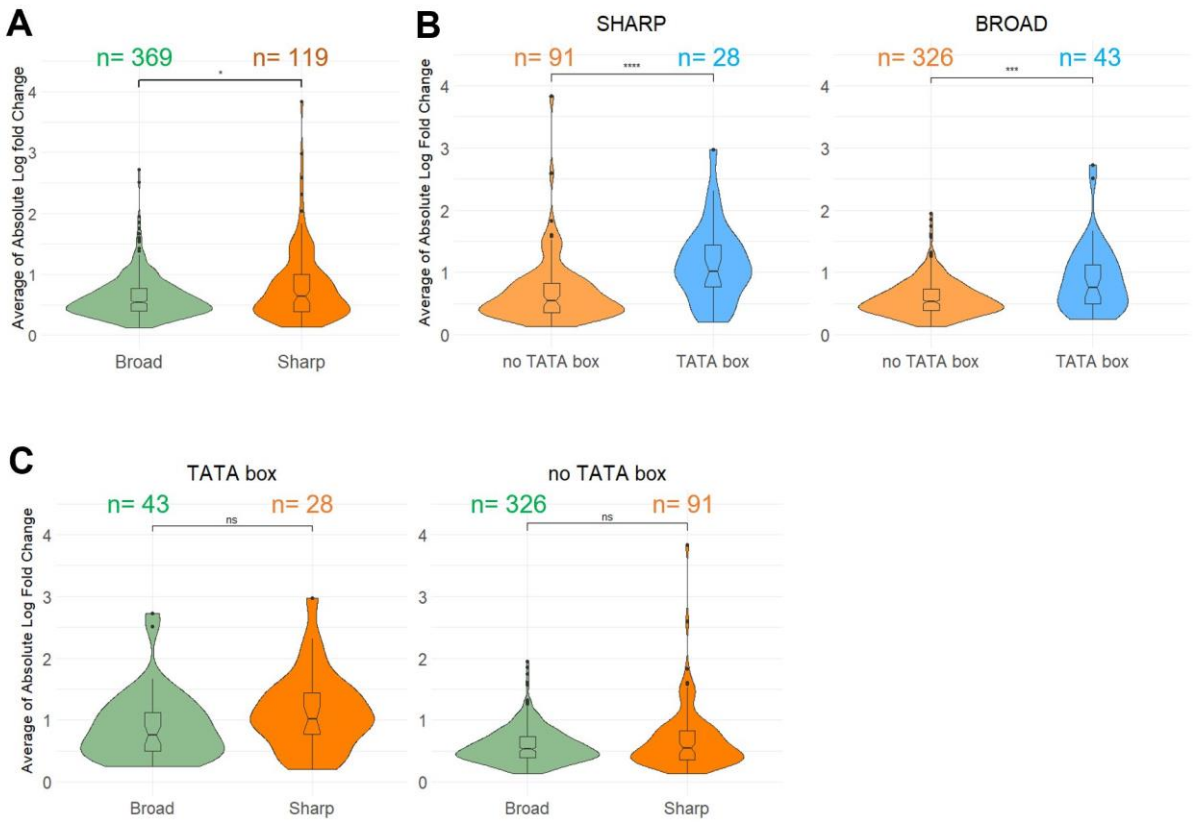



**Figure 2. Relationship between TSS cluster shape, TATA-box and gene expression dynamic. (**A) Magnitude of gene expression change of Sharp cluster genes and Broad cluster genes is depicted as violin plots and box plots. Average of Absolute Log Fold Change in 4 growth conditions compared to WT (main text) is on y-axis. Sample sizes n are indicated for each group. Significance was computed using Wilcoxon rank sum test, p-value <0.05. (B) Stratification analysis comparing TATA-box-harboring genes and TATA-box-devoid genes within those having Sharp clusters (left panel) and those with Broad clusters (right panel). Sample sizes n are indicated for each group. Significance was computed using Wilcoxon rank sum test, p-value < 0.0001 for both comparisons. (C) Stratification analysis comparing genes with Sharp clusters and genes with Broad clusters within TATA-box-harboring genes (left panel) and TATA-box-devoid genes (right panel). Sample sizes n are indicated for each group. Significance was computed using Wilcoxon rank sum test, p-value > 0.05 for both comparisons.

## Alternative TSS cluster identification in *Cryptococcus*

To evaluate alternative TSS usage in *Cryptococcus*, we applied the same subclusterization procedure to all previously identified TSS clusters (14). We then merged the overlapping TSS clusters obtained from the four growth conditions in triplicate. The resulting TSS cluster reference GFF file defines the genomic coordinates of 7,213 TSS clusters associated with 4,931 coding genes in *C. neoformans* (72.6% of the coding genes) (Supplementary Table S4). Among these clusters, we defined 3,817 as alternative TSS clusters associated with 2,431 genes. We performed the same analysis using the *C. deneoformans* data and defined 4,064 alternative TSS clusters associated with 2,581 genes (Supplementary Table S4).

We classified these alternative TSS clusters in three categories depending on their position relative to the annotated start codon (aATG) : within the Transcript Leader (TL), in close proximity with the aATG, or far downstream of aATG (Figure 3).

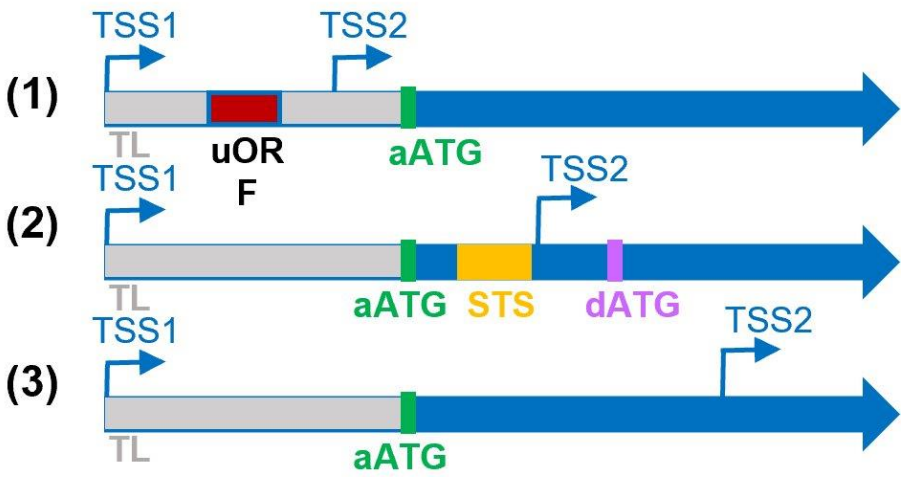



**Figure 3.** Scheme of three scenarios where the position of TSS2 (alternative TSS cluster) is (1) within the Transcript leader, (2) aATG-associated, (3) far downstream of the aATG.

*altTSS within the TL sequence*

A large proportion of these alternative TSS clusters (37% (n = 1,405) in *C. neoformans* and 31% (n = 1,256) in *C. deneoformans*) are positioned within the TL sequence. Their usage regulates the length of the TL sequence (Figure 3) and can result in the inclusion or exclusion of protein binding sites, secondary structures or uORFs, potentially impacting mRNA subcellular localization, stability or translation efficiency (14,15,58). We previously reported that *Cryptococcus* TL sequences are very rich in uORFs and that uORF containing mRNAs are more prone to be degraded by the NMD pathway than uORF-free mRNAs, thus reducing translation efficiency (14). Here, we identified 542 and 633 genes in *C. neoformans* and *C. deneoformans* respectively, for which the sequence between the annotated TSS cluster and the alternative TSS cluster located in the TL contains one or several uORFs. In these genes, alternative TSS usage can regulate the presence or absence of uORFs within the mRNA. In these cases, we reasoned that the production of a short transcript devoid of uORFs would be a way to stabilize it by preventing its degradation by the NMD pathway. Accordingly, among the genes containing one uORF within their TL sequence, the ones containing an alternative TSS potentially skipping it are less prone to be up-regulated by the deletion of the major NMD factor *UPF1* than those devoid of alternative TSS within their TL sequence (Figure 4A) (Wilcoxon rank sum test, *p*-value = $1.246 \times 10^{-10}$). For instance, usage of an alternative TSS within the TL sequence of the CNA08250 locus skips 11 uORFs and results in the production of a short isoform immune to the NMD, whereas the long isoform can only be revealed by the deletion of *UPF1* (Figure 4B).

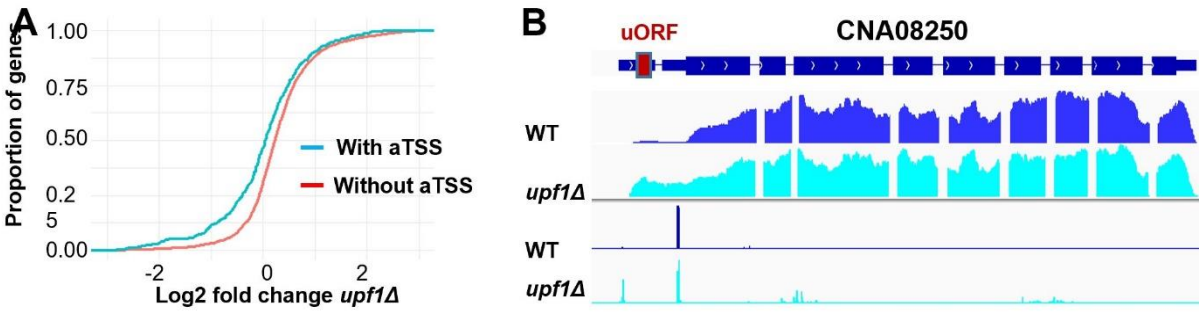


**Figure 4. altTSS usage regulates the presence or absence of uORFs within the mRNA impacting its sensitivity to NMD.** Among 3,055 uAUG containing genes in *C. deneoformans*, genes having alternative TSS within their TL sequence are more resistant to NMD. (A) Cumulative distribution plot of Log fold change between *upf1Δ* mutant and WT at 30°C are generated for genes which use an altTSS at 30°C (blue line) and genes that do not use an altTSS at 30°C (red line). A one-sided Kolmogorov–Smirnov test shows that these differences are significant (p-value = $1.95 \times 10^{-10}$). (B) IGV visualization of RNA-seq and TSS-seq at the CNA08250 locus. The gene CNA08250 possesses an alternative TSS that skips the uORFs of the transcript leader sequence. At 30°C, the alternative TSS is prominently used, resulting in the short mRNA isoform devoid of uORF.

*altTSS close to the annotated start codon*

We identified 220 alternative TSS clusters associated with 213 genes in *C. neoformans* (208 TSS clusters and 199 genes in *C. deneoformans*) positioned between -20 and +90 nt from the aATG, when counting bases on the spliced mRNA. The usage of each of these altTSSs results in the transcription of an mRNA which can code a protein truncated from its N-termini, thereby potentially lacking an N-terminal targeting motif (Figure 3). Accordingly, DeepLoc 2.0 (59) analysis predicts that whereas for 169 of these genes the long isoform has a defined subcellular localization, for 87 of them the usage of an altTSS is predicted to result in the production of a shorter isoform with a different subcellular localization (Table S5 and Figure 5). For instance, we identified an altTSS associated with the gene *MAE102* (CNAG_06374) which encodes a putative mitochondrial malate dehydrogenase (Figure 6A). Depending on the TSS used, the predicted translated protein contains a mitochondrial targeting signal (MTS) or not. Interestingly, the usage of this altTSS is regulated by the growth condition, the long isoform encoding the putative mitochondrial protein being transcribed in stationary phase at

30°C, whereas the exponential phase condition triggers mostly the transcription of the short isoform predicted to code a protein devoid of MTS. Accordingly, a C-terminal mNeonGreen-tagged version of Mae102 protein accumulates into mitochondrial-like particles in stationary phase (Figure 6B) whereas, in exponential phase it localizes in the cytosol and the nucleus (Figure 6B). Moreover, mutation of the first ATG into CGT (M1R) or deletion of the sequence between the two ATGs (MTSΔ) is associated with the loss of the Mae102 protein mitochondrial-like localization in stationary phase confirming the functionality of the MTS sequence. In contrast, the exponential phase specific nuclear localisation of the protein was conserved suggesting an independent mechanism. Overall, these results are consistent with the alternative localization of the Mae102 isoforms depending on the alternative TSS usage regulated by the growth phase.

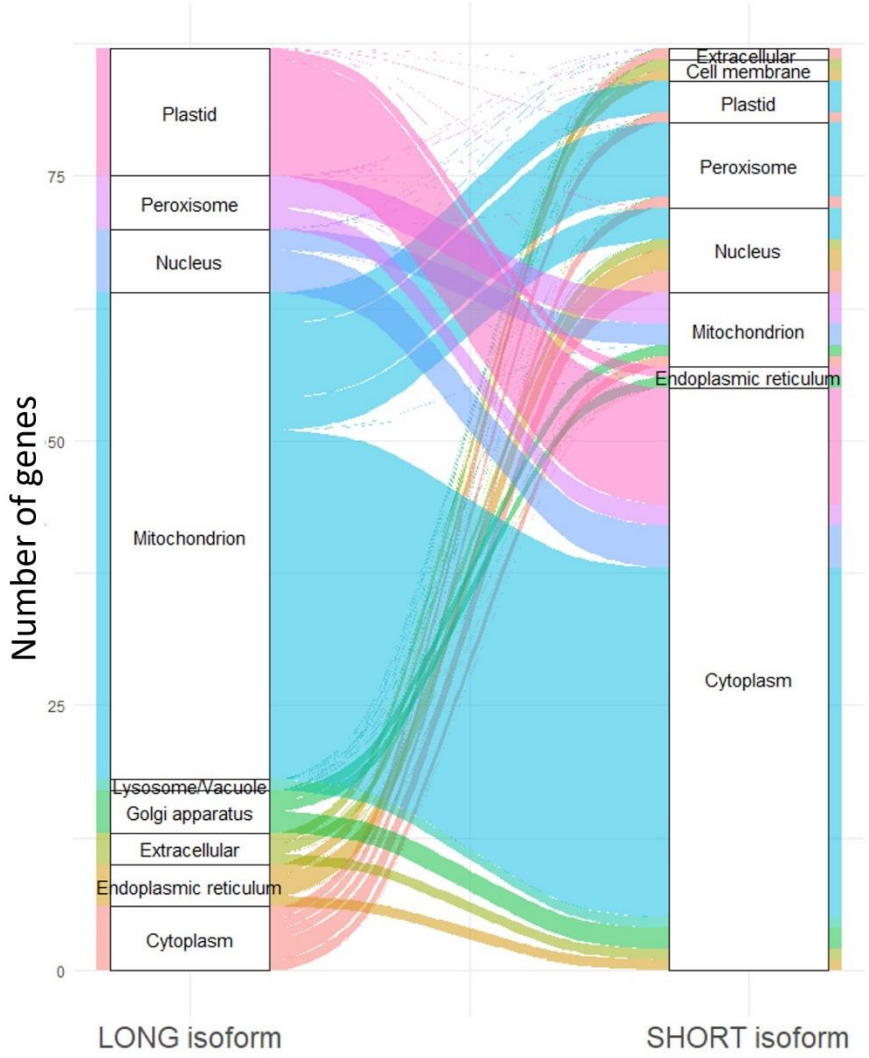



**Figure 5. altTSS usage can result in the production of proteins shorter than the annotated ones, lacking N-terminal targeting motif** and targeted to a different organelle. Alluvial diagram depicting the predicted localization of long and short protein isoforms as predicted by DeepLoc 2.0 (59).

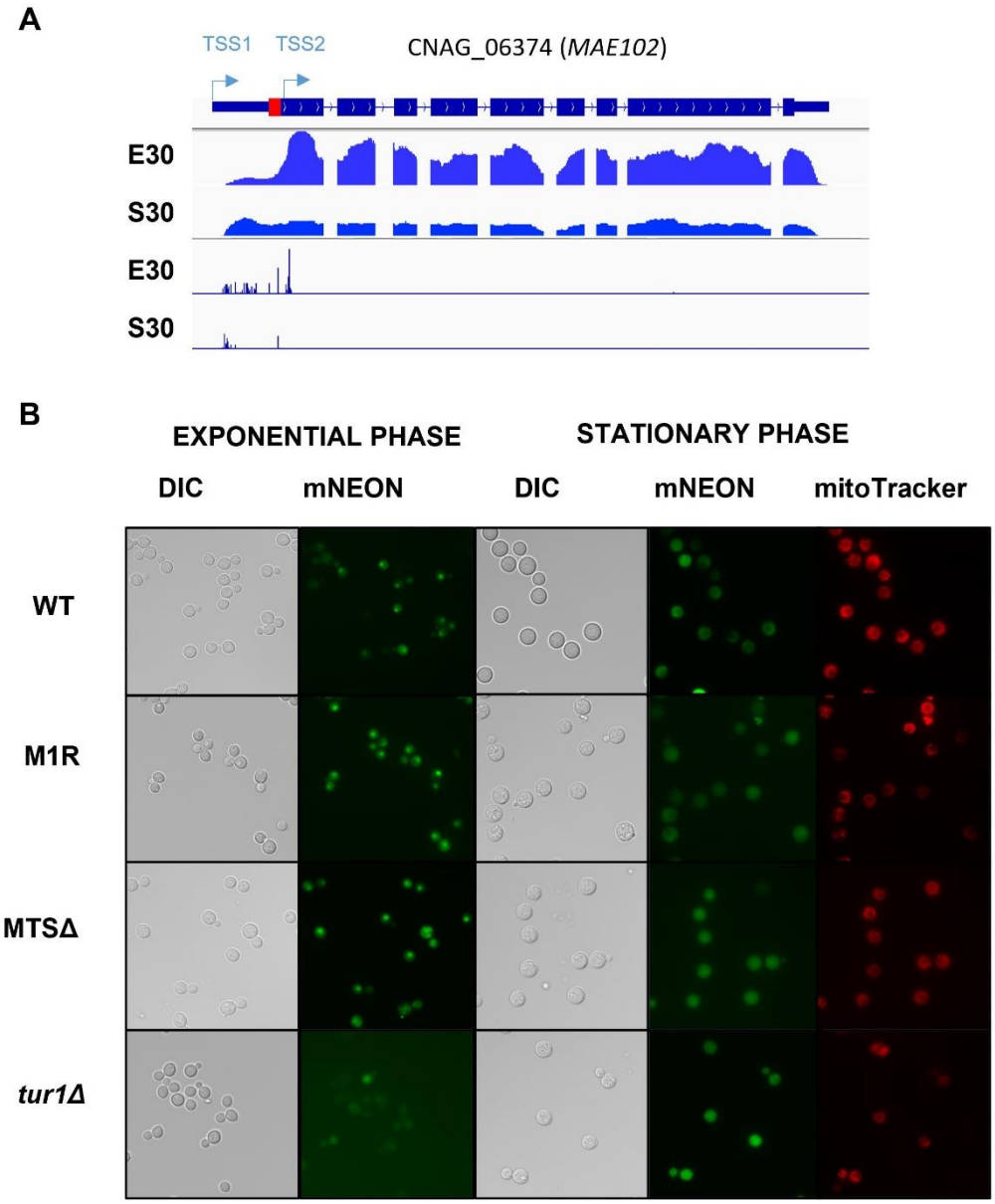



**Figure 6. Mae102 subcellular localization is regulated through alternative TSS usage.** (A) IGV visualization of RNA-seq and TSS-seq at the *MAE102* locus. In exponential phase at 30°C cells mainly use an altTSS (TSS2) located 19 bp downstream of the annotated ATG. The shorter mRNA produces a protein isoform lacking the mitochondria targeting signal (MTS, red box) as predicted by DeepLoc 2.0 and MitoFates (59,60). In stationary phase at 30°C, the annotated TSS (TSS1) is mainly used to produce a full-length transcript which is translated in a protein containing an MTS at its N-terminal. In exponential phase at 30°C, usage of the downstream TSS2 produces a protein expected to go mainly to cytosol. (B) DIC and fluorescent microscopy images of *C. neoformans* cells expressing a Mae102-mNeonGreen fusion protein (green) in

different mutants: WT, *mae102-MTSΔ*, *mae102-M1R* and *tur1Δ*. Mitochondria are stained using MitoTracker (red).

*altTSS downstream of the start codon*

The 2,181 remaining TSS clusters (2,576 in *C. deneoformans*) are positioned at least 90 bp after the annotated start codon and promote the transcription of shorter RNAs than the annotated ones (Supplementary Table S4). We named this last category of transcripts TRASS for Transcript Resulting from Alternative Start Sites. One spectacular example of TRASS is given in Figure 7. At the CNC03460 locus in *C. deneoformans*, cells growing in exponential phase at 30°C use mainly an altTSS located within the sixth intron of the gene and positioned within a CDS-intron 1395 bp downstream of the annotated TSS. In contrast, cells in stationary phase mainly use the annotated altTSS as confirmed by Northern-blot analysis (Figure 7B).

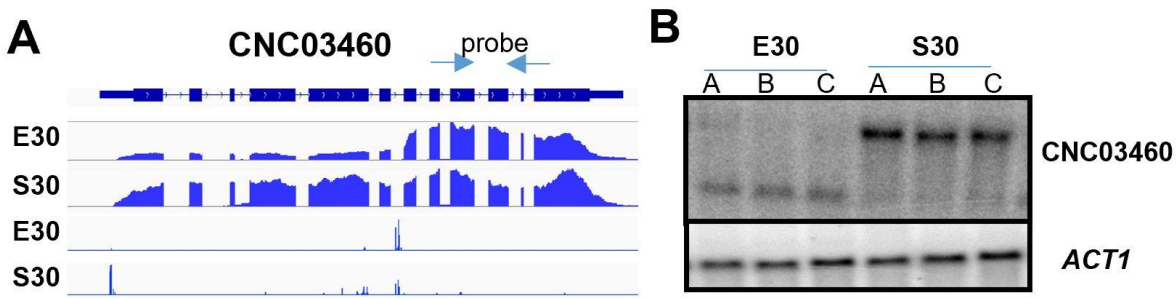



**Figure 7. altTSS usage regulates the production of TRASS in *Cryptococcus*** (A) IGV visualization of RNA-seq and TSS-seq at the CNC03460 locus when *C. deneoformans* cells were cultivated at 30°C under exponential (E30) or stationary (S30) phase, respectively. (B) Northern-blot validation of the alternative RNA molecule production. For CNC03460 the probe was amplified using primers specific of the second half of the gene (arrows). The actin gene (*ACT1*) was used as reference.

Although the function of these RNAs was unknown, we noticed that nearly all the TRASS have coding capacity. Overall, 1,523 potential new proteins could be encoded by these RNAs in *C. neoformans*. Most of these new proteins would be in frame with the annotated ones and

would thus be completely ignored though classical proteomic analysis. To gain insights into the coding capacities of the TRASS, we performed N-terminomic analysis (44) using proteins extracted from cells growing in stationary phase at 30°C. We identified 844 peptide sequences corresponding to N-terminal sequences associated with 12% of the coding genes (n = 810) in *C. neoformans*. As expected, most of the N-terminal peptides (97%; n = 818) correspond to the annotated N-terminal sequence of 784 proteins (Table S6A). Also as expected, we identified 19 peptides corresponding to 19 proteins presenting an alternative N-terminal sequence produced from an altTSS close to the annotated ATG (Supplementary Table S6B). Finally, 7 N-terminal sequences are likely the products of translation of 7 TRASS (Supplementary Table S6B). A striking case is the gene CNAG_04307 (*URO1*) coding Urate oxidase for which we identified a peptide that is out-of-frame with the main protein. Of course, these represent only a few examples of peptides translated from these RNAs and it is probable that a large part of the TRASS are lncRNAs. Nevertheless, N-terminomic analysis, although powerful and efficient in selecting protein N-termini, is not powerful enough for low abundance alternative N-termini suggesting that much more TRASS encoded peptides are produced in these cells.

**Alternative TSS usage in *Cryptococcus***

To further explore the dynamics of alternative TSS usage in *Cryptococcus*, we evaluated how each TSS is employed depending on the growth phase (Exponential or Stationary) and temperature (30°C or 37°C). TSS cluster usage was evaluated considering an ”expression level” in each replicate of each condition measured by the number of TSS-seq reads associated within a given TSS cluster. These numbers were then normalized by the total amount of TSS-seq reads aligned to each coding gene. To limit the potential bias associated with this type of normalization, we limited our analysis to genes for which we could count at least 20 TSS reads

in each considered condition. Overall, we considered 3,648 TSS clusters belonging to 1,366 genes in *C. neoformans* under four growth conditions (4,311 TSS clusters belonging to 1,621 genes in *C. deneoformans*) for this analysis. We confidently identified 1478 altTSS usage regulations by growth condition in 627 *C. neoformans* genes (Figure 8A, Supplementary Table S7). Alternative TSS usage seems to be more dynamic in *C. deneoformans* with 2,369 significant regulations in 1,070 genes. Interestingly, 25% of the altTSS usage regulations are specific to the considered comparison (Figure 8A). Explicit examples of these regulations are given on the Figure 8B. At the gene *PKP1* (CNAG_00047), two TSS are used alternatively depending on the phase of growth. In stationary phase at 30°C, an altTSS within the second intron is mainly used and promotes the transcription of a TRASS. Strikingly, we also identified *SOD1* and *SOD2* genes as regulated by altTSS usage during this phase transition (Supplementary Figure S5). In that case, the pattern of altTSS usage in stationary phase is similar to the one previously reported in copper limiting condition using gene specific experiments (33). Alternative TSS usage at these two genes has been reported to be regulated by copper and is dependent on the transcription factor Cuf1 which can recognize a Cu-responsive element (CuRE) found in their promoters (33). It is possible that the stationary phase condition somehow mimics the copper shortage as used by the Thiele laboratory thus altering altTSS usage at these loci in the same way. We also observed regulation of altTSS by the temperature. For instance, at the locus CNAG_00812, the usage of a downstream TSS is favored in stationary phase at 37°C thus promoting the production of a TRASS. Moreover, at the loci CNAG_01272 and CNAG_03239, a change in temperature in stationary and exponential phase, respectively, alters the usage of an altTSS located within the TL sequence (Figure 8B). Our original idea was to compare the results obtained with the two studied sibling species of *Cryptococcus*. We observed an apparent low conservation of these regulations with only 16.21% and 10.12% of the altTSS usage regulations observed in *C. neoformans* upon the

transition of the growth phase and temperature change, respectively conserved in *C. deneoformans* (Supplementary Table S4). However, visual examination of some of the "non-conserved" events revealed that the apparent non-conservation was mainly due to the poor level of one or the other TSS in one species which did not pass the different thresholds of our bioinformatic pipelines (Supplementary Figure S6). Indeed, in some cases, the position of the altTSS is conserved but regulation is reversed. In others, the altTSS exists but its position is not conserved (Supplementary Figure S6). These observations suggest that the degree of conservation of these regulations might be more significant than they appear at first glance.

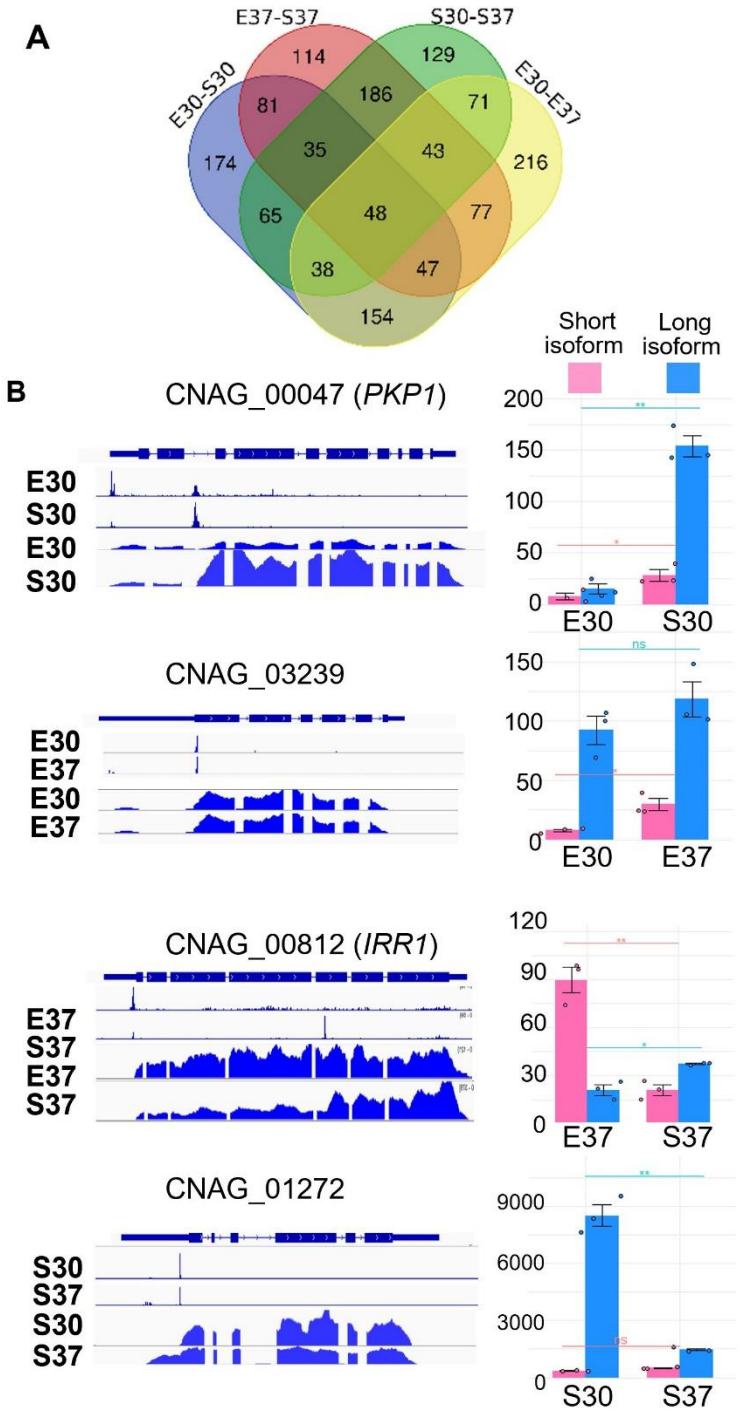



**Figure 8. Differential analysis of altTSS usage in response to the growth phase and temperature changes** (A) Venn diagram illustration the overlap between the altTSS usage regulation (FDR < 0.05) in 4 conditions: exponential phase (E30), stationary phase (S30) at 30°C, exponential phase (E37), stationary phase (S37) at 37°C in *C. neoformans*. (B) IGV visualization of RNA-seq and TSS-seq of examples altTSS usage regulated by the temperature and the growth phase. qRT-PCR confirmation of these regulations was performed using primers specific of one or both isoforms. Y-axis shows 1000*fold-change compared to the number of *ACT1* mRNA molecules. Experiments were performed in biological triplicates and technical duplicates. Error bars are shown median +/-standard deviation.

**Tur1 regulates altTSS usage at the *PKP1* gene**

Our analysis revealed a spectacular widespread regulation of altTSS usage in *Cryptococcus* regulating both transcriptome and proteome structures in response to modifications of the growth condition. Although the mechanisms associated with these regulations remain unknown, the specificity of some of these regulations suggests that alternative TSS usage is mediated by precise regulators. We reasoned that at least some of these regulators might be transcription factors (TF) as previously observed in mammals and in yeast, (9,10,33,61). To identify altTSS usage regulators in *Cryptococcus*, we screened a library of 155 TF mutant strains (46) using a RT-qPCR assay with primers specific of the long or both the long and the short isoforms identified at the *PKP1* gene (CNAG_00047). This gene encodes a protein that shares homology with the *S. cerevisiae* mitochondrial protein kinase Pkp1 (62). In *C. neoformans*, *PKP1* transcription can start at two altTSS clusters with different expression pattern in exponential and stationary phase at 30°C (Figure 8B). RNA was extracted from the wild type strain (WT) and the 155 TF mutant strains grown in stationary phase at 30°C. The level of the long and the short mRNA isoforms was then evaluated by RT-qPCR. We identified a single mutant strain altered for altTSS usage at this locus. This mutant strain displaying no growth defect is mutated at the locus CNAG_05642 encoding a zinc cluster TF previously designated Fzc37 (46). Sequence conservation analysis revealed that this TF is basidiomycete specific with no predicted function. In this mutant strain, the short isoform is strongly down-regulated in stationary phase compared to the WT strain (*t*-test *p-value* $< 0.01$) whereas the expression of the long isoform is not altered (*t*-test *p-value* $> 0.05$) (Figure 9A). We therefore re-named this gene *TUR1* (TSS Usage Regulator 1). As expected, complementation of the *tur1Δ* mutation restored the WT expression of the short isoform (Figure 9A). We confirmed the role of Tur1 in the regulation of altTSS usage at the *PKP1* gene by constructing a conditional mutant

in which a 2xFlag-CBP tagged version of *TUR1* is expressed under the control of the *C. neoformans GAL7* promoter (53) (Figure 9C). As expected, the expression of the *PKP1* short isoform was only observed in stationary phase when the cells were grown with galactose as the sole source of carbon. No regulation was observed when cells were grown in glucose (Figure 9B) a condition in which the *GAL7* promoter is repressed and no Tur1 protein is produced (Figure 9B).

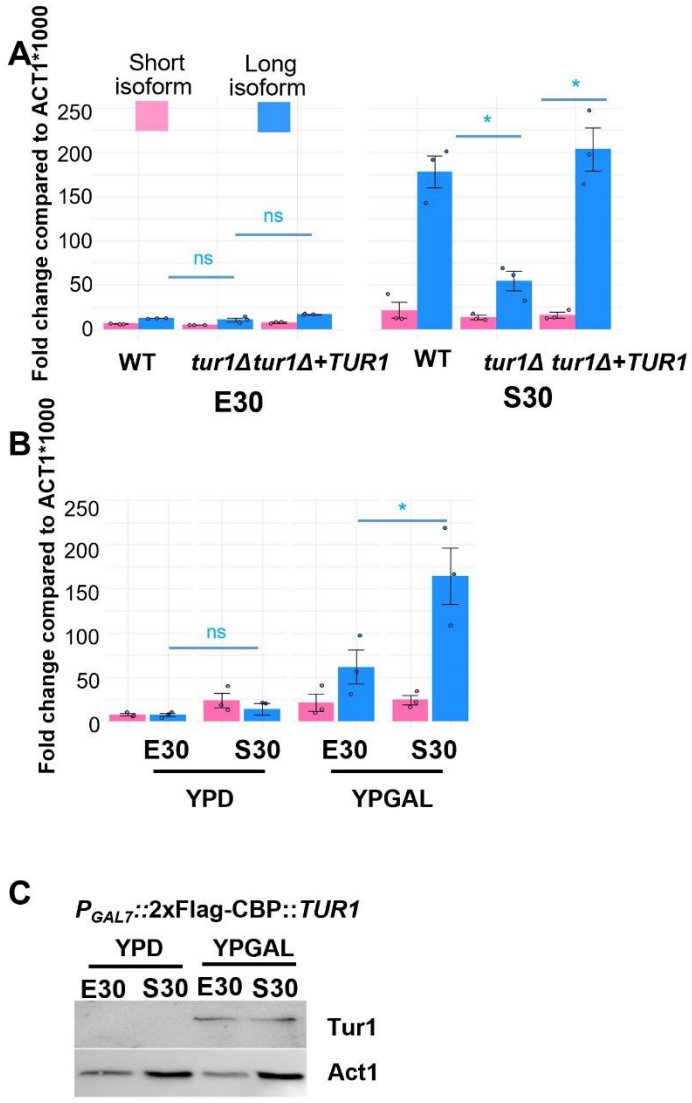



**Figure 9. Tur1 regulates the altTSS usage at *PKP1* during the transition from exponential phase to stationary phase at 30°C.** (A) RT-qPCR measuring the level of the long and short mRNA isoforms of *PKP1* in WT, *tur1Δ*, and the complemented *tur1Δ+TUR1* strains grown in stationary or exponential phase at 30°C. Y-axis shows 1000*fold-change compared to the

number of *ACT1* mRNA molecules (B) RT-qPCR measuring the level of the long and short mRNA isoforms of *PKP1* in a $P_{GAL7}$-*2xFlag-CBP-TUR1* strain in exponential phase and stationary phase at 30°C, cultured in YPD (glucose) or YPGAL (galactose). Y-axis shows 1000*fold-change compared to the number of *ACT1* mRNA molecules. Experiments were performed in biological triplicates and technical duplicates. Error bars are shown median +/- standard deviation. (C) Western-blot analysis of *TUR1* conditional expression. Tur1 was detected using an anti-flag antibody and actin was used as control.

**Tur1 regulated transcriptome structure in *C. neoformans***

To get more insights into the role of *TUR1* in *C. neoformans* biology, we performed spiked-in RNA-seq experiments using RNA extracted from *tur1Δ* cells grown in either stationary or exponential phase. In the wildtype background, the transition between exponential phase to stationary phase is associated with a strong decrease of the transcriptome size with 82% (n= 5,582) of the genes downregulated at least 2-fold (Figure 10A; Supplementary Table S8). Our estimate is that the number of RNA molecules is reduced by 3.8-fold per genome during this phase transition. This is similar to that observed previously in *S. cerevisiae*, in which the overall transcription rate is reduced about three to five times during the exponential to stationary phase transition (63). In the *tur1Δ* mutant strain, this slowdown of the metabolism is less marked than in the WT strain, and we measured only a 1.5-fold reduction of the total number of RNA molecules per genome and identified only 27% (n =1,814) of the genes downregulated by at least two-fold (Figure 9A; Supplementary Table S8). Interestingly, nearly 97% of these downregulated genes in the mutant strain were also downregulated in the WT background (Figure 9B). GO-term analysis showed that this set of conserved downregulated genes is enriched in genes coding translation and ribosome biosynthesis linked functions. In contrast, the set of genes downregulated in the WT but not the *tur1Δ* strain is enriched in genes coding proteins linked to transcription regulation suggesting that this transcription factor is necessary for such regulation.

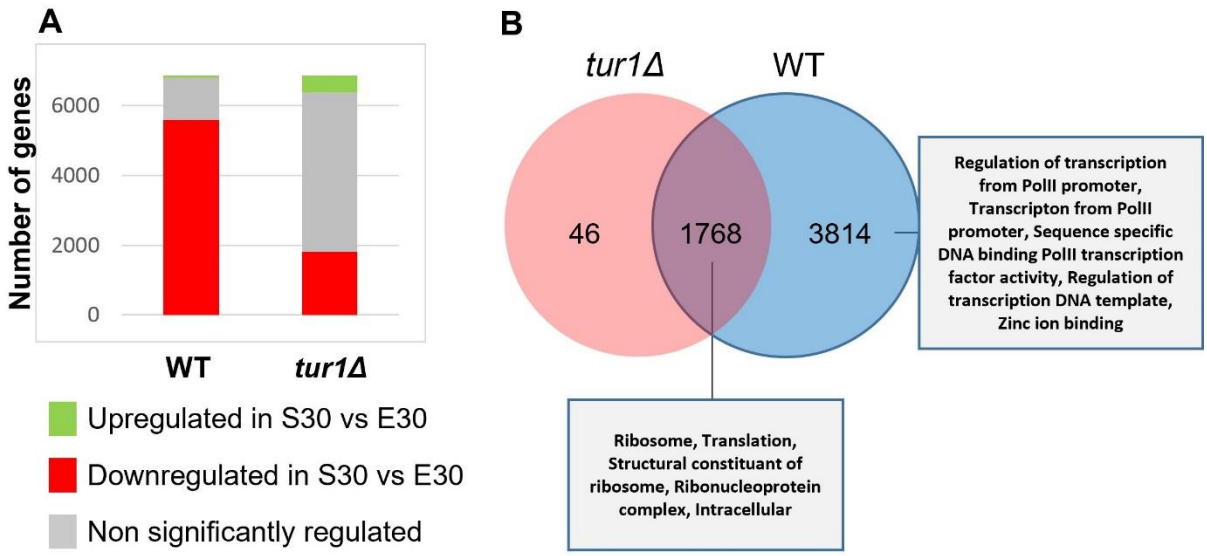



**Figure 10. Comparative differential gene expression analysis during the transition exponential to stationary phase in the WT and *tur1Δ* strains.** (A) Number of genes significantly regulated in WT and *tur1Δ* strains during the exponential to stationary phase transition at 30°C determined by DeSEQ2 (2-fold change; adjusted *p*-value < 0.05). (B) Venn diagram describing the overlap of genes downregulated during the exponential to stationary phase transition at 30°C in WT and *tur1Δ* strains.

This spectacular effect of *TUR1* deletion on transcriptome structure regulation during exponential to stationary phase transition suggested that this TF could have a genome-wide effect on TSS regulation. To discover whether Tur1 would regulate altTSS at any other locus, we performed TSS-seq in the *tur1Δ* mutant strain using cells grown in exponential and stationary phase at 30°C. Here again, in a very conservative way, we restricted our analysis to only 804 genes with at least 2 TSSs and at least 20 TSS reads in all four samples (WT E30, WT S30, *tur1Δ* E30 and *tur1Δ* S30) (Supplementary Table S9). As expected, the stationary phase enhanced usage of the second TSS of *PKP1* in the WT strain whereas usage of this distal TSS is nearly abolished in the *tur1Δ* background confirming the Tur1-dependent altTSS usage regulation at this locus (Figures 10A, 10B). Strikingly, 88% (n = 150) of the 171 genes with altTSS significantly regulated by the growth phase in the WT, were not regulated anymore in a

*tur1Δ* background (Figure 11A). For instance, at the *IDP1* gene (CNAG_03920), which encodes a putative mitochondrial isocitrate dehydrogenase, cells growing in exponential phase mostly use the first TSS which promotes the production of a long mitochondrial isoform whereas in stationary phase, the second TSS is mainly used promoting the synthesis of a short cytosolic isoform (Figure 11B). In the absence of *TUR1*, this growth phase dependent regulation is lost, and the long isoform is expressed in both exponential and stationary phases (Figure 11B). In contrast, at some loci, deletion of *TUR1* has little impact on altTSS usage (Figure 10A). For instance, Tur1 has only a minor role if any on the altTSS usage regulation by the growth phase at the *SOD1* and *SOD2* genes (Figure 11B). *TUR1* deletion also revealed new growth phase dependent altTSS usage regulation in *C. neoformans*. Thus, we identified 5 genes for which this regulation can be only observed in the *tur1Δ* context (Figure 11). The *YSA1* gene coding for an ADP-ribose pyrophosphatase is such an example. In that case, the long isoform produced in stationary phase in the *tur1Δ* strain contains a MTS absent in the current *C. neoformans* genome annotation (Figure 11B).

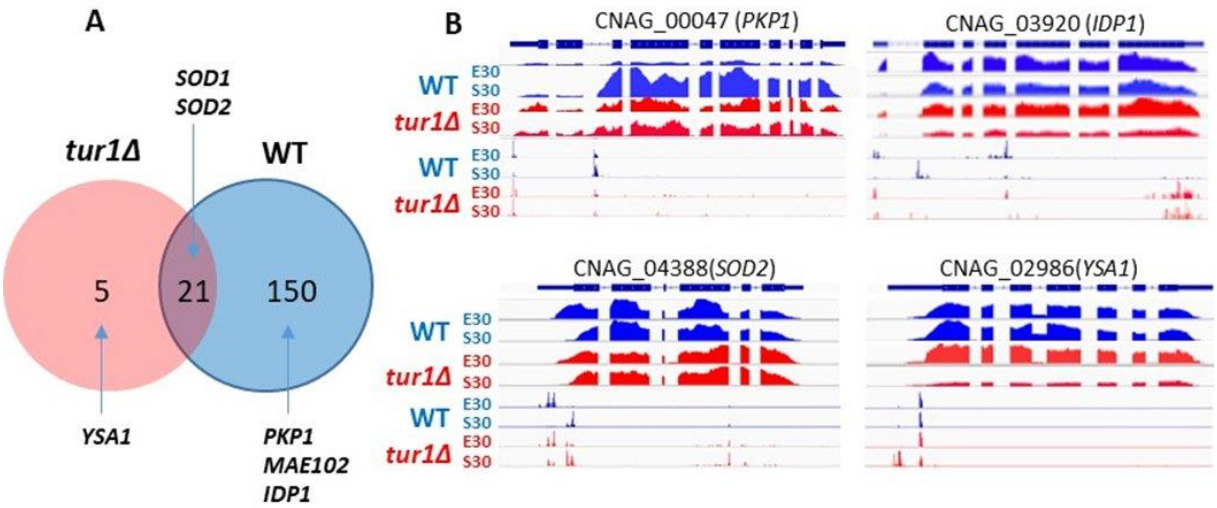



**Figure 11. Tur1 is a global regulator of altTSS usage in *C. neoformans*.** (A) Venn diagram depicting the overlap between altTSS regulated genes during transition from exponential phase to stationary phase in the WT and *tur1Δ* strains. (B) IGV visualization of RNA-seq and TSS-

seq data at representative altTSS regulated genes in WT and *tur1Δ* strains grown in exponential phase (E30) and stationary phase (S30) at 30°C

**A Tur1-dependent Mae102 subcellular targeting**

As seen above, in a wild-type context, a growth phase dependent altTSS usage regulation was observed at the *MAE102* gene. In that case, the second TSS is mostly used in exponential phase whereas in stationary phase, the long isoform targeted to the mitochondria is expressed using the most upstream TSS (Figure 5B). *TUR1* deletion completely reversed this pattern (Figure 12A). Indeed, *tur1Δ* mutant grown in stationary phase preferentially expresses the short isoform, whereas in exponential phase it mostly uses the upstream TSS, only expressing the long isoform albeit at a low level (Figure 12B). To study the impact of *TUR1* deletion on Mae102 protein subcellular targeting, we deleted *TUR1* in a strain expressing a C-terminal mNeonGreen-tagged version of the Mae102 protein. In this genetic background, in exponential phase, Mae102 does not localize in the nucleus in agreement with the absence of production of the short isoform (Figure 5C). In stationary phase, under which the short isoform is strongly induced in *tur1Δ*, Mae102 seems to be poorly targeted to the mitochondria. It is noteworthy that Mae102 is not targeted to the nucleus either suggesting that additional factors specifically present in exponential phase are necessary for targeting this protein to the nucleus. Overall, this analysis exemplifies how Tur1 regulates altTSS usage to control protein subcellular targeting during the transition between exponential to stationary phase (Figure 5C).

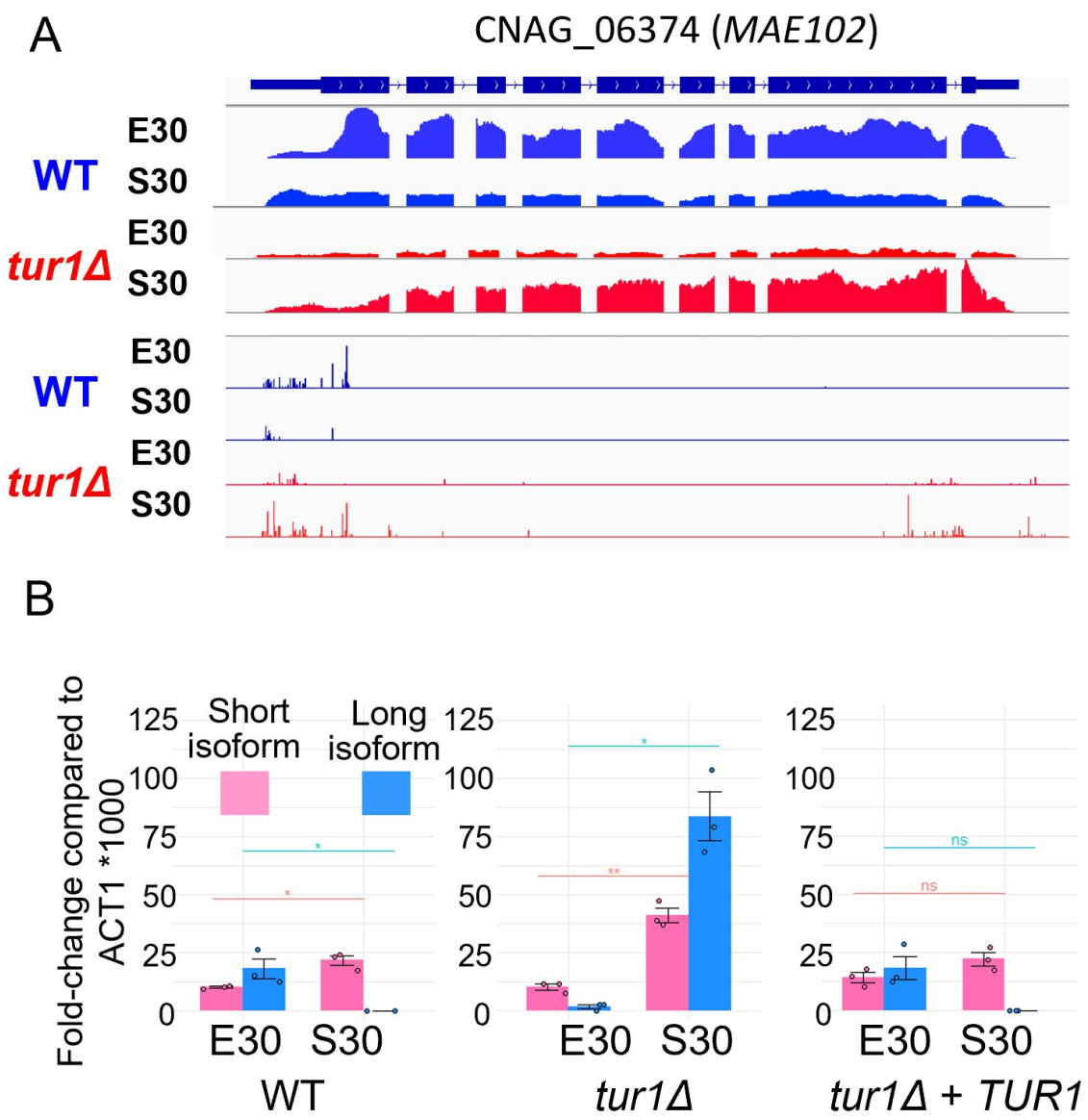



**Figure 12. Tur1 regulates altTSS usage at *MAE102*.** (A) IGV visualization of RNA-seq and TSS-seq of CNAG_06374 in WT and *tur1Δ* strain grown in exponential phase (E30) and stationary phase (S30) at 30°C. (B) RT-qPCR analysis of the level of the long and short *MAE102* mRNA isoforms in WT, *tur1Δ*, and complemented *tur1Δ* + *SH1*::*TUR1* strains grown in exponential or stationary phase at 30°C. Y-axis shows 1000*fold-change compared to the number of *ACT1* mRNA molecules. Experiments were performed in biological triplicates and technical duplicates. Error bars are shown median +/-standard deviation.

**Tur1 regulates oxidative stress sensitivity and resistance to macrophage phagocytosis**

Tur1 was previously reported to regulate mating efficiency, sensitivity to menadione and cadmium, as well as virulence in a mouse infection model (46). We confirmed the sensitivity of *tur1Δ* mutant to menadione, a commonly used superoxide generator and, this sensitivity was reverted in the *tur1Δ+TUR1* complemented strain (Figure 13A). However, the *tur1Δ* mutant does not display growth defect in the presence of reactive oxygen species (ROS) such as $H_2O_2$ (46), 0.2 mM Cumen hydroperoxide, or 2 mM $NaNO_2$ (Figure S7). This suggests the protective function of Tur1 against ROS generator is specific to superoxide generator menadione, consistent with the fact that cells employ distinct mechanisms to adapt to different reactive oxygen species (64). Similarly, we observed no growth defect of *tur1Δ* strain in the presence of an inhibitor of superoxide dismutase (sodium dithiocarbamate (100 mM)) (Figure S7). This result is in good agreement with the fact that Tur1 does not affect neither the expression of *SOD1* and *SOD2* nor the altTSS usage at these loci (Figure 11). STRING 1.5 (65) analysis of the genes regulated by altTSS in a Tur1-dependent manner showed an enrichment in GO-Terms related to translation (large and small ribosome subunits assembly, translation initiation) but also genes related to glycolysis and glycogenesis (glycolytic and pyruvate metabolic processes). We also identified several mitochondrial protein encoding genes among those regulated by altTSS in a Tur1-dependent manner. We thus thought to test the sensitivity of *tur1Δ* to inhibitors of different components of the respiratory chain: 0.4 mM rotenone (complex I), 10 mM Dimethyl malonate (complex II), 5 µg/mL antimycin A (complex III), and 0.2 µg/ml oligomycin (complex V). We found that the growth of the *tur1Δ* mutant is severely impeded in the presence of 5 µg/mL antimycin A, a known inducer of mitochondrial superoxide production (66), but not affected by the other agents (Figure 13A and Supplementary Figure S7). These results suggest that Tur1 might contribute to oxidative stress defense in *C. neoformans*. Interestingly, *tur1Δ* mutant cells are also phagocytosed significantly more efficiently than both the WT and the complemented strain (Figure 13B). Collectively, these results suggest that Tur1

might regulate *C. neoformans* virulence by somehow regulating its sensitivity to ROS and its resistance to phagocytosis.

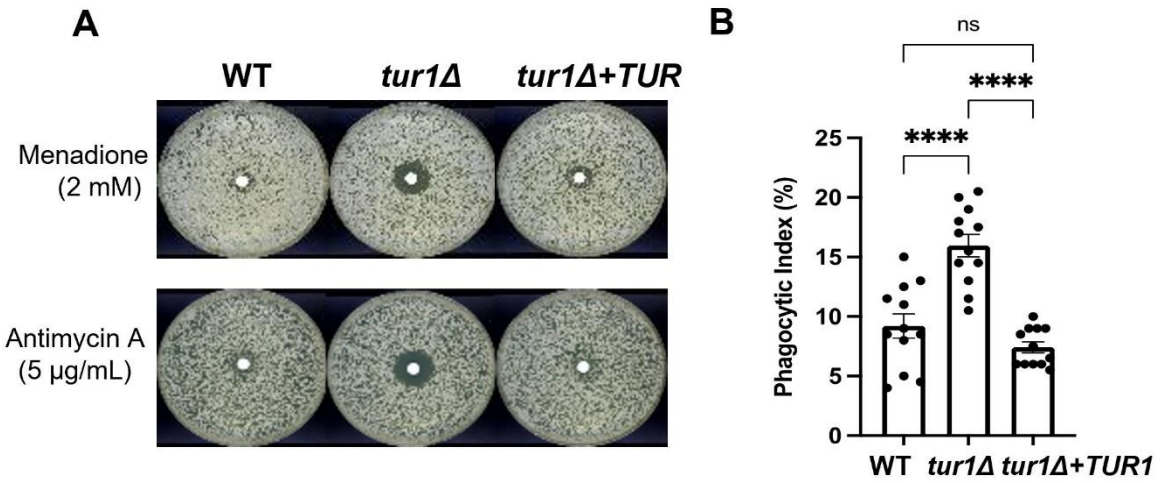



**Figure 13. Tur1 regulates resistance to menadione, antimycin A and macrophage phagocytosis.** (A) Results of resistance disc assays. 2 mM menadione and 5 µg/mL antimycin A filter discs were used on YPD plates. Plates were incubated at 30°C for 3 days and photographed. (B) Macrophage phagocytosis assay shows the phagocytic index for unopsonized *C. neoformans* WT, *tur1Δ*, and complemented *tur1Δ+TUR1* strains as described in Material and Methods.

## Discussion

There is no accepted standard bioinformatics pipeline to analyze TSS-seq data mostly because the structure and density of genes are very different between organisms and because the strategies used to produce this type of sequencing data are diverse (5,17,39,67,68). Here we report a *Cryptococcus*-adapted bioinformatics protocol to analyze TSS-seq data. We thus identified thousands of TSS clusters associated with coding genes in two species of pathogenic *Cryptococcus*. The structure of these TSS clusters (*i.e.*, broad, or sharp) and the preferential association of sharp TSS clusters with TATA-box containing promoters and highly regulated genes resemble results previously obtained in some other reference organisms (2,8,12,41,67,69-71). This suggests that our strategy is robust and that the TSS clusters identified are genuine. Moreover, alternative TSS usage analysis identified thousands of alternative usage events which impact nearly a quarter of the coding genes in both *Cryptococcus* species. Because we have been very conservative in our bioinformatics analysis and because we only compared four conditions of growth for this study, our data suggest that altTSS usage might be far more widespread. In fact, altTSS usage is probably a major regulator of gene expression as well as transcriptome and proteome diversity in *Cryptococcus* as observed in metazoans (5,6,11,41) and for some aspects, in the two model yeasts (16,17,19).

Regulation by alternative TSS can impact gene expression by altering transcript stability as it has been previously shown for alternative splicing (27). Although both regulations have the potential to greatly impact the transcriptome by inducing transcripts degradation, they differ in several aspects. First, altTSS-mediated regulation is predicted to mostly rely on cytoplasmic degradation by NMD whereas the decay induced by intron-retention is NMD-independent and occurs in the nucleus (27). Besides, the extent of the impact of these two means of regulation is dramatically different. Indeed, we previously showed that the impact of alternative splicing is probably minor whereas the data we present in this study suggest a fantastic regulatory

potential for altTSS-mediated regulation. First, as previously reported in baker's yeast (72,73) and more recently genome-wide in plants (74), number of altTSS clusters are positioned close to the annotated ATG in *Cryptococcus*. As shown, growth conditions can regulate altTSS usage at *MAE102* and promote the synthesis of protein isoforms with different N-terminal sequences and different subcellular targeting. Second, in both *Cryptococcus* species, we identified close to two thousand altTSSs positioned within the coding sequence of genes and promoting TRASS transcription. These internal TSS have been previously reported in metazoans (75) and a few examples in model yeasts in which their usage can be also regulated by the growth conditions and/or the genetic background (16,19,22,76,77). For example, in *S. cerevisiae* two studies identified a few dozen genes where internal TSS usage is regulated by cell-fate transition during gametogenesis (19,77). However, the coding potential of the resulting transcripts has been difficult to evaluate. Since most peptides potentially encoded by these truncated transcripts, including the ones we identified in this study, are in frame with the full-length protein, classical proteomics is not able to specifically identify them. Nevertheless, indirect evidence for the existence of these truncated proteins has been presented. For instance, in *S. cerevisiae*, cell-fate transition specific truncated transcripts have been shown to overlap with ribosome protected fragments (19) and in *S. pombe*, "disruptive" internal altTSS located within or downstream of conserved protein domains are generally poorly used. Yet only a few examples of functional truncated proteins have been reported (75,77,78). Here, N-terminomic approach strongly suggests that a significant part of these transcripts is translated. Our results are in good agreement with *S. cerevisiae* data showing that some "internal cryptic" transcripts exhibiting 5' truncation can be translated from downstream in-frame start codons and thus giving rise to shorter polypeptides (22). As for baker's yeast, the function of these TRASS translated peptides is not known. Similarly, the function of all the other TRASS which might not be translated is unknown, though we can speculate that in some cases at least, the usage of one TSS might

impact nucleosome positioning and chromatin modifications to regulate positively or negatively the transcription of the other TSS as shown in yeast (19,79,80). In some other cases, the TRASS might have a function in trans which remains to be discovered.

Our analysis reveals an extended regulation of altTSS usage in response to change in growth conditions in *Cryptococcus*. In that sense, these results are similar to results obtained in *S. pombe* in which 42 % of the genes are regulated by altTSS usage in response to environmental cues (16). The study of *S. pombe* alternative TSS usage identified specific motifs associated with altTSS according to the type of stress or the condition of growth tested (16) suggesting specific regulation of altTSS by transcription factors. This is in good agreement with several other studies in baker's yeast. For instance, Ndt80 regulates the expression of internal transcripts during gametogenesis (19,77) whereas Ume6 controls the expression of long undecoded transcript isoforms (LUTIs) during meiosis (79). In Valine and Isoleucine restricted conditions, Gcn4 can bind within the CDS or the 3'UTR of about 300 genes where it can promote bidirectional transcription (76). These data suggest the existence of a programmed regulation of altTSS usage depending on the growth condition in addition to the classically identified regulons. Similarly in plants, light activates phytochromes which will act on transcription factor activities to coordinate altTSS usage impacting both the uORF-mediated inhibition of gene expression and protein localization (74,81). In humans, a specific pattern of altTSS usage has been reported according to cell type, developmental stage, or cancer (82,83). Accordingly, we here identified a basidiomycete specific transcription factor regulating altTSS usage during the transition from exponential phase to stationary phase in *C. neoformans*. In the absence of this transcription factor, the altTSS usage during this change in cell physiology is globally impaired. In pathogenic and nonpathogenic yeasts, entry to stationary phase is usually characterized by a slowdown of cell metabolism through the action of several transcription

factors (84-86). It is predictable that such program alteration of altTSS usage exists in all these fungi tuning the transcriptome and cell metabolism toward a quiescent state.

Tur1 regulates virulence in *C. neoformans* but does not regulate the expression of its main virulence factors (*i.e.*, capsule biosynthesis, melanin production or fitness at 30°C or 37°C). However, *TUR1* deletion impacts resistance to superoxide stress as well as the phagocytosis of *C. neoformans* by macrophages, phenotypes previously shown to impact virulence in *C. neoformans* (87,88). More specifically, antimycin A treatment is known to drastically induce extra-mitochondrial superoxide release (89). Given that Tur1 regulates the localization of mitochondrial enzymes like Mae102, it could potentially contribute to mitochondrial metabolism's defense against oxidative stress by facilitating the distribution of these enzymes. However, further experiments are necessary to establish a conclusive link and determine a direct impact.

Besides the model yeasts and *Cryptococcus*, altTSS has been poorly analyzed in fungi (18). Nevertheless, specific examples in *Neurospora crassa*, *Aspergillus oryzae*, *Fusarium graminearum* and *Candida glabrata* suggest a widespread although completely ignored regulation of the altTSS usage in pathogenic fungi (31,90-92). We identified thousands of new RNA molecules resulting from the usage of altTSS and showed that at least a subset of them are coding for new proteins or proteins isoforms with a different subcellular localization in *Cryptococcus*. We can anticipate that the same will be true for all other fungal pathogens. Our work suggests that the impact of altTSS usage in their biology and virulence might be major. This study opens up new possibilities for exploring the interactions between host pathogens and emerging players, as well as uncovering insights into gene regulation and the structure of the transcriptome in fungi.

## AVAILABILITY and ACCESSION NUMBERS

Raw and summarized sequencing data are available on GEO under accession number GSE237320 (RNA-seq, TSS-seq, DNA-seq).

**Acknowledgements**

Thi Tuong Vi Dang was supported by the Pasteur-Paris University (PPU) International PhD Program.

# References


1. Haberle, V. and Stark, A. (2018) Eukaryotic core promoters and the functional basis of transcription initiation. *Nat Rev Mol Cell Biol*, **19**, 621-637.
2. Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A.M., Taylor, M.S., Engström, P.G., Frith, M.C. *et al.* (2006) Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet*, **38**, 626-635.
3. Lu, Z. and Lin, Z. (2021) The origin and evolution of a distinct mechanism of transcription initiation in yeasts. *Genome Res*, **31**, 51-63.
4. Adiconis, X., Haber, A.L., Simmons, S.K., Levy Moonshine, A., Ji, Z., Busby, M.A., Shi, X., Jacques, J., Lancaster, M.A., Pan, J.Q. *et al.* (2018) Comprehensive comparative analysis of 5′-end RNA-sequencing methods. *Nature Meth*, **15**, 505-511.
5. Kimura, K., Wakamatsu, A., Suzuki, Y., Ota, T., Nishikawa, T., Yamashita, R., Yamamoto, J., Sekine, M., Tsuritani, K., Wakaguri, H. *et al.* (2006) Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes. *Genome Res*, **16**, 55-65.
6. Nepal, C., Hadzhiev, Y., Previti, C., Haberle, V., Li, N., Takahashi, H., Suzuki, A.M., Sheng, Y., Abdelhamid, R.F., Anand, S. *et al.* (2013) Dynamic regulation of the transcription initiation landscape at single nucleotide resolution during vertebrate embryogenesis. *Genome Res*, **23**, 1938-1950.
7. Cooper, S.J., Trinklein, N.D., Anton, E.D., Nguyen, L. and Myers, R.M. (2006) Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. *Genome Res*, **16**, 1-10.
8. Forrest, A.R.R., Kawaji, H., Rehli, M., Kenneth Baillie, J., de Hoon, M.J.L., Haberle, V., Lassmann, T., Kulakovskiy, I.V., Lizio, M., Itoh, M. *et al.* (2014) A promoter-level mammalian expression atlas. *Nature*, **507**, 462-470.
9. Bharti, K., Liu, W., Csermely, T., Bertuzzi, S. and Arnheiter, H. (2008) Alternative promoter use in eye development: the complex role and regulation of the transcription factor MITF. *Development*, **135**, 1169-1178.
10. Patta, I., Madhok, A., Khare, S., Gottimukkala, K.P., Verma, A., Giri, S., Dandewad, V., Seshadri, V., Lal, G., Misra-Sen, J. *et al.* (2020) Dynamic regulation of chromatin organizer SATB1 via TCR-induced alternative promoter switch during T-cell development. *Nucleic Acids Res*, **48**, 5873-5890.
11. Pal, S., Gupta, R., Kim, H., Wickramasinghe, P., Baubet, V., Showe, L.C., Dahmane, N. and Davuluri, R.V. (2011) Alternative transcription exceeds alternative splicing in generating the transcriptome diversity of cerebellar development. *Genome Res*, **21**, 1260-1272.
12. Lu, Z. and Lin, Z. (2019) Pervasive and dynamic transcription initiation in *Saccharomyces cerevisiae*. *Genome Res*, **29**, 1198-1210.
13. Aoyama, T., Nakayama, H., Ueno, K., Inukai, T., Tanabe, K., Nagi, M., Bard, M. and Chibana, H. (2014) Genome-wide survey of transcriptional initiation in the pathogenic fungus, *Candida glabrata*. *Genes Cells*, **19**, 478-503.
14. Wallace, E.W.J., Maufrais, C., Sales-Lee, J., Tuck, L.R., de Oliveira, L., Feuerbach, F., Moyrand, F., Natarajan, P., Madhani, H.D. and Janbon, G. (2020) Quantitative global studies reveal differential translational control by start codon context across the fungal kingdom. *Nucleic Acids Res.*, **48**, 2312-2331.
15. Arribere, J.A. and Gilbert, W.V. (2013) Roles for transcript leaders in translation and mRNA decay revealed by transcript leader sequencing. *Genome Res*, **23**, 977-987.

979 16. Thodberg, M., Thieffry, A., Bornholdt, J., Boyd, M., Holmberg, C., Azad, A., Workman, C.T.,
980 Chen, Y., Ekwall, K., Nielsen, O. *et al.* (2019) Comprehensive profiling of the fission yeast
981 transcription start site activity during stress and media response. *Nucleic Acids Res*, **47**, 1671-
982 1691.
983 17. Pelechano, V., Wei, W. and Steinmetz, L.M. (2013) Extensive transcriptional heterogeneity
984 revealed by isoform profiling. *Nature*, **497**, 127-131.
985 18. Dang, T.T.V., Colin, J. and Janbon, G. (2022) Alternative Transcription Start Site Usage and
986 Functional Implications in Pathogenic Fungi. *J Fungi (Basel)*, **8**.
987 19. Chia, M., Li, C., Marques, S., Pelechano, V., Luscombe, N.M. and van Werven, F.J. (2021) High-
988 resolution analysis of cell-state transitions in yeast suggests widespread transcriptional tuning
989 by alternative starts. *Genome Biol*, **22**, 34.
990 20. Kaplan, C.D., Laprade, L. and Winston, F. (2003) Transcription elongation factors repress
991 transcription initiation from cryptic sites. *Science*, **301**, 1096-1099.
992 21. Carrozza, M.J., Li, B., Florens, L., Suganuma, T., Swanson, S.K., Lee, K.K., Shia, W.-J., Anderson,
993 S., Yates, J., Washburn, M.P. *et al.* (2005) Histone H3 Methylation by Set2 Directs Deacetylation
994 of Coding Regions by Rpd3S to Suppress Spurious Intragenic Transcription. *Cell*, **123**, 581-592.
995 22. Wei, W., Hennig, B.P., Wang, J., Zhang, Y., Piazza, I., Pareja Sanchez, Y., Chabbert, C.D., Adjalley,
996 S.H., Steinmetz, L.M. and Pelechano, V. (2019) Chromatin-sensitive cryptic promoters
997 putatively drive expression of alternative protein isoforms in yeast. *Genome Res*, **29**, 1974-
998 1984.
999 23. Rajasingham, R., Smith, R.M., Park, B.J., Jarvis, J.N., Govender, N.P., Chiller, T.M., Denning,
1000 D.W., Loyse, A. and Boulware, D.R. (2017) Global burden of disease of HIV-associated
1001 cryptococcal meningitis: an updated analysis. *The Lancet Infectious Diseases*, **17**, 873-881.
1002 24. Idnurm, A., Bahn, Y.S., Nielsen, K., Lin, X., Fraser, J.A. and Heitman, J. (2005) Deciphering the
1003 model pathogenic fungus *Cryptococcus neoformans*. *Nat Rev Microbiol*, **3**, 753-764.
1004 25. Johnston, S.A. and May, R.C. (2013) Cryptococcus interactions with macrophages: evasion and
1005 manipulation of the phagosome by a fungal pathogen. *Cell Microbiol*, **15**, 403-411.
1006 26. Janbon, G., Ormerod, K.L., Paulet, D., Byrnes III, E.J., Chatterjee, G., Yadav, V., Mullapudi, N.,
1007 Hon, C.C., Billmyre, R.B., Brunel, F. *et al.* (2014) Analysis of the genome and transcriptome of
1008 *Cryptococcus neoformans* var. *grubii* reveals complex RNA expression and microevolution
1009 leading to virulence attenuation. *PLoS Genet*, **10**, e1004261.
1010 27. Gonzalez-Hilarion, S., Paulet, D., Lee, K.-T., Hon, C.-C., Lechat, P., Mogensen, E., Moyrand, F.,
1011 Proux, C., Barboux, R., Bussotti, G. *et al.* (2016) Intron retention-dependent gene regulation in
1012 *Cryptococcus neoformans*. *Sci. Rep.*, **6**, 32252.
1013 28. Gröhs Ferrareze, P.A., Maufrais, C., Silva Araujo Streit, R., Priest, S.J., Cuomo, C.A., Heitman, J.,
1014 Staats, C.C. and Janbon, G. (2021) Application of an optimized annotation pipeline to the
1015 *Cryptococcus deuterogattii* genome reveals dynamic primary metabolic gene clusters and
1016 genomic impact of RNAi loss. *G3 (Bethesda)*, **11**.
1017 29. Janbon, G. (2018) Introns in *Cryptococcus*. *Mem. Inst. Oswaldo Cruz*, **113**, e170519-e170519.
1018 30. Goebels, C., Thonn, A., Gonzalez-Hilarion, S., Rolland, O., Moyrand, F., Beilharz, T.H. and
1019 Janbon, G. (2013) Introns regulate gene expression in *Cryptococcus neoformans* in a Pab2p
1020 dependent pathway. *PLoS Genet*, **9**, e1003686.
1021 31. Kaur, J.N. and Panepinto, J.C. (2016) Morphotype-specific effector functions of *Cryptococcus*
1022 *neoformans PUM1*. *Sci. Rep.*, **6**, 23638.
1023 32. Verma, S. and Idnurm, A. (2013) The Uve1 endonuclease Is regulated by the white collar
1024 complex to protect *Cryptococcus neoformans* from UV Damage. *PLoS Genet*, **9**, e1003769.
1025 33. Smith, A.D., Garcia-Santamarina, S., Ralle, M., Loiselle, D.R., Haystead, T.A. and Thiele, D.J.
1026 (2021) Transcription factor-driven alternative localization of *Cryptococcus neoformans*
1027 superoxide dismutase. *J Biol Chem*, **296**, 100391.
1028 34. Winston, F., Dollard, C. and Ricupero-Hovasse, S.L. (1995) Construction of a set of convenient
1029 *Saccharomyces cerevisiae* strains that are isogenic to S288C. *Yeast*, **11**, 53-55.

1030 35. Varma, A. and Kwon-Chung, K.J. (1991) Rapid method to extract DNA from *Cryptococcus*
1031 *neoformans*. *J Clin Microbiol*, **29**, 810-812.
1032 36. Moyrand, F., Lafontaine, I., Fontaine, T. and Janbon, G. (2008) *UGE1* and *UGE2* regulate the
1033 UDP-glucose/UDP-galactose equilibrium in *Cryptococcus neoformans*. *Eukaryot Cell*, **7**, 2069-
1034 2077.
1035 37. Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat*
1036 *Methods*, **9**, 357-359.
1037 38. Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. and Salzberg, S.L. (2013) TopHat2:
1038 accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions.
1039 *Genome Biol*, **14**, R36-R36.
1040 39. Malabat, C., Feuerbach, F., Ma, L., Saveanu, C. and Jacquier, A. (2015) Quality control of
1041 transcription start site selection by nonsense-mediated-mRNA decay. *Elife*, **4**, e06722.
1042 40. Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and
1043 dispersion for RNA-seq data with DESeq2. *Genome Biol*, **15**, 550.
1044 41. Hoskins, R.A., Landolin, J.M., Brown, J.B., Sandler, J.E., Takahashi, H., Lassmann, T., Yu, C.,
1045 Booth, B.W., Zhang, D., Wan, K.H. *et al.* (2011) Genome-wide analysis of promoter architecture
1046 in *Drosophila melanogaster*. *Genome Res*, **21**, 182-192.
1047 42. Li, P. and Chen, J. (2010) Testing the Order of a Finite Mixture. *Journal of the American*
1048 *Statistical Association*, **105**, 1084-1092.
1049 43. Bailey, T.L., Johnson, J., Grant, C.E. and Noble, W.S. (2015) The MEME Suite. *Nucleic Acids Res*,
1050 **43**, W39-49.
1051 44. Kleifeld, O., Doucet, A., Prudova, A., auf dem Keller, U., Gioia, M., Kizhakkedathu, J.N. and
1052 Overall, C.M. (2011) Identifying and quantifying proteolytic events and the natural N
1053 terminome by terminal amine isotopic labeling of substrates. *Nat Protoc*, **6**, 1578-1611.
1054 45. Cox, J. and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized
1055 p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature*
1056 *Biotechnology*, **26**, 1367-1372.
1057 46. Jung, K.-W., Yang, D.-H., Maeng, S., Lee, K.-T., So, Y.-S., Hong, J., Choi, J., Byun, H.-J., Kim, H.,
1058 Bang, S. *et al.* (2015) Systematic functional profiling of transcription factor networks in
1059 *Cryptococcus neoformans*. *Nat Com*, **6**, 6757.
1060 47. Lin, J., Fan, Y. and Lin, X. (2020) Transformation of *Cryptococcus neoformans* by electroporation
1061 using a transient CRISPR-Cas9 expression (TRACE) system. *Fungal Genet Biol*, **138**, 103364.
1062 48. Arras, S.D., Chitty, J.L., Blake, K.L., Schulz, B.L. and Fraser, J.A. (2015) A genomic safe haven for
1063 mutant complementation in *Cryptococcus neoformans*. *PLoS One*, **10**, e0122916.
1064 49. Summers, D.K., Perry, D.S., Rao, B. and Madhani, H.D. (2020) Coordinate genomic association
1065 of transcription factors controlled by an imported quorum sensing peptide in *Cryptococcus*
1066 *neoformans*. *PLoS Genet*, **16**, e1008744.
1067 50. Saveanu, C., Rousselle, J.C., Lenormand, P., Namane, A., Jacquier, A. and Fromont-Racine, M.
1068 (2007) The p21-activated protein kinase inhibitor Skb15 and its budding yeast homologue are
1069 60S ribosome assembly factors. *Mol Cell Biol*, **27**, 2897-2909.
1070 51. Huang, M.Y., Joshi, M.B., Boucher, M.J., Lee, S., Loza, L.C., Gaylord, E.A., Doering, T.L. and
1071 Madhani, H.D. (2022) Short homology-directed repair using optimized Cas9 in the pathogen
1072 *Cryptococcus neoformans* enables rapid gene deletion and tagging. *Genetics*, **220**.
1073 52. Upadhya, R., Lam, W.C., Maybruck, B.T., Donlin, M.J., Chang, A.L., Kayode, S., Ormerod, K.L.,
1074 Fraser, J.A., Doering, T.L. and Lodge, J.K. (2017) A fluorogenic *C. neoformans* reporter strain
1075 with a robust expression of m-cherry expressed from a safe haven site in the genome. *Fungal*
1076 *Genet Biol*, **108**, 13-25.
1077 53. Wickes, B.L. and Edman, J.C. (1995) The *Cryptococcus neoformans GAL7* gene and its use as an
1078 inducible promoter. *Mol Microbiol*, **16**, 1099-1109.
1079 54. Ralph, P. and Nakoinz, I. (1975) Phagocytosis and cytolysis by a macrophage tumour and its
1080 cloned cell line. *Nature*, **257**, 393-394.

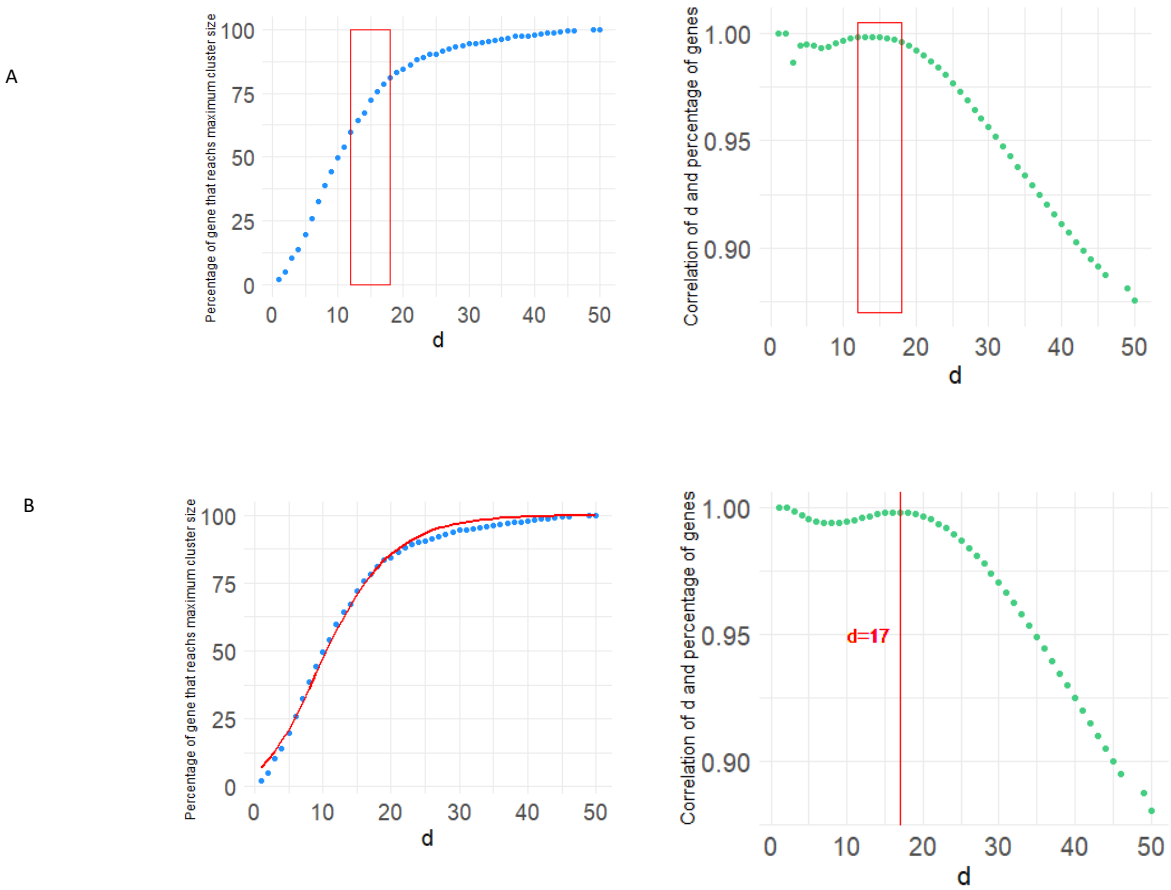
1081 55. Yamamoto, Y.Y., Yoshioka, Y., Hyakumachi, M. and Obokata, J. (2011) Characteristics of core
1082 promoter types with respect to gene structure and expression in *Arabidopsis thaliana*. *DNA*
1083 *Res*, **18**, 333-342.
1084 56. Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y. and Hume, D.A. (2007)
1085 Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nature*
1086 *Reviews Genetics*, **8**, 424-436.
1087 57. Ni, T., Corcoran, D.L., Rach, E.A., Song, S., Spana, E.P., Gao, Y., Ohler, U. and Zhu, J. (2010) A
1088 paired-end sequencing strategy to map the complex landscape of transcription initiation.
1089 *Nature Meth*, **7**, 521-527.
1090 58. Hinnebusch, A.G., Ivanov, I.P. and Sonenberg, N. (2016) Translational control by 5'-
1091 untranslated regions of eukaryotic mRNAs. *Science*, **352**, 1413-1416.
1092 59. Thumuluri, V., Almagro Armenteros, J.J., Johansen, Alexander R., Nielsen, H. and Winther, O.
1093 (2022) DeepLoc 2.0: multi-label subcellular localization prediction using protein language
1094 models. *Nucleic Acids Res.*, **50**, W228-W234.
1095 60. Fukasawa, Y., Tsuji, J., Fu, S.C., Tomii, K., Horton, P. and Imai, K. (2015) MitoFates: improved
1096 prediction of mitochondrial targeting sequences and their cleavage sites. *Mol Cell Proteomics*,
1097 **14**, 1113-1126.
1098 61. Sehgal, A., Hughes, B.T. and Espenshade, P.J. (2008) Oxygen-dependent, alternative promoter
1099 controls translation of tco1+ in fission yeast. *Nucleic Acids Res*, **36**, 2024-2031.
1100 62. Krause-Buchholz, U., Gey, U., Wünschmann, J., Becker, S. and Rödel, G. (2006) YIL042c and
1101 YOR090c encode the kinase and phosphatase of the *Saccharomyces cerevisiae* pyruvate
1102 dehydrogenase complex. *FEBS Lett*, **580**, 2553-2560.
1103 63. Choder, M. (1991) A general topoisomerase I-dependent transcriptional repression in the
1104 stationary phase in yeast. *Genes Dev*, **5**, 2315-2326.
1105 64. Thorpe, G.W., Fong, C.S., Alic, N., Higgins, V.J. and Dawes, I.W. (2004) Cells have distinct
1106 mechanisms to maintain protection against different reactive oxygen species: oxidative-stress-
1107 response genes. *Proc Natl Acad Sci U S A*, **101**, 6564-6569.
1108 65. Szklarczyk, D., Gable, A.L., Nastou, K.C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N.T., Legeay,
1109 M., Fang, T., Bork, P. *et al.* (2021) The STRING database in 2021: customizable protein-protein
1110 networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic*
1111 *Acids Res*, **49**, D605-d612.
1112 66. Wang, L., Duan, Q., Wang, T., Ahmed, M., Zhang, N., Li, Y., Li, L. and Yao, X. (2015)
1113 Mitochondrial Respiratory Chain Inhibitors Involved in ROS Production Induced by Acute High
1114 Concentrations of Iodide and the Effects of SOD as a Protective Factor. *Oxid Med Cell Longev*,
1115 **2015**, 217670.
1116 67. Morton, T., Petricka, J., Corcoran, D.L., Li, S., Winter, C.M., Carda, A., Benfey, P.N., Ohler, U.
1117 and Megraw, M. (2014) Paired-End Analysis of Transcription Start Sites in *Arabidopsis* Reveals
1118 Plant-Specific Promoter Signatures. *Plant Cell*, **26**, 2746.
1119 68. Li, H., Hou, J., Bai, L., Hu, C., Tong, P., Kang, Y., Zhao, X. and Shao, Z. (2015) Genome-wide
1120 analysis of core promoter structures in *Schizosaccharomyces pombe* with DeepCAGE. *RNA*
1121 *Biol.*, **12**, 525-537.
1122 69. Yamashita, R., Sathira, N.P., Kanai, A., Tanimoto, K., Arauchi, T., Tanaka, Y., Hashimoto, S.,
1123 Sugano, S., Nakai, K. and Suzuki, Y. (2011) Genome-wide characterization of transcriptional
1124 start sites in humans by integrative transcriptome analysis. *Genome Res*, **21**, 775-789.
1125 70. Rach, E.A., Winter, D.R., Benjamin, A.M., Corcoran, D.L., Ni, T., Zhu, J. and Ohler, U. (2011)
1126 Transcription initiation patterns indicate divergent strategies for gene regulation at the
1127 chromatin level. *PLoS Genet*, **7**, e1001274.
1128 71. Lenhard, B., Sandelin, A. and Carninci, P. (2012) Metazoan promoters: emerging characteristics
1129 and insights into transcriptional regulation. *Nat Rev Genet*, **13**, 233-245.
1130 72. Natsoulis, G., Hilger, F. and Fink, G.R. (1986) The *HTS1* gene encodes both the cytoplasmic and
1131 mitochondrial histidine tRNA synthetases of *S. cerevisiae*. *Cell*, **46**, 235-243.

1132 73. Carlson, M. and Botstein, D. (1982) Two differentially regulated mRNAs with different 5' ends
1133 encode secreted with intracellular forms of yeast invertase. *Cell*, **28**, 145-154.
1134 74. Ushijima, T., Hanada, K., Gotoh, E., Yamori, W., Kodama, Y., Tanaka, H., Kusano, M., Fukushima,
1135 A., Tokizawa, M., Yamamoto, Y.Y. *et al.* (2017) Light Controls Protein Localization through
1136 Phytochrome-Mediated Alternative Promoter Selection. *Cell*, **171**, 1316-1325.e1312.
1137 75. Nieke, S., Yasmin, N., Kakugawa, K., Yokomizo, T., Muroi, S. and Taniuchi, I. (2017) Unique N-
1138 terminal sequences in two Runx1 isoforms are dispensable for Runx1 function. *BMC*
1139 *Developmental Biology*, **17**, 14.
1140 76. Rawal, Y., Chereji, R.V., Valabhoju, V., Qiu, H., Ocampo, J., Clark, D.J. and Hinnebusch, A.G.
1141 (2018) Gcn4 Binding in Coding Regions Can Activate Internal and Canonical 5' Promoters in
1142 Yeast. *Mol Cell*, **70**, 297-311.e294.
1143 77. Zhou, S., Sternglanz, R. and Neiman, A.M. (2017) Developmentally regulated internal
1144 transcription initiation during meiosis in budding yeast. *PLoS One*, **12**, e0188001.
1145 78. Nuckolls, N.L., Nidamangala Srinivasa, A., Mok, A.C., Helston, R.M., Bravo Núñez, M.A., Lange,
1146 J.J., Gallagher, T.J., Seidel, C.W. and Zanders, S.E. (2022) *S. pombe* wtf drivers use dual
1147 transcriptional regulation and selective protein exclusion from spores to cause meiotic drive.
1148 *PLoS Genet*, **18**, e1009847.
1149 79. Tresenrider, A., Morse, K., Jorgensen, V., Chia, M., Liao, H., van Werven, F.J. and Ünal, E. (2021)
1150 Integrated genomic analysis reveals key features of long undecoded transcript isoform-based
1151 gene repression. *Mol Cell*, **81**, 2231-2245.e2211.
1152 80. Chen, J., Tresenrider, A., Chia, M., McSwiggen, D.T., Spedale, G., Jorgensen, V., Liao, H., van
1153 Werven, F.J. and Ünal, E. (2017) Kinetochore inactivation by expression of a repressive mRNA.
1154 *Elife*, **6**.
1155 81. Kurihara, Y., Makita, Y., Kawashima, M., Fujita, T., Iwasaki, S. and Matsui, M. (2018) Transcripts
1156 from downstream alternative transcription start sites evade uORF-mediated inhibition of gene
1157 expression in *Arabidopsis*. *Proc Natl Acad Sci U S A*, **115**, 7831-7836.
1158 82. Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao,
1159 X., Schmidl, C., Suzuki, T. *et al.* (2014) An atlas of active enhancers across human cell types and
1160 tissues. *Nature*, **507**, 455-461.
1161 83. Demircioğlu, D., Cukuroglu, E., Kindermans, M., Nandi, T., Calabrese, C., Fonseca, N.A., Kahles,
1162 A., Lehmann, K.V., Stegle, O., Brazma, A. *et al.* (2019) A Pan-cancer Transcriptome Analysis
1163 Reveals Pervasive Regulation through Alternative Promoters. *Cell*, **178**, 1465-1477.e1417.
1164 84. Mao, Y. and Chen, C. (2019) The Hap Complex in Yeasts: Structure, Assembly Mode, and Gene
1165 Regulation. *Front Microbiol*, **10**, 1645.
1166 85. Werner-Washburne, M., Braun, E., Johnston, G.C. and Singer, R.A. (1993) Stationary phase in
1167 the yeast *Saccharomyces cerevisiae*. *Microbiol Rev*, **57**, 383-401.
1168 86. Sagot, I. and Laporte, D. (2019) The cell biology of quiescent yeast - a diversity of individual
1169 scenarios. *J Cell Sci*, **132**.
1170 87. Wang, Y., Pawar, S., Dutta, O., Wang, K., Rivera, A. and Xue, C. (2022) Macrophage Mediated
1171 Immunomodulation During *Cryptococcus* Pulmonary Infection. *Front Cell Infect Microbiol*, **12**,
1172 859049.
1173 88. Brown, S.M., Campbell, L.T. and Lodge, J.K. (2007) *Cryptococcus neoformans*, a fungus under
1174 stress. *Curr Opin Microbiol*, **10**, 320-325.
1175 89. Muller, F.L., Liu, Y. and Van Remmen, H. (2004) Complex III releases superoxide to both sides
1176 of the inner mitochondrial membrane. *J Biol Chem*, **279**, 49064-49073.
1177 90. Colot, H.V., Loros, J.J. and Dunlap, J.C. (2005) Temperature-modulated alternative splicing and
1178 promoter use in the Circadian clock gene frequency. *Mol Biol Cell*, **16**, 5563-5571.
1179 91. Inoue, T., Toji, H., Tanaka, M., Takama, M., Hasegawa-Shiro, S., Yamaki, Y., Shintani, T. and
1180 Gomi, K. (2020) Alternative transcription start sites of the enolase-encoding gene enoA are
1181 stringently used in glycolytic/gluconeogenic conditions in *Aspergillus oryzae*. *Curr Genet*, **66**,
1182 729-747.

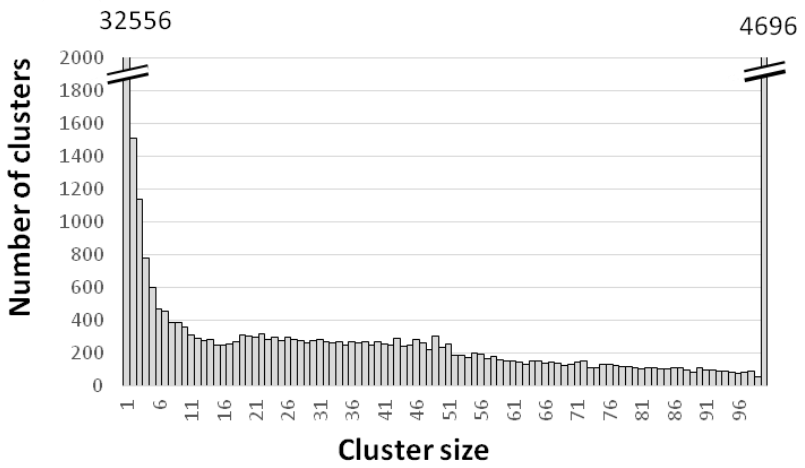
1183 92. Park, A.R., Son, H., Min, K., Park, J., Goo, J.H., Rhee, S., Chae, S.K. and Lee, Y.W. (2015)
1184 Autoregulation of ZEB2 expression for zearalenone production in *Fusarium graminearum*. *Mol*
1185 *Microbiol*, **97**, 942-956.

1186

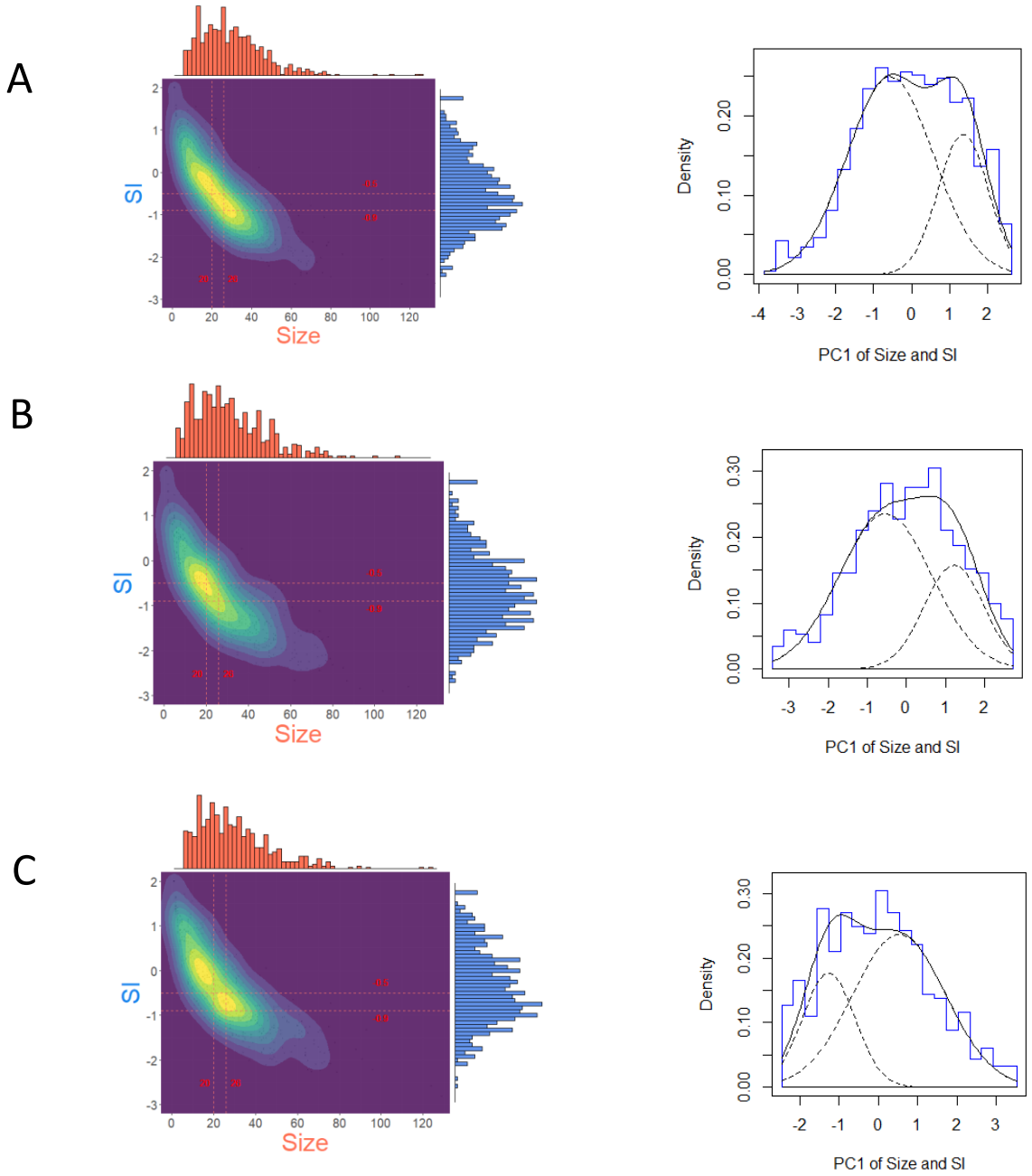
1187



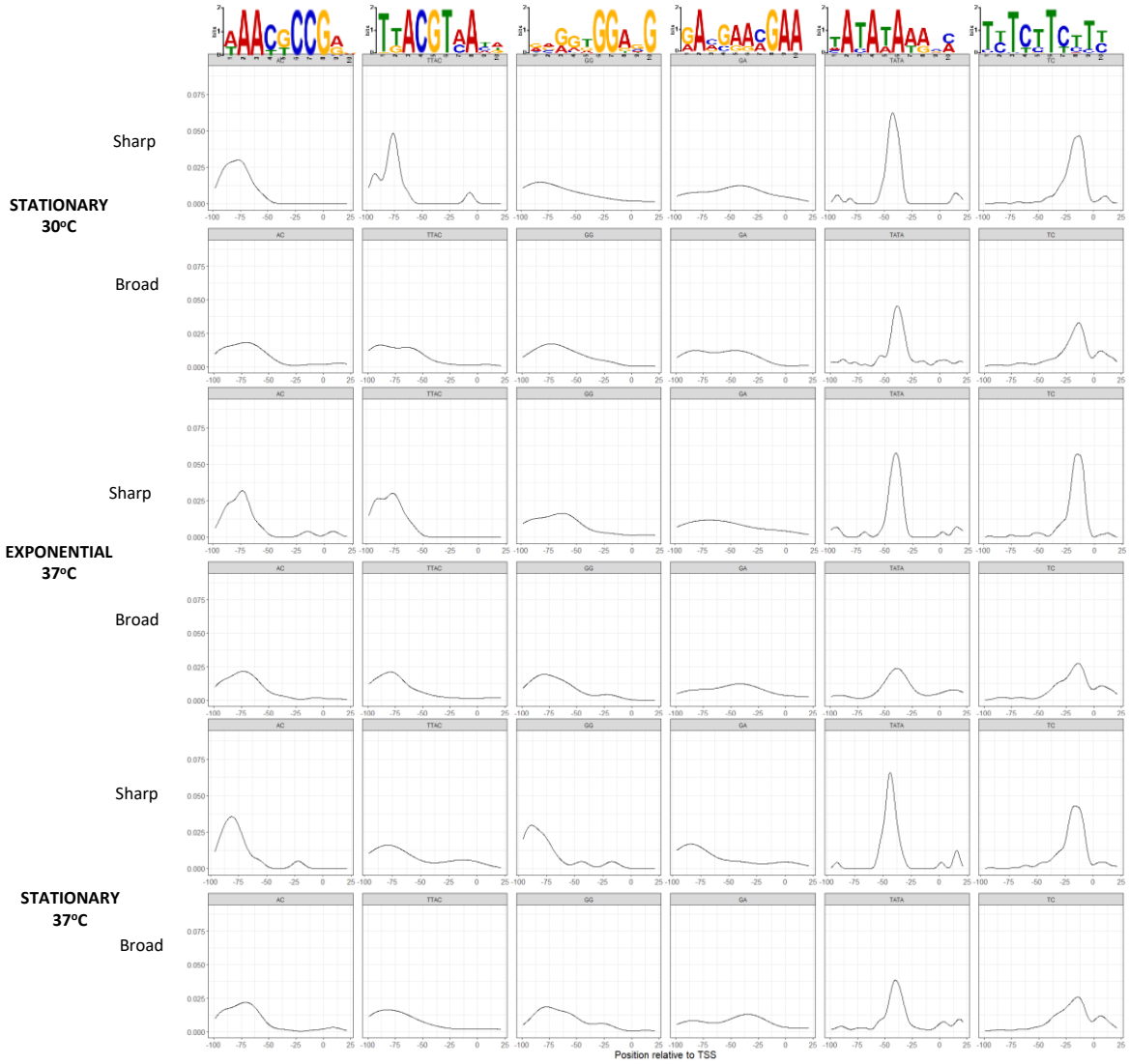

**Supplementary Figure S1. Optimization of TSS data clusterization.** (A) Left panel: Cumulative percentage of clusters that reach size when testing a range of d values from 1 to 50. Right panel: Correlation between d and percentage of clusters that reach a maximum size. The red rectangle illustrates the areas where d runs from 12 to 18 in both panels. (B) Left panel: The data in (A) is modelized as the red smooth curve. Right panel: Correlation between d and modelized percentage of clusters that reach maximum size, which reaches the maximum value at d=17 (red line).



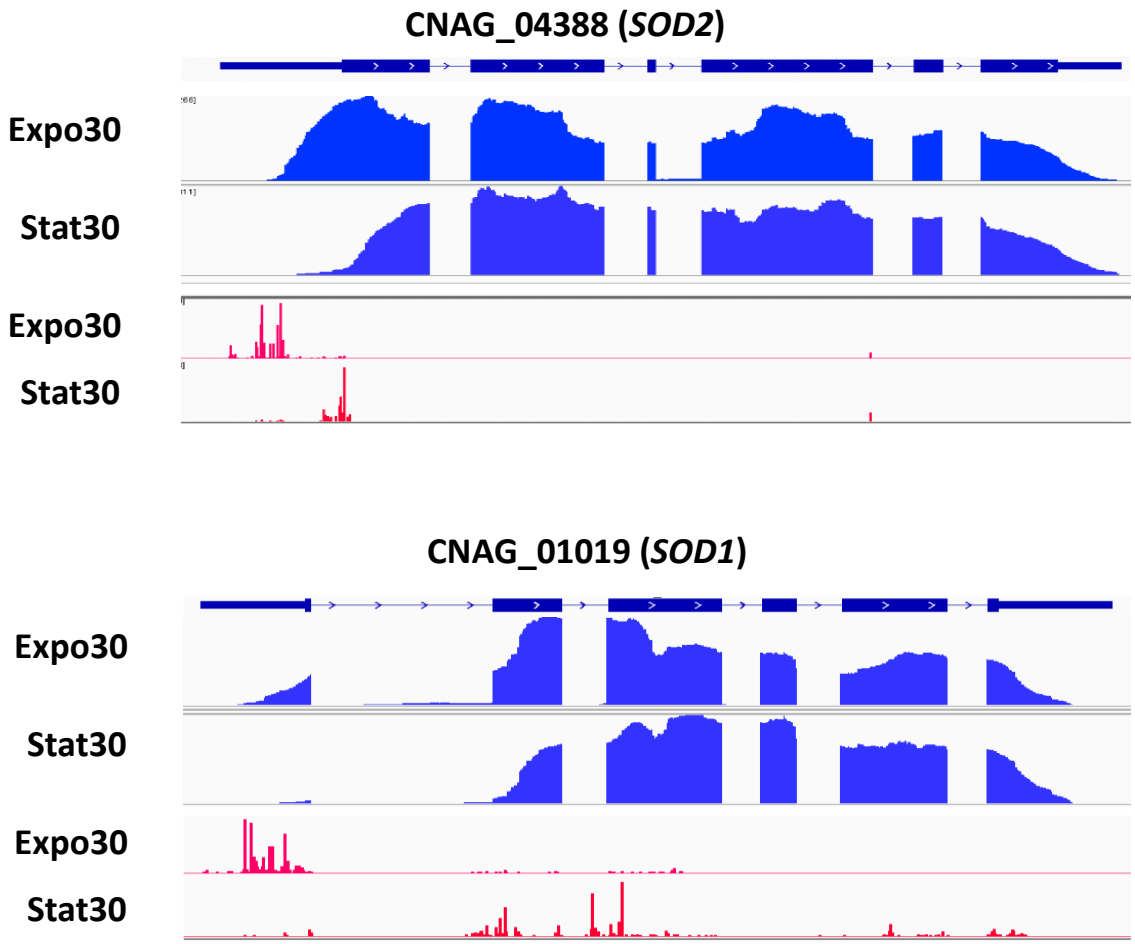

**Supplementary Figure S2 : TSS cluster sizes obtained after directly merging the twelve gff files originally generated in *Wallace et al* (14).**



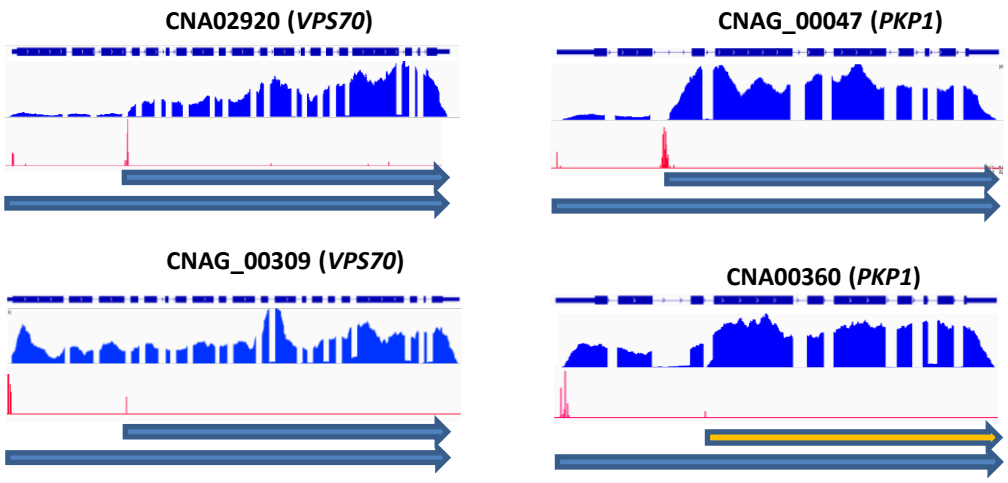

**Supplementary Figure S3. TSS clusterization using *C. neoformans* TSS-seq data obtained in three other growth conditions also revealed two types of TSS (broad and sharp).** (A) Stationary phase 30°C (B) exponential phase 37°C, and (C) stationary phase 37°C**.** Left. Distribution of the TSS clusters in terms of cluster width (Size) and shape index (SI) is represented as a 2D density plot in which Size and SI distribution of TSS clusters display a bimodal pattern. Corresponding histogram of Size and SI are projected on x-axis and y-axis, respectively. Right. Histogram (bold blue line) and density plot (bold black line) of the first principal component (PC1) of Size and SI. Two subpopulations of PC1 are detected by statistical test using "MixtureInf" R package and represented as the theoretical density plot (dotted black line).



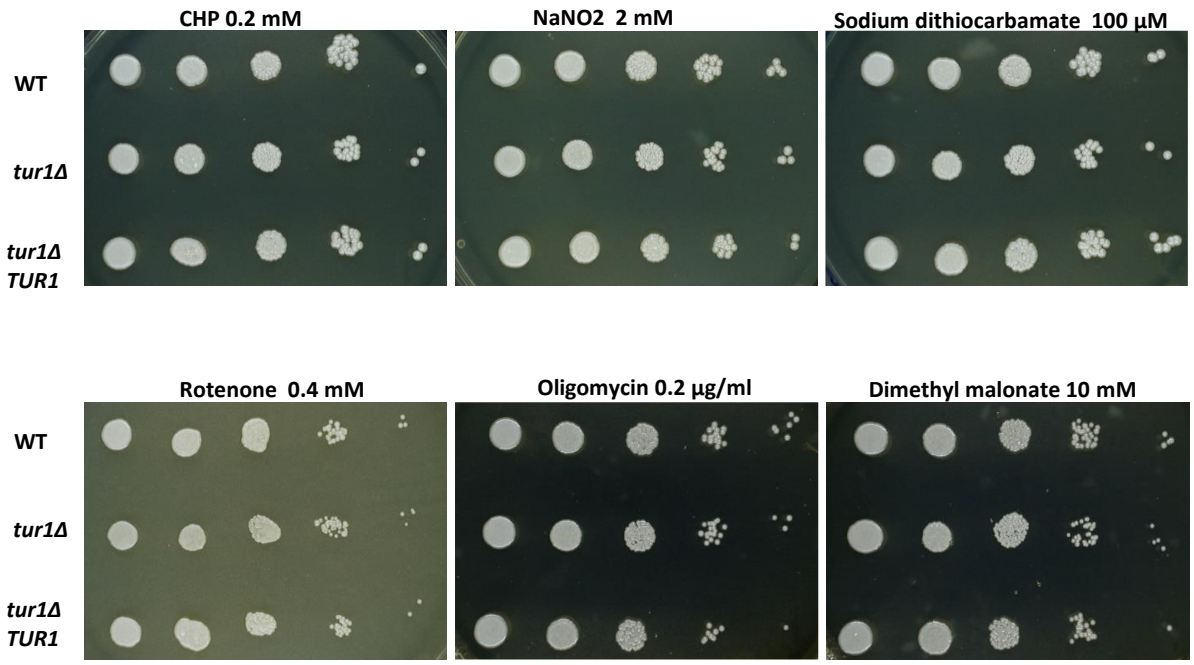

**Supplementary Figure S4. Similar motif enrichment was observed associated with broad and sharp clusters generated from each set of TSS-seq data**. Enrichments and positions relative to the major position within the TSS cluster of 6 detected motifs within the TL sequence.




**Supplementary Figure S5. altTSS usage regulation of *SOD1* and *SOD2* genes in exponential and stationary phases.** IGV visualization of RNA-seq (upper tracks) and TSS-seq (bottom tracks) at the CNAG_01019 (*SOD1*) and CNAG_04388 (*SOD2*) loci when *C. neoformans* cells were cultivated at 30°C under exponential or stationary phase, respectively.




**Supplementary Figure S6. Apparent "non-conserved" altTSS can be conserved between *C. neoformans* and *C. deneoformans*.** IGV visualization of RNA-seq and TSS-seq at the *VPS70* and *PKP1* loci of *C. neoformans* and *C. deneoformans* obtained when cells were cultivated at 30°C under stationary phase. At the *C. deneoformans VPS70* gene, the same altTSS observed in *C. neoformans* is visible albeit very poorly used. At the *C. deneoformans PKP1* gene, an altTSS is also visible but not at the same position as the one observed in *C. neoformans*. Interestingly, in both cases, the regulation of altTSS is reversed between species.




**Supplementary Figure S7. Sensitivity of the *tur1Δ* mutant to oxidative agents.** 10-fold serial dilutions (starting with $10^7$ cells/mL) of the wildtype (WT), *tur1* mutant (*tur1Δ)* and complemented strain *(tur1Δ TUR1)* cells were spotted on YPD medium containing Cumen hydroperoxide (CHP), $NaNO_2$, Sodium dithiocarbamate, rotenone, Dimethyl malonate, or oligomycin. Plates were incubated 3 days at 30°C before being photographed

**Supplementary tables:**

https://drive.google.com/drive/folders/1rEq51tcgo9lmBJ52mDDSduTQkNd9rlpE?usp=sharing

**Supplementary Table S1. List of primers used in this study**

**Supplementary Table S2. List of strains used in this study**

**Supplementary Table S3. List of plasmids used in this study**

**Supplementary Table S4. List of genes with altTSS clusters identified in *C. neoformans* and *C. deneoformans.***

**Supplementary Table S5. List of genes for which the condition dependent regulation of altTSS usage results in the production of protein isoforms with different subcellular localization as predicted by Deeploc2.0**

**Supplementary Table S6. N-terminomic analysis. (A) Peptides identified by N-terminomic. (B) Peptides specific of mRNAs resulting from altTSS usage identified by N-terminomic.**

**Supplementary Table S7. List of genes regulated by altTSS in response to change in temperature or/and growth phase.**

**Supplementary Table S8. List of genes regulated by the exponential to stationary phase transition in the WT and *tur1Δ* in *C. neoformans***

**Supplementary Table S9. List of genes with at least 2 TSSs and at least 20 TSS reads in all four samples (WT E30, WTS30, *tur1Δ* E30 and *tur1Δ* S30)**