Comprehensive profiling of the fission yeast transcription start site activity during stress and media response

Malte Thodberg[1], Axel Thieffry[1], Jette Bornholdt[1], Mette Boyd[1], Christian Holmberg[2], Ajuna Azad[1], Yun Chen[1], Karl Ekwall[3], Olaf Nielsen[2#], Albin Sandelin[1#]

[1] The Bioinformatics Centre, Department of Biology and Biotech Research and Innovation Centre, University of Copenhagen, Denmark
[2] Cell cycle and genome stability group, Department of Biology, University of Copenhagen, Denmark
[3] Department of Biosciences and Nutrition, Karolinska Institute, Sweden

Abstract

Fission yeast, *Schizosaccharomyces pombe,* is an attractive model organism for transcriptional and chromatin biology research. Such research is contingent of accurate annotation of transcription start sites (TSSs). However, comprehensive genome-wide maps of TSS and their usage across commonly applied laboratorial conditions and treatments for *S. pombe* are lacking. To this end, we profiled TSSs activity genome-wide in *S. pombe* cultures exposed to heat shock, nitrogen starvation, peroxide and two commonly applied media, YES and EMM2, using Cap Analysis of Gene Expression (CAGE). We show that CAGE-based annotation of TSS is substantially more accurate than existing PomBase annotation, and that *S. pombe* core promoters commonly initiate transcription bidirectionally at the edges of nucleosomes, but do not show sharp and dispersed TSS distributions as higher eukaryotes. Our data recapitulate known *S. pombe* stress expression response patterns, but also identify stress-responsible alternative TSSs. Importantly, growth media change induces changes of similar magnitude as some stressors. Integration of genetic variation data shows a link between nucleosome occupancy and genetic variation, and that the proximal promoter region is genetically diverse between *S. pombe* strains compared to the TATA and TSS regions. Our data constitute a central resource for *S. pombe* gene regulation research.

Introduction

Yeast cells have been central models for understanding eukaryotic gene regulation. Historically, baker's yeast (*Sacharomyces cerevisiae*) has been the unicellular model of choice, but the remotely related fission yeast *Schizosaccharomyces pombe* has in many ways turned out to be a more relevant model for mammalian cells (1). The architecture of *S. pombe* and human chromosomes are similar, featuring large repetitive centromeres and regions of RNAi-dependent heterochromatin (2). *S. pombe* also utilizes histone modifications and chromatin remodelling enzymes similar to multicellular eukaryotes (2). On the other hand, both yeast types have highly related signal transduction pathways responding to environmental stress (3).

A central challenge for cells is to react on changing environments through the regulated activation of genes. The most commonly studied states in yeast cells are environmental stress through chemicals (e.g. peroxide, sorbitol, cadmium), physical conditions (e.g. heat), or changes in available nutritional substances (nitrogen starvation, change of growth media, etc.). This is highly relevant for unicellular organisms, where all cells must directly sense and respond to a wide range of different environmental states. Previous work on stress response in *S. pombe* and *S. cerevisiae* has focused on the distinction between a general environmental stress response (Core Environmental Stress Response, CESR) versus specific response to each individual type of stress (Specific Environmental Stress Response, SESR). CESR is comprised of metabolic genes related to carbohydrate metabolism and genes involved in protein stability such as anti-oxidants, proteases and heat shock proteins, while SESR is comprised of genes with more specific functions related to the given type of stress (4-7) .

In *S. pombe*, the timing and maintenance of the CESR is controlled by a conserved signal transduction pathway that ultimately activates the Sty1 protein kinase(*sty1*)(8), a member of the stress-activated MAP kinase family (SAPK family), which also includes HOG1 from *S. cerevisiae* and human *MAPK14* and cJun-N terminal kinase (*MAPK8*). Sty1 in turn activates key transcription factors such as Atf1, Pap1 and Hsf1 (8-10).

Stress-specific gene regulation has been studied by comparing genome-wide changes in expression profiles in wild-type cells and mutants lacking key signaling components or transcription factors (4, 11). While some types of stress (for example heat shock) induce a quick and transient response, peroxide and alkylating agents induce a more lengthy response (4). Furthermore, both stress exposure and environmental cues such as nutritional composition modulate the growth rate and size at which fission yeast cells enter mitosis(12). This size regulation occurs partly through the Sty1 SAPK and also via the TOR (target of rapamycin) pathway (13, 14). Extreme nutritional stresses, in particular nitrogen starvation, induces cells to enter sexual development, a process which is intertwined with the CESR (15).
Due to its importance in chromatin biology research and stress response, it is surprising that no comprehensive maps of transcription start sites (TSSs) across different states have been reported for *S. pombe*. Accurate maps of TSSs and their activity have been instrumental in understanding gene regulation and core promoter activity, as well as the evolution of gene regulation in mammals, birds, insects and plants (e.g. (16-21)). Such data sets have also been highly beneficial in deciphering the relationship between chromatin and transcription initiation (e.g. (22-25)).

Cap Analysis of Gene Expression (CAGE), based on sequencing the first 20-30 bp of capped, full-length cDNAs (26) is arguably the most used technique for locating TSSs and their transcriptional activity genome-wide. Such so-called CAGE tags can, when mapped to a

reference genome, identify TSSs with single nucleotide resolution and quantify their level of transcription, as the number of CAGE tags mapping to a location will be proportional to the concentration of the originating mRNA (Fig.1A). CAGE tags can thus be used for expression profiling, but on TSS rather than gene level. In this sense, it is complementary to RNA-seq, which has the advantage that splicing can be assessed but, on the other hand, is not precise in locating TSSs. Previous studies have shown that RNA-seq and CAGE have comparable accuracy in terms of expression profiling (27). Because CAGE is not contingent on existing gene models, it can both refine existing gene models and locate novel TSSs, both within and outside of known genes. In eukaryotes, alternative TSS usage is common (28): for instance, in human and mouse >50% of genes have two or more distinct TSSs, many of which are used in a tissue- or context-specific manner(21). Such alternative TSSs are interesting as they may confer additional, independent regulatory inputs for genes and/or change the protein-coding potential of the resulting mRNA, for instance by excluding exons coding for protein domains (29, 30).

Since stress response is highly studied in *S. pombe*, having accurate and genome-wide TSS maps for stress states would be beneficial for understanding their gene regulation and associated processes. Here, we used CAGE to define a TSS atlas of unprecedented resolution and scope for *S. pombe*, across three environmental stressors (heat shock, nitrogen starvation and peroxide stress) and two commonly used growth media, Edinburgh Minimal Medium (EMM2) and Yeast Extract Medium with supplements (YES). We show that this atlas substantially expands and refines current state-of-the-art *S. pomb*e TSS annotation, allowing for more detailed interpretation of a range processes, including nucleosome positioning, histone modification and transcription levels. Additionally, our CAGE-based annotation allows the analysis of stress-specific and shared stress response regulation and growth media adaptation, and enables refined analysis of *S. pomb*e genetic variation data.

Materials and Methods

Culture conditions and RNA preparation

Triplicate cultures of the wild-type strain h^{-S} (972) were grown at 30°C in either EMM2 (Edinburgh minimal medium) or YES (yeast extract with supplements) to a density of 5 X 10^6 cells/ml (31). For nitrogen starvation, cells were transferred by vacuum filtration from EMM2 to EMM2 without ammonium chloride and incubated for 16 hours at 30°C. Heat shock was imposed by transferring YES cultures to 39°C for 15 minutes. Oxidative stress was inflicted by treating YES cultures with 0.5 mM H_2O_2 for 15 minutes at 30°C. Total RNA was extracted from 10^8 cells. In brief, cells were harvested by 5 minutes of centrifugation at 3000 rpm. Pellets were resuspended in TES (10 mM TrisHCl pH7.5, 10 mM EDTA and 0.5% SDS) and transferred to 65˚C preheated acidic phenol-chloroform (Sigma P-1944). After 1 hour of incubation at 65˚C with mixing every 10 minutes, RNA was extracted with chloroform-isoamyl alcohol (Sigma C-0549), ethanol precipitated and re-suspended in water. All RIN-values were above 8.8. As CAGE requires a certain amount of input RNA concentration, RNA from nitrogen-starved cultures was additionally up-concentrated by vacuum centrifuge (reported RIN-scores are after up-concentration). See Table.S1 for an overview of libraries.

CAGE analysis and mapping

CAGE libraries were prepared from the 15 yeast cultures as in (26), using 5 µg total RNA as starting material. Libraries were run individually with the following four barcode: #2(CTT), #3(GAT), #4(CACG) and #8(ATC). Prior to sequencing, four CAGE libraries with different barcodes were pooled and applied to the same sequencing lane. Sequencing of the libraries

was performed on a HiSeq2000 instrument from Illumina at the National High-Throughput DNA Sequencing Centre, University of Copenhagen. To compensate for the low complexity in 5′ ends of the CAGE libraries, 30% Phi-X spike-ins were added to each sequencing lane, as recommended by Illumina. CAGE reads were assigned to their respective originating sample according to identically matching barcodes. Using the FASTX Toolkit (v0.0.13), assigned reads were i) 5'-end trimmed to remove linker sequences (10+1 bp to account for the CAGE protocol G-bias at the first 5' base), ii) 3'-end trimmed to a length of 25 bp and iii) filtered for a minimum sequencing quality of 30 in 50% of the bases. Trimmed reads were mapped using Bowtie(32) (Version 0.12.7) with parameters –t –best –strata –v –k 10 –y –p 6 –phred33-quals –chunksmbs 512 –e 120 –q –un to ASM294v2.26. To obtain bp resolution CAGE Transcription Start Sites (CTSSs), the number CAGE tag 5'-ends were counted for each genomic position. CTSSs coordinates were offset by 1 bp to account for the G-bias trimming. Only chromosomes I, II and III were used for analysis. Mapping statistics are available in Table.S1.

Quantification and annotation of TSSs and genes

For each library, CTSS counts were normalized into tag-per-million (TPM) values and the sum of TPM-values across all libraries were calculated for each base pair. Nearby CTSSs within 20 bp of each other were merged in Tag Clusters (TCs). Only CTSSs passing a TPM-threshold (0.04516233) was considered; this threshold was chosen to maximize the total number of TCs across the genome (as implemented in the CAGEfightR R-package (version 0.3, under preparation)). Expression was quantified as the number of tags in each TC for each sample. TCs with less than 1 TPM in any 3 libraries were discarded.

Annotations (ASM294v2.26) was downloaded from PomBase(33, 34) as a GTF-file and imported as a TxDb object(35). Regions were extracted the transcriptsBy-family of functions and genes using the genes-function. TCs were annotated by their overlap to gene annotation categories defined from PomBase using transcript models using the hierarchical models shown in Fig.2A. Promoters were defined as the +/-100 bp region around a PomBase TSS, while Proximal was defined as -1000 to -100 from annotated TSS. 5'UTRs, 3' UTRs and CDS regions were defined as in PomBase. Exonic refers to non-coding exons, including lncRNAs. Antisense was defined as intragenic but on opposite strand. TCs were aggregated to gene-level by summing all counts within or -1000 upstream of annotated genes. In case a TC overlapping more than one gene, the gene with the nearest annotated TSS was chosen.

To make TSSs obtained from PomBase comparable to CAGE-defined TSSs in terms of expression, the same approach was used: CTSSs were quantified around PomBase TSSs (first bp of the 5'-UTR) +/- 100 bp and were then filtered by discarding TCs with less than one 1 TPM in 3 libraries. CTSSs, TCs and transcript models were visualized using CAGEfightR and the Gviz R-package (36) . TCs, and bp-resolution tracks for each condition are deposited in the GEO database (GSE110976).

Biological sequence analysis

The getSeq function from the BSgenome package was used to extract sequences from the reference genome. Di- and trinucleotide frequencies were counted using vmatchPattern function and sequence logos were made using the ggseqlogo package(37). TATA-Box and INR-motifs were obtained from JASPAR(38) via the JASPAR2016 R-package(39). The motifScanScores from the seqPattern R-package (http://bioconductor.org/packages/seqPattern/) was used to scan for motif occurrences: 60% of maximum score considered a hit in Fig.2D and average percentage in Fig.3B. CpG frequency (CpG dinucleotides per bp) was calculated in a -100/+100 window around TSS peaks.

Average CAGE, MNase-seq, H3K4me3 ChIP-seq and NET-seq signal calculation
MNase-seqand H3K4me3 ChIP data was obtain from (40, 41) (accession numbers GSM1374060 and GSE49575, respectively). Only wildtype strains were used in both cases. Net-seq data was obtained from(42) (BAM-files were generously supplied by the authors). The GenomicAlignments package(35) was used to import and calculate the average signal. Only NET-seq signal for TSSs corresponding to annotated promoters were plotted, sorted by the amount of sense-direction activity in the nucleosome deficient region (Fig.3 and Fig.S3C). In all cases, data was imported with rtracklayer and coverage calculated with the coverage function from IRanges(35). NET-seq and CAGE profiles were plotted after removing the 1% most highly expressed sites, to remove the influence of a few extreme outliers.

Comparison with mammalian TSSs
We obtained CTSSs from five control replicates for CAGE assays on mouse lung from(43). TSSs were obtained from the CTSSs using CAGEfightR (as above). As these libraries were prepared in the same lab as the *S. pombe* libraries presented here and analyzed using the same pipeline, they are highly comparable. TC widths were calculated as the distance between the two positions in the TC corresponding to the 10th and 90th percentile of the pooled TPM across all libraries.

Differential expression analysis and GO-term enrichment
All analysis was carried out using the edgeR(44) and limma(45) R-packages. The same analysis was run on TSS- and gene-level. Initial normalization using the TMM-method from edgeR revealed a consistent shift in the overall distribution of expression values in the Nitro-EMM libraries. This could either reflect a large global shift in transcriptional output or an artifact of the additional RNA-purification steps in these samples (see above). Subsequent analysis using only TMM-normalization revealed a large number of up-regulated compared to down-regulated TSSs as well as a global shift in fold changes for highly expressed genes. For the final analysis we therefore chose the conservative approach of also normalizing expression using the cyclicloess-method prior to voom-transformation to model the mean variance relationship inherent to count data. Heat maps are based on the dual-normalized expression values. MDS plots were made using the plotMDS-function using pairwise distance between the top 1000 genes.

We modeled expression using ~treatment+medium with Ctrl-EMM libraries as reference levels. We used robust estimation at the empirical Bayes stage and used the treat method to test against a minimum required $\log_2$ fold change of 0.5 (corresponding to ~41% higher of lower expression between states). Resulting *P*-values were corrected for multiple testing using Benjamini-Hochberg. Fig.S5-6 show diagnostic plots for the differential expression analysis on TSS and gene level.

Differential transcript usage was assessed using the diffSplice-approach, with the Simes-method to aggregate results at the gene level. Resulting *P*-values were corrected for multiple testing using Benjamini-Hochberg.

GO-terms were downloaded from the PomBase, and the kegga-function was to test for enrichment, with mean expression used as the covariate. Resulting *P*-values were corrected for multiple testing using Benjamini-Hochberg.

Comparison with previously defined CESR response and alternative 5' UTR TSSs
CESR genes were downloaded from(4) (http://bahlerweb.cs.ucl.ac.uk/projects/stress/); 68/62 genes for the CESR Up/Down response could be matched to the CAGE data based on gene

IDs. Log_2 fold changes estimated with limma-voom were plotted for each CESR gene. Genes utilizing alternative 5'-UTR usage was obtained from(46). 24/28 genes could be matched by gene IDs. As the original data set did not provide coordinates for novel TSSs, we compared our gene-level differential transcript usage results from limma-voom with diffSplice to investigate agreements.

Motif enrichment analysis in promoter regions

Promoter region DNA sequences for TSSs were extracted as sequences -300 to +100 of TSS peaks (Fig.S1A) via BSgenome and getSeq as described above. Homer(47) was used to detect de-novo enriched motifs in each differentially expressed set compared to a background set consisting of TSSs not part of any differentially expressed set (default setting except size=given, S=9, p=40 and mset=yeast). *de-novo* motifs were compared to Homer's database of *S. cerevisiae* motifs. The top 5 enriched *de-novo* motifs with a foreground set frequency of more than 0.9% were plotted using ggseqlogo (see above). *S. cerevisiae* motifs was also used to perform known motif enrichment for the same sets (Dataset.S1).

Single Nucleotide Variation analysis

Genetic variation data from(48) was downloaded from EnsemblFungi. Only Single Nucleotide Variations ("TSA=SNV;Jeffares_SNPs") were used for analysis. IRanges(35) were used to calculate the average number of SNVs per basepair in promoter regions (TSS peaks -300:+100 base pairs). GC-content and TATA-box positions were calculated as above. Background single nucleotide variation (SNV) frequencies were calculated as the total number of SNVs overlapping CDS regions divided by the total size in basepairs of all CDS regions. The number of SNVs within promoter regions as a function of differential expression status and annotation category was modeled using negative binomial (NB) regression (due to overdispersion compared to a Poisson distribution) using the glm.nb function from the MASS R-package (49). Exponentiated effect estimates and confidence intervals were extracted using the tidy-function from the broom R-package (https://CRAN.R-project.org/package=broom). Including average TSS expression in the model did not affect have any major effect on the results (data not shown).

Protein domains in genomic space

Pfam protein domains(50) were obtained from PomBase. Amino acid positions were multiplied by 3 and mapped back to genome using the mapFromTranscripts function from GenomicFeatures(35). Domain disruptive TSSs were defined as TSSs that either: a) overlapped a domain or b) were downstream of a domain within the same gene.

Western Blot analysis

Protein extracts from a Cds1-HA tagged strain and a wild type control treated as above (Ctrl – EMM, Ntro-EMM, Heat-YES and Ctrl-YES for both strains) were made by the TCA precipitation method. Extracts were fractionated by SDS-PAGE. After semi-dry transfer to a nitrocellulose membrane, Cds1 was detected using primary anti-HA (12CA5) monoclonal antibodies and HRP-coupled secondary anti-mouse antibodies.

Results

An TSS atlas for *S. pombe*

We prepared three biological replicates of *S. pombe* cultures growing under five different conditions: cells growing exponentially in EMM2 medium (designated as "Ctrl-EMM"), cells growing exponentially in YES ("Ctrl-YES"), heat-shocked cells growing in YES ("Heat-YES"), nitrogen starved cells growing in EMM2 ("Nitro-EMM") and peroxide-stressed cells growing in

YES ("H_2O_2-YES") (Fig.1B). After RNA purification, we constructed three CAGE libraries per condition, corresponding to each biological replicate (the CAGE method is outlined in Fig.1A). The average yield was 16.9 million uniquely mapping tags to the ASM294v2 assembly (Table.S1).

Mapped CAGE tags closely located to each other on the same strand were grouped into Tag Clusters (TC), similarly to (51). The expression of TCs across libraries were assessed as the number of CAGE tags within each cluster, normalized by sequencing depth in the respective library as Tags Per Million (TPM). We retained only TCs that had at least 1 TPM in three or more libraries. After cutoffs, we detected 13.003 TCs. For simplicity, we will refer to these TCs as "CAGE-defined TSSs". For each CAGE-defined TSSs, we identified the "TSS peak" as the single bp position with the highest TPM value across all libraries (Fig.S1A).

In order to interpret the activity of CAGE-defined TSS, it is useful to annotate them in reference to known transcript and gene models. As an example, Fig.1C shows the annotated TSSs of the *use1* and *pep7* genes, which are positioned bidirectionally ~1kb apart. Both annotated TSSs were detected by CAGE in all conditions, albeit consistently slightly downstream of the annotated TSSs (see below for a systematical investigation of this phenomenon). However, an unannotated TSS for *pep7* was detected 106 bp downstream of the annotated TSSs. Unlike the annotated TSS, this TSS was only strongly expressed in H_2O_2-YES libraries. It is thus an H_2O_2-YES-specific novel alternative TSS.

Fig.1D shows a more complex locus, where three annotated genes (*nhe1, prl19 and SPNCRNA.602*) occupy the same region, on both strands. Here, all annotated TSSs were detected by CAGE, and were primarily expressed in Nitro-EMM cells. However, two downstream alternative TSSs for *nhe1*, both placed in the extended 5' UTR, were detected, one which was exclusively used in Nitro-EMM and one that was highly used in all conditions.

CAGE-defined TSSs refine PomBase TSS annotations

To systematically investigate the overlap with existing gene models, each CAGE-defined TSS was annotated into one of nine categories, depending on their overlap with transcript models as defined by PomBase(33, 34) (Fig.2A). CAGE-defined TSS most commonly overlapped PomBase-annotated promoter regions (+/- 100 bp from the start of PomBase annotated TSSs); 34% of all CAGE-defined TSS overlapped these regions. CAGE-defined TSSs also commonly overlapped the proximal upstream region of genes (up to 1kbp upstream of PomBase annotated TSSs, *14%*), 5'-UTRs (*13%*), and coding exons (CDS) (*18%*). Notably, CAGE-defined TSSs overlapping annotated promoters or 5'-UTRs were generally more highly expressed than CAGE-defined TSSs within other categories. This reflects observations in mouse and human (e.g. (43, 51)) and indicates that novel TSSs on average are less expressed, although clear exceptions to this rule exist (e.g. see Fig.1C).
We noted that while CAGE-defined TSSs often overlapped the regions around PomBase-annotated gene TSSs, the peak (the position having the most CAGE tags) of the corresponding CAGE tag cluster was often located slightly downstream of PomBase TSSs (exemplified in Fig.1C). In order to investigate whether this was a general trend, we plotted the average CAGE signal around PomBase-defined TSSs. This confirmed the existence of a small but consistent shift between CAGE and PomBase TSSs, where the majority of CAGE tags fell 50-75 bp downstream of annotated TSSs (Fig.2B and Fig.S2A). This systematic difference could be either due to systematic overestimation of transcript lengths in the PomBase annotation, or a systematic bias in our CAGE data. We reasoned that the most correct TSS set should better recapitulate known biology of gene regulation at DNA-sequence and chromatin levels.

Previous experiments have shown that TATA and INR core promoter elements have a strong positional preference around TSS (~-32/-25 for TATA and immediately at the TSS for INR) when present (52, 53). A sequence logo (54) constructed by aligning the DNA sequence +-50bp around all peaks of CAGE-defined TSS (Fig.2C, left) showed enrichment of T/A-rich sequences at -32 to -25 and a strong pyrimidine-purine dinucleotide enrichment at -1/+1. These correspond to the TATA box and the central part of the INR element, and this pattern was highly similar to corresponding analysis in human and mouse (21, 53). While this pattern was present across CAGE-defined TSS in all annotation categories as defined above (Fig.S2B), it was not detected when constructing a sequence logo based on PomBase-defined TSSs (Fig.2C, right). Analogously, when scanning the +/-500 bp sequence around CAGE TSS peaks with experimentally established TATA and INR motifs from the JASPAR database(38), strong TATA and INR enrichments were observed at the expected locations (Fig.2D, left). Corresponding analysis based on PomBase TSSs only showed mild enrichments at corresponding locations (Fig.2D, right).
Dinucleotide analysis in the same regions showed a 150-bp cyclical pattern with higher AT/TA frequency downstream of CAGE TSSs (Fig. 2E, left). A similar, but weaker, pattern was present when centering on PomBase TSSs (Fig. 2E, right). Nucleosome structure in *S. pombe* is highly DNA sequence dependent, determined by relative frequencies of CG and TA dinucleotides (55-57). We reasoned that if the 150-bp cyclical AT-rich patterns reflected nucleosomal placement, then nucleosome alignment should be stronger and better when positioned by CAGE data compared to PomBase. To test this, we assessed nucleosomal positions by micrococcal nuclease digestion followed by sequencing (MNase-seq) in Ctrl-EMM cells (40). Indeed, MNase-seq signal displayed the same cyclic pattern downstream of CAGE-based TSSs as predicted from diucleotide frequencies, where the amplitude, but not phase, correlated with the CAGE expression strength and the position of the +1 nucleosome edge was adjacent to the CAGE TSS (Fig. 2F, left). H3K4me3 ChIP-seq(41) signals from corresponding cells(41) showed similar patterns (Fig. 2F, right).
Together these observations (sequence content, nucleosome occupancy and histone modifications) strongly indicate that CAGE-defined TSSs are more accurate than current PomBase annotation. Thus, our data not only increase the number of known TSSs, and thereby promoters of *S. pombe*, but additionally refines the positional accuracy of existing genome-wide data sets. The latter is important for any study relying on nucleotide-resolution TSS-anchored analysis and visualization; including profiling of histone marks, transcriptional activity, initiation and elongation. Our data thereby constitute a valuable new source of information for investigating *S. pombe* transcriptional regulation and chromatin biology.

S. pombe promoters initiate transcription bidirectionally, but lack broad TSS distributions

One of the main advantages of *S. pombe* compared to *S. cerevisiae* as a model organism is its higher similarity to mammalian genomes in terms of gene regulation and chromatin biology. It was therefore important to assess whether core promoter features in mammalian genomes also were present in *S. pombe*. In particular, we compared the shape of TSS distributions, the prevalence of bidirectional transcription initiation and the relation between TSSs and nucleosomes in *S. pombe* vs. mammalian genomes.

In mammals, promoters can be distinguished by their distribution of TSS into “broad” and “sharp“ classes. Sharply defined TSS distributions, where CAGE tags are focused on one or a few nearby nucleotides, are associated with TATA boxes and cell/tissue-specific transcription,

while broader TSS distributions are often overlapping CpG islands and are associated with more ubiquitous expression (reviewed in(58)). Notably, in mammals, broad TSS distributions are more common than sharp. We measured the brodness of nearby TSSs as the width of CAGE tag clusters (see Fig.S1A) in *S. pombe* and a set of representative mammalian CAGE libraries produced and processed in the same way as the *S. pombe* set presented here (five pooled *M. musculus* lung CAGE libraries (43)).

While both *S. pombe* and *M. musculus* had CAGE tag cluster widths spanning from a single to hundreds of bps, *S. pombe* did not show the same bimodal distribution of tag cluster widths found in mouse; in particular, 30-50 bp wide tag clusters were common in *M. musculus* but not in *S. pombe*. This observation was correlated with the underlying DNA sequence content: while *M. musculus* tag cluster width as previously reported was positively correlated with CpG dinucleotide content (Fig.3A), this relation was not present in *S. pombe*, perhaps related to the general low CpG dinucleotide content in *S. pombe* promoters (Fig.S3A). Similarly, in *M. musculus* we observed a clear tendency for stronger TATA-box matches upstream of sharp TSS distributions (width 1-5 bp); this pattern was absent in *S. pombe* (Fig.3B); on the other hand, *S. pombe* tag clusters had, regardless of their width, as high average TATA match scores as sharp *M. musculus* tag clusters. This is likely partially due to the overall higher T/A content in *S. pombe* promoters. INR motif scores showed no relation to tag cluster width in either species (Fig.S3B). Overall, our results imply that *S. pombe* does not display broad TSS distributions as frequently as mammals, which may be due to the overall higher TA content and, in extension, higher TATA-box content.

Recent studies have shown that transcription in mammals initiates at both edges of the nucleosome deficient regions (NDR), producing mRNAs in the sense direction and typically short and exosome-sensitive RNAs in the reverse direction. In mammals, such reverse strand transcripts have been termed PROMPTS or uaTSS (59, 60); similar observations has been made in *S. cerevisiae*, where the transcripts are called CUTs or SUTs (reviewed in (61)).To see whether *S. pombe* promoters shared this property, we analyzed recent native elongating transcript sequencing (NET-seq) data from wild type cells (42). NET-seq signal anchored on CAGE-defined TSS showed bidirectional transcription, consistent with transcription initiation at the edges of the NDR, mirroring the properties of PROMPTs in mammals (**Fig.3C-D**). The only exception to this was in cases where the region upstream of the TSS also had high NET-seq signal on the sense strand (presumably due to upstream TSSs on the same strand) (Fig.S3C). Because of our high CAGE sequencing depth, we could often detect PROMPTs upstream of gene TSSs: 36% of highly expressed TSSs corresponding to annotated PomBase promoters had an additional TSS detected within 250 bp on the opposite strand (see also see upstream reverse strand TSS in Fig.2B).

Thus, the above analyses indicate that *S. pombe* TSSs share some, but not all mammalian properties: *S. pombe* promoters generally support bidirectional initiation, but do not exhibit a clear distinction between broad and sharp TSS distributions.

Identification of stress-responding TSSs

In the above analysis, we treated all TSSs equally, even though they may be ubiquitously expressed across conditions, specifically expressed in one condition, or somewhere between these two extremes. In order to investigate patterns of differential TSS expression between different stress and growth conditions, we first visualized overall differences between libraries using a multidimensional dimensional (MDS) scaling plot (Fig.4A). This showed that cells subjected to the same type of stress and medium, clustered tightly together, indicating low

biological variance within groups and consistent differences between groups. Nitro-EMM libraries, followed by Heat-YES, appeared the most different from Ctrl-EMM. Interestingly, we observed a clear difference between control libraries growing on the different growth media (Ctrl-EMM and Ctrl-YES), which was of comparable magnitude as the difference between Ctrl-EMM and H_2O_2-EMM. Thus, YES medium appears to alter the TSS landscape to an extent comparable to some stressors.

To identify individual TSSs that showed statistically significant changes in expression, we analyzed differential expression on TSS-level using the linear model framework implemented in the R-package limma (45). The advantage of this approach, compared to multiple pairwise tests, is that the individual effect of media and stress treatment can be assessed separately. Because of the high sequencing depth combined with the low variance within groups, we had power to detect even weak levels of differential TSS expression across all ranges of TSS expression. However, even though subtle changes could reliably be detected, we reasoned that changes with a substantial fold change were the most biologically relevant. Thus, to focus our analysis on only the most strongly responding TSSs, we only considered TSSs with a $\log_2$ fold change cutoff of 0.5 (corresponding to ~40% higher or lower expression between conditions) and *FDR*<0.05 to be significantly differentially expressed between two or more states. Using Ctrl-EMM as a baseline, we defined four sets of differentially expressed TSSs responding by either increasing or decreasing expression: $Nitro_{up}/Nitro_{down}$ for nitrogen starvation, $Heat_{up}/Heat_{down}$ for heat shock, H_2O_{2up}/ H_2O_{2down} for peroxide stress and YES_{up}/YES_{down} for YES media (Table.S2). The results recapitulated the MDS plot: the $Nitro_{up}$ and $Nitro_{down}$ sets had the highest number of differentially expressed TSSs, followed by $Heat_{up}/Heat_{down}$, YES_{up}/YES_{down} and $H2O2_{up}/H2O2_{down}$ (Fig.4B) Notably, while YES_{up}/YES_{down} and $Nitro_{up}/Nitro_{down}$ had roughly as many up- as down-regulated TSSs, $Heat_{up}/Heat_{down}$ and H_2O_{2up}/ H_2O_{2down} showed a substantially higher number of up-regulated vs. down-regulated TSSs.

Differentially expressed TSSs across the three stress conditions overlapped with each other to a much higher degree than with YES_{up}/YES_{down} (Fig. 4C-D), indicating that a substantial part of stress response was generic, and while YES medium results in the change of expression of a comparable number of TSS, the response is distinct from stress response. Overall, 6375 of the detected TSSs (49%) were differentially expressed in at least one stress condition (excluding YES response), indicating that almost half of *S. pombe* TSSs are directly or indirectly involved in stress response.

To investigate how the above results correspond to known fission yeast biology, we related differential expression to existing Gene Ontology (GO) terms. Since GO-term annotations are defined at gene level, we measured gene expression by summing CAGE expression contribution from all CAGE-defined TSSs within a gene on the same strand, and used these expression values to perform gene-wise test for differential gene expression, in the same way as the TSS-level expression analysis above (Fig.S4).

We obtained GO terms from PomBase, and used these to perform enrichment tests for GO-terms within each of the sets of differentially expressed genes (see Fig.4A for the top 10 enriched GO-terms in each category, and Table.S3 for the complete results). Most differentially expressed gene sets were enriched for expected terms, i.e. H_2O_{2up} was enriched for “cellular response to oxidative stress”, $Heat_{up}$ was enriched for “protein folding”, $Nitro_{up}$ was enriched for “autophagy” and YES_{down} was enriched for “thiamine biosynthetic process”. In addition to these stress-type specific GO-terms, we also observed shared GO-term enrichments: H_2O_{2up} and $Heat_{up}$ genes were enriched for “cellular detoxification” while $H2O2_{down}$, $heat_{down}$ and $nitro_{down}$ were enriched for “ribosome biogenesis”. Interestingly, this shared ribosome biogenesis GO

term over-representation was not due to a shared set of downregulated genes across sets. Instead, the gene sets that were responsible for the enrichment were distinct for two groups of stressors (Fig.4B): while H_2O_{2down} and $Heat_{down}$ primarily included nuclear ribosomal-related genes, $Nitro_{down}$ primarily included cytosolic ribosomal genes (based on molecular function GO terms). Ribosome biogenesis and maturation is a complex process that begins in the nucleus and continuous in the cytosol (62), and general downregulation or destabilization of ribosomal protein RNAs and/or proteins have been reported previously in stress response (63-65). The split between cytosolic and nuclear ribosome genes for H_2O_{2down} and $Heat_{down}$ vs $Nitro_{down}$ may reflect the elapsed duration of the stress (a scale of minutes for H_2O_2-Yes and Heat-YES and hours for Nitro-EMM libraries, see Fig.1B). This may also represent a preference to downregulate synthesis of new ribosomes in the nucleus (H_2O_{2down} and $Heat_{down}$) as opposed to inactivation of maturation or replacement of existing cytosolic ribosomes ($Nitro_{down}$).

We observed an enrichment of GO-terms related to meiosis in genes upregulated after nitrogen starvation (e.g. “conjugation with cellular fusion” and “meiosis 1”). This is expected, as sexual development in *S. pombe* is intimately linked to nitrogen starvation (66, 67), and nitrogen starvation in EMM2 medium is particularly efficient in inducing this response. The activation of genes related to sexual development in addition to the stress response explains the overall larger number of differentially expressed TSSs and genes in $Nitro_{up}$/$Nitro_{down}$ compared to the other types of stress.

The above results were in general agreement with a previous study of the *S. pombe* stress response based on microarrays (4), with a focus on defining a set of CESR genes. Despite that we had a smaller range of stressors, we observed substantial overlap between the upregulated genes in our set and the CESR-upregulated set: 38% (26/68) CESR genes were in both H_2O_{2up}, $Heat_{up}$, $Nitro_{up}$ sets, while only seven CESR genes were not in any of our differentially up-regulate sets. Only 6/62 downregulated CESR genes were found in H_2O_{2down}, $Heat_{down}$ and $Nitro_{down}$ (Fig.S7A). This most likely reflects our stricter criteria for differential expression, since only five genes showed a different direction of differential expression in H_2O_{2up}/H_2O_{2down}, $Heat_{up}$/$Heat_{down}$ compared to (4) (Fig.S7B). A larger number of genes in $Nitro_{up}$/$Nitro_{down}$ showed a different direction of change; this is not surprising as (4) did not include nitrogen starvation as a stressor.

Since we found that almost half of all TSSs were changing due to one or more stressors, we also wished to identify TSSs and genes that were not significantly changing across any stress conditions. Such TSSs/genes would be useful as candidates for “housekeeping TSSs”, and thereby references for e.g. qPCR analysis. To select such TSSs or genes we required an FDR >0.05 in any comparison to Ctrl-EMM (using an F-test against a $\log_2$ fold change different from 0), resulting in 19% (2530) of TSSs and 16% (950) of genes. These genes were enriched for GO-terms related to protein transport and the Golgi-apparatus, e.g. “intracellular protein transport” and “ER to Golgi vesicle-mediated transport” **(Fig.S7C**). A list of these genes and TSSs are shown in Table.S4.

Stress responding TSSs provide insights into *S. pombe* gene regulation

An outstanding question is how the various stress responses identified above are regulated. Because CAGE data enables both accurate localization of TSSs, and precise quantification of

TSS usage across conditions, it is ideally suited for prediction of transcription factor binding sites (TFBSs) responsible for stress-specific gene regulation in upstream promoter regions.

We used HOMER (47) to identity enriched DNA motifs in the promoters around TSSs up- or down-regulated by each stressor (defined as the -300 to +100 bp around TSS peaks). Because there is no comprehensive resource of transcription factor motifs in fission yeast, we performed the analysis *de novo* (identifying over-represented motifs with no prior information). To determine whether any of the *de novo* motifs resembled known motifs, we compared them to HOMER's database of motifs from *S. cerevisiae*.

As with GO terms, we observed motifs enriched in specific types of stress as well as motifs common across all stressors. In many cases, the enriched patterns reflected known *S. pombe* or *S. cerevisiae* biology (see Fig.6 for a summary and full analysis at Dataset.S1). $Heat_{up}$, H_2O_{2up} and $Nitro_{up}$ promoters all shared enrichment for the Atf1 motif (homolog of S. cerevisiae CST6); Atf1 is widely regarded as the central regulator of the CESR (4, 68). $Heat_{up}$ promoters were highly enriched for a HSF1-like motif, while H_2O_{2up} promoters were enriched for a PAP1-like motif (*S. cerevisiae* homolog of YAP1). Heat shock factors and PAP1 are well-known transcription factors associated with the specific response to heat shock and peroxide stress, respectively, in both *S. pombe* and *S. cerevisae*(69) (70)

$Nitro_{down}$ promoters were enriched for the SFP1 motif, while H_2O_{2down} was enriched for the similar SUM1 motif (while SFP1 was not found *de-novo* in $heat_{down}$, when doing a complimentary enrichment analysis using known *S. cerevisiae* motifs, SFP1 was enriched both $heat_{down}$, $H2O2_{down}$ and $Nitro_{down}$). SFP1 was previously implicated in regulating the synthesis of ribosomal gene expression (71), supporting the consistent enrichment of the "ribosome biogenesis" GO-term in corresponding down-regulated genes described above.

As discussed above, $Nitro_{up}$ were enriched for genes associated to sexual development in *S. pombe*. In agreement, the Ste11 motif (AZF1 homolog) (67) was highly enriched in $Nitro_{up}$ as well as a GATA binding site (GAT1 homolog). Ste11 and the GATA transcription factor Gaf1 have previously been shown to be regulated in a coordinated fashion in *S. pombe* (72).

YES_{up} and YES_{down} were enriched for different motifs compared the three types of stresses, reflecting an overall different regulation of this response. We are not aware of any comprehensive study of TFBSs involved in the response to YES medium, but our analysis suggests a possible role for several novel *S. pombe* TFBSs similar to known TFBSs in *S. cerevisae*. For example, YES_{up} were enriched for a PDR1 motif and a general forkhead motif. PDR1 is annotated as specific to *S. cerevisiae*, while at least three forkhead-domain-containing genes exist in *S. pombe* (*fhl1, mei4 and sep1*), which may have a role cell adjustment to rich media. The Sep1 transcription factor controls transcription of a cluster of genes peaking in expression during mitosis (73), including *plo1*, encoding the *S. pombe* homolog of POLO kinase. Interestingly, Plo1 was previously shown to enable the stress response pathway to modulate the timing of mitotic commitment in response to nutrient status (74).YES_{down} promoters were enriched for GCR1, STB4 and PHD1 motifs. GCR1/2 were previously shown not to have *S. pombe* homologs, but regulate the glycolytic pathway in *S. cerevisiae* (75).

Single nucleotide variation around *S. pombe* TSSs

TFBSs underlying the enriched motifs above were primarily detected in the -50 to -150 bp region relative to TSSs, suggesting that much of the regulatory signals for stress response depend on this region. We were interested in how this over-representation related to genetic

variation across *S. pombe* populations. To investigate this, we overlapped the regions the CAGE-defined TSSs with available single nucleotide variations (SNVs), generated from comparison of multiple global *S. pombe* strains (48) (Fig.7A). Overall, and consistent with (48), we observed an higher number of SNVs in intergenic regions (i.e. upstream of TSSs) compared to within genes (downstream of TSSs). However, the precise location of the CAGE-defined TSSs revealed additional SNV patterns.

First, downstream of TSSs, SNV abundance was highly correlated to nucleosome occupancy (as defined by MNase-seqdata, see Fig.2F) and anti-correlated to CG dinucleotide content (Fig.7A). It is interesting to note that while CG dinucleotides in general are hyper-mutable, the TA-rich regions within the first 6-7 nucleosomes downstream of the TSS showed higher average SNVs than then CG-rich linker regions. A possible underlying mechanisms is that DNA with high nucleosome occupancy have higher mutation rates becasue nucleosomes block the access of the DNA repair machinery, as observed in human (76).

Second, SNV abundance was highest -400 to -100 bp upstream of TSSs, but rapidly declined towards two minima corresponding to the TATA box position at around -30 and the TSS at +1 (See zoom-in panel of Fig.7A). The low frequency of observed SNVs around the TATA-box and TSS was similar the SNVs fequency in coding exons. These results indicate that while the regions immediately surrounding the TSS and TATA-box is similar between *S. pombe* strains, the region upstream of the TSS is more variable: this is somewhat unexpected as this region overlapped most predicted TFBSs. To investigate whether this pattern was related to differential expression, we examined whether the regions around differentially expressed TSSs had more or less SNVs vs. non-differentially expressed TSSs. Specifically, we modeled the number of SNVs in the -300 to +100 region around the TSSs as a function of i) their differential expression ($Heat_{up}$/$Heat_{down}$, YES_{up}/YES_{down}, etc.; as in Fig.4A) and ii) their overlap with gene annotation (annotated TSS, 5' UTR, coding sequence, etc., as in Fig.2A) using negative binomial regression (Fig.7B). The advantage with this approach is that the effect of TSS location and differential expression status can be separated. Compared to a baseline consisting of non-differentially expression TSSs near annotated promoters, intergenic TSSs (annotated as proximal or intergenic) had more SNVs, while intragenic SNVs (intron, antisense, 3'-UTR and CDS) had less (in agreement with Fig.6A). Correcting for these gene structure effects through the regression model, we found that some differentially expressed sets of TSSs were associated with substantial increases in the number of SNVs. In particular, YES_{up} and YES_{dowm} promoters had an increase in SNVs bigger than the effect of being in an intergenic region. H_2O2_{up}, $Heat_{down}$ and $Nitro_{up}$ promoters were also enriched for SNVs, while H_2O_{2down}, $Heat_{up}$ and $Nitro_{down}$ where neither enriched nor depleted. We speculate that this increase in SNVs in differentially expressed promoters in particular could be related to positive selection or genetic drift of SNVs in these regions (see Discussion).

Stress-specific activation of alternative TSSs

Alternative TSSs are common in multicellular organisms (28, 29, 77). Principally, genes may utilize alternative TSSs for at least three reasons: i) having multiple TSSs and thereby promoters may increase the overall expression of the gene in question. ii) If alternative TSS are located within genes, their usage may produce RNAs that exclude coding exons. This may result in truncated proteins, which may have a functional impact. iii) Multiple TSSs offer higher regulatory flexibility, where two TSSs may respond to different stimuli or context due to different transcription factor binding sites around them. Notably, none of these scenarios are mutually exclusive, and combinations of scenarios ii and iii are particularly interesting.

To our knowledge there have been no systematic attempts to profile alternative TSS usage in *S. pombe* genome under different conditions. We reasoned that since CAGE can detect and

quantify both known and novel TSS, our data set is well suited for detecting differential TSS usage within genes across stress conditions.

First we investigated the occurrence of multiple TSS inside genes across all conditions. Around 54% of genes had >1 TSSs (Fig.8A), indicating that multiple TSSs per gene is a widespread phenomenon in *S. pombe*. For many genes, a single dominant TSS might be expressed at much higher levels than other TSS(s) within the same gene. We therefore identified the dominant TSS within each gene (defined as the TSS with the highest overall expression across all libraries) and analyzed where dominant and non-dominant TSSs were located respective to annotation (Fig.8B). For genes having >1 TSSs, the dominant TSS most often overlapped the annotated TSS region or 5' UTRs, much like the location of TSSs with only a single (and thereby dominant) TSS. Non-dominant TSSs were most commonly occurring in CDS or proximal upstream regions, a pattern consistent with the overall expression of TSSs in the same regions (Fig.2A).

Next, we wanted to investigate how common it is that multi-TSS genes use different TSS under different conditions? For each gene, we analyzed whether any of its TSSs showed a different response across conditions compared to the rest of the TSSs in the same gene (using the diffSplice approach from limma, see Methods). Nitro resulted in the highest number of genes showing differential TSS usage, followed by heat, H_2O_2 and YES, mirroring the overall gene expression results (Fig.8C-D). Overall, 77% percent of genes with >1 TSS showed differential TSS expression compared to the rest of TSS within the gene in at least one condition. It is thus common for genes to have at least two TSSs that are regulated differently in *S. pombe*. There was no clear preference for differential TSS usage in any specific part of genes (Fig.8D), although TSSs in coding regions were less likely to be in this category. This pattern was consistent across all conditions.

Many genes showed both differential expression and differential TSS usage as defined above (Fig.S8C), indicating that these processes are coupled. Adding an additional filter to only analyze the top most expressed TSSs within genes (only TSSs making up at least 10 % of total gene expression in at least 3 libraries), showed the same overall pattern (Fig.S8A-B and Fig.S8D).

In several cases, differential TSSs usage detected by CAGE was verifying previous single-gene studies. We show two such cases in Fig.8E-F. We detected two TSSs in the 5' UTR of the Superoxide Dismutase 1 (sod1) gene where most the upstream TSS showed little change between conditions, but the downstream TSS (~1kb downstream of the annotated sod1 TSS) showed high expression in all EMM2 media conditions and low expression in YES media conditions (Fig.8E). Ref.(78) previously identified two *sod1* mRNA isoforms: a shorter transcript that was gradually replaced by a larger when cultures growing in YEP entered stationary phase. The lengths of both transcripts detected in their Northern blot correspond to the distance between the two TSSs detected by CAGE, suggesting that a switch from EMM2 to YES can create a similar change (Fig.8E, inset).

For the GTPase activating protein 1 gene (*gpa1*) we found two strongly expressed CAGE-defined TSSs, where one was expressed across all conditions but further induced by nitrogen starvation, and the other 500bp downstream TSS was expressed only by nitrogen starvation (Fig.8F). These observations are consistent with the findings of Obara *et al* (79), who detected two *gpa1* specific transcripts, respectively, 2kb and 1.6 kb long, where the shorter was only observed in nitrogen starved cells capable of generating a pheromone signal. Hence, even though we did not analyse pheromone-stimulated cells, the CAGE data allowed us to detect a

similar induction in nitrogen-starved cells, emphasizing the overlapping signalling pathways between these two stimuli (80).

We were also interested in the specific case where differential TSS usage could lead to an altered protein product, corresponding to scenario ii). We reasoned the most dramatic effect on the final protein product of a gene would be the exclusion or truncation of DNA sequence coding for protein domains by the use of alternative TSSs. To investigate this, we mapped annotated *S. pombe* protein domains from PomBase onto the genome to determine whether a given TSS overlapped a domain or was located upstream of one, potentially giving rise to, respectively, a domain truncation or exclusion. We refer to such a TSS as being ”disruptive” in terms of protein domain content: 64% of CDS TSSs, 56% of intronic TSSs and 75% of 3’-UTR TSSs were classified as disruptive using this approach. The average expression level of disruptive TSSs was low, and disruptive TSSs generally made low contributions to total gene expression (Fig.9A). We detected 52 dominant disruptive TSSs, and 478 disruptive TSSs which contributed more than 10% of the expression of their host gene in at least three libraries. 356 of these TSSs were differentially used in at least one condition (Fig.9B). This indicates that while alternative TSS usage affecting protein domain inclusion is relatively rare, there are clear cases of this phenomenon. Below we present three such examples:

Fig.9C shows the *arg11* gene, which had one TSS in the 5’-UTR and another in the coding region downstream of a NAT- and amino acid kinase domain. Cells from all treatments expressed both TSSs, but the upstream TSS was greatly preferred in Nitro-EMM. If transcripts from both TSSs were translated, Nitro-EMM would express higher concentrations of the protein harboring the NAT- and amino acid kinase domains.

Fig.9D shows the *fus1* gene, containing two highly expressed TSSs. While the annotated TSS was weakly expressed across all treatments, Nitro-EMM expressed a 5’-UTR TSS and Heat-YES expressed a bidirectional TSS inside the coding sequence downstream of GTPase-binding domain. If the sense transcript is translated, most *fus1* transcripts expressed in Heat-YES would lack the GTPase domain.

Fig.9E shows the *cds1* gene, which had two highly expressed TSSs: one in the 5’ UTR which was active across all conditions, and one in the fourth protein coding exon, expressed only in Nitro-EMM and Heat-YES. The latter TSS would produce an RNA that cannot encode the forkhead domain encoded in the upstream exons. A previous study showed that this domain is important for the activation of Cds1 through binding to Mrc1 (81).

An importance caveat is that it is not possible based on CAGE data alone to determine if the transcript produced by an alternative TSS is translated into a stable protein product. While CAGE can be used as a hypothesis generator, protein-based experiments are needed to verify candidates for protein-altering alternative TSSs. Illustrating this caveat, a western blot for Cds1 after heat shock and nitrogen starvation did not detect a protein product corresponding to the Heat-YES/Nitro-EMM specific downstream TSS discussed above, despite the high RNA levels detected by CAGE (Fig.S9).

Discussion

Here we have used CAGE to define a comprehensive atlas of TSSs in *S. pombe* and their activity across a range of different growth conditions and stresses.

We compared our TSSs atlas to existing *S. pombe* annotation from PomBase, which is primarily based on RNA-seq. We show that our TSS atlas can refine the location of known TSSs, including TSSs for protein coding genes, as well as more lowly expressed non-coding transcripts. This is in line with previous work (46), showing pervasive transcription of many unstable transcripts, including many anti-sense transcripts as exemplified in Fig.1D.

Accurate TSSs are essential for studying chromatin biology, for which *S. pombe* is a widely used model organism. CAGE-defined TSSs had expected core promoter motifs at expected spacing, and showed a much better agreement with MNase-seq and ChiP-seq data for relating transcription initiation with nucleosome positioning and histone modifications respectively. We showed that *S. pombe* chromatin biology shares many features with mammals, including the positioning of TSSs adjacent to nucleosomes at both strands, supporting bidirectional transcription.

The only genome-wide TSS experiment previously reported for *S. pombe* is, to our knowledge, the single replicate CAGE experiment presented in (82). Compared to this study, we do not only have a much larger number of libraries (15 CAGE libraries) which allows for the assessment of variability, but our analysis shows that each of our individual libraries is of better quality, with a higher number of mapped reads and less reads mapping to rRNA (Table.S1). More importantly, our TSS atlas covers multiple commonly used experimental states, allowing for analysis of regulatory elements responsible for changes in gene expression.

Our differential expression analysis on gene and TSS level shows that nearly half of TSS are affected by stress response, and their expression is consistent with previous microarray studies (4). The overlap is noteworthy, since our study only measures steady state RNA levels at a single time point following stress application, and thus cannot detect genes showing complex responses over time. Interestingly, we also observe a large transcriptional effect of YES compared to EMM2 media, comparable in overall magnitude to some stresses. A similar phenomenon was reported in (46). The implication of this is that the two most commonly used growth media for *S. pombe* are not interchangeable, and thus, biological conclusions based on experiments performed on YES growing cells cannot necessarily be extrapolated to EMM2 conditions and vice versa.

This is also the first dedicated genome wide study of differentially expressed alternative TSSs in *S. pombe* across stress conditions. We show that differential TSS usage is widespread in the *S. pombe* genome: 54% of genes utilized more than one TSS, and 77% percent of these showed differential alterantiv TSS usage in at least one condition. The large number of alternative TSSs reported here disagrees with a previous effort of detecting genes with alternative TSS in their 5' UTRs using RNA-seq (46), which identified only 30 genes. Using our data, we could confirm 23/28 of these cases. The discrepancy in numbers is likely due to technical issues: since RNA-seq is based on random fragmentation, it is challenging to identify distinct 5' ends of full-length RNAs, as opposed to a dedicated technique like CAGE that specifically selects for them through cap-trapping. It is thus likely that *S. pombe* uses a large repertoire of alternative TSSs. Because multi-exonic genes, required for the generation of alternative splice isoforms, are uncommon in *S. pombe* (only 19.5% percent of genes have >2 exons), our data suggests that alternative TSSs are a greater source of transcript diversity than alternative splicing in *S. pombe*, unlike in mammals where the two processes seem be used to a similar extent (28, 83).

We show that most differential TSS usage, where two or more TSSs in the same gene show different transcriptional patterns, involves TSSs located upstream of the annotated TSS or within the 5'-UTR. Different regulation between such TSSs across conditions likely does not

change the protein product of the resulting transcript, but may confer regulatory flexibility because the two regions can evolve independently.

Binding of transcription factors to sites proximal to TSSs is key process in gene regulation, and our TSS atlas allows for linking between accurate promoter regions and differential expression. Using such region, we identified likely TFBSs involved in the stress response, ranging from well-known stress-associated transcription factors to possible candidate transcription factors that may regulate the response to YES media. We reasoned that the regulatory importance of these promoter regions may be related to their evolution; indeed, we observed a strong conservation of the immediate TSS and TATA-box regions (comparable to that of coding sequence), while regions upstream of the TATA-box were more genetically variable between *S. pombe* populations. A simple explanation for this observed pattern is a lack of negative selection in the predominantly intergenic upstream regions, leading to a higher rate of genetic drift between *S. pombe* populations. However, we also observe that differentially expressed TSSs in many cases show more genetic variability, in particular YES-responding TSSs. Therefore, an alternative model is that while the relatively constrained CDS and core promoter regions are under negative selection, positive selection may be acting on genetic variants affecting transcription factor binding upstream of the TATA-box, giving rise to variation in gene regulation between *S. pombe* strains, rather than variation in the structure of genes themselves. Because YES-media is richer than EMM2, we speculate this higher diversity may be related to gradual adaptation to different nutritional environments for the different *S. pombe* strains, while the more acute stress (e.g. heat shock) response patterns may have a higher conservation of their regulatory elements.

While most of the work in this paper has focused on genome-level interpretation of the TSS atlas, we believe one of the main uses will be as a resource for detailed investigation of individual genes and processes. To make the TSS atlas as easily available as possible to researchers, it will be made publically available via the PomBase genome browser.

Acknowledgements
Work in the AS laboratory was supported by grants from the Lundbeck Foundation and the Novo Nordisk Foundation. Work in ON's laboratory was funded by the Villum Foundation. Work in the KE laboratory was supported by the Swedish Research Council and Cancerfonden

Author contributions
MT, AT, AA, YC, AS analyzed data. JL, MB made CAGE libraries. ON, CH made *S. pombe* cultures and Western blotting. MT, KE, AN, AS interpreted results. MT, ON, AS wrote the paper with input from all authors.

Data availability
CAGE raw data, as well as processed tag cluster and CTSS datasets have been deposited in the GEO database (accession number: GSE110976).

Supplementary material
Supplementary figures, tables and datasets are available at FigShare:
https://figshare.com/projects/Comprehensive_profiling_of_the_fission_yeast_transcription_start_site_activity_during_stress_and_media_response/30298

Figure legends:

Fig 1. Overview of CAGE experiment

A: Schematic overview of the CAGE protocol. Capped full-length cDNAs are isolated from total RNA through cap-trapping, followed by cleaving off the first 30 nt, which is sequenced and mapped back to the reference genome, identifying TSS locations and their relative usage. Such 30 bp reads are referred to as CAGE tags. Nearby CAGE tags on the same strand are grouped into tag clusters.

B: Experimental design. Five sets of biological *S. pombe* triplicates were prepared, covering three types of stress (heat shock, peroxide stress and nitrogen starvation) and two common growth media: Edinburgh Minimal Media (EMM2) and Yeast Extract (YES). Labels for samples/libraries are shown in the left column. Growth conditions are shown in the right column.

C: Genome-browser example of CAGE defined TSS landscape of the *pep7* and *use1* gene locus. X-axis represents the genome. Panels represent different types of data mapped to the genome. Data on forward strand is indicated by blue color, reverse strand by red. The first and second panels show PomBase gene models. The third to seventh panels show the average CAGE signal from each experimental group (indicated by color code to the left) as bar plots where the Y-axis shows CAGE expression (TPM-normalized number of 5'-ends of CAGE tags at a given bp), where positive values and red color indicate forward strand CAGE-tags, and negative values and blue color indicate reverse strand CAGE-tags. Arrows show the direction of transcription for highly used TSSs. The last panel shows the location of clusters of nearby CAGE tags, where thicker lines indicate the TSS peaks, i.e. the position with the highest CAGE signal. We identified ubiquitously expressed TSS for *use1* and *pep7* located just downstream of the PomBase gene model-defined TSSs, but also a novel *pep7* alternative TSS only active in H_2O_2-treated cells.

D: Genome-browser examples of CAGE defined TSS landscape of the nhe1 locus. Panels are arranged as in panel C. The plot shows a more complex landscape of TSS on both strands, where some correspond to known transcripts, including antisense long non-codig RNAs. See main text for further details.

Fig 2. Properties of CAGE-defined TSS and comparison to PomBase annotation

A: Location of CAGE-defined TSS vs. gene annotation. CAGE-defined TSSs were grouped into one of nine categories based on overlap with PomBase reference annotation (see schematic on top: categories are ordered by priority: in case a CAGE-defined TSS overlapped two or more categories, it was assigned to the one with the highest priority). Color is based on whether regions overlap genes on same strand, antisense strand or are intergenic. Lower left panel shows a bar plot counting the number of CAGE-defined TSS assigned to each category (from a total of 13003 expressed TSSs). Right lower panel shows the distribution of CAGE expression within each category (average TPM across all replicates and conditions) as violin plots, colored as above. Marks within distributions indicate 1st quantile, median and 3rd quantile.

B: Comparison of CAGE and PomBase TSS locations. X-axis show distance relative to annotated PomBase TSSs in bp. Y-axis shows average CAGE signal as TPM across all libraries: CAGE tags on the reverse strand in relation to the annotated TSS are shown as negative values (red), CAGE tags on forward strand are show as positive values (blue). Dotted line indicates PomBase TSS locations.

C: Comparison of CAGE and PomBase promoter patterns. Each panel shows a sequence logo of the +/-50 genomic region around TSS peaks defined by CAGE (left panel) or PomBase annotation (right panel). Y-axis shows information content in bits. X-axis show distance around the TSS in bp. Colors indicate DNA-bases. Insets show a zoom-in of the +/-2 region in each panel.

D: Comparison of CAGE and PomBase core promoter TATA and INR frequency. Sequences of +/- 500 bp around TSSs were scanned for occurrences of TATA (left panel) and

INR (right panel) sites using position weight matrix models, recording hits with >60% of maximal score. X-axis indicates the distance from the TSS, Y-axis shows the frequency of sequences having a motif hit at a given relative location. Color indicates analyses anchored on CAGE-defined (purple) or PomBase-defined (grey) TSSs.

E: Comparison of CAGE and PomBase promoter di-nucleotide frequencies. Y-axis shows the frequency of di-nucleotides (as indicated by color) around TSSs defined by CAGE or PomBase; thick lines show loess regression trend lines. X-axes show distance relative to CAGE-defined TSSs (right panel) or PomBase-defined TSSs (right panel) in bp.

F: Comparison of MNAse-seq and H3K4me3 ChIP-seq signals anchored at CAGE or PomBase TSS. X-axis indicates the distance from TSSs defined by CAGE or PomBase. Y-axis shows the average MNase-seq or H3K4me3 ChIP-seq signal, where larger values indicate highly positioned nucleosomes (left) or modified histones (right). Colors indicate whether signals are anchored on CAGE-defined TSSs (solid lines, stratified by expression quintiles) or PomBase-defined TSSs (dashed).

Figure 3: Analysis of TSS distribution shape and bidirectional transcription in *S. pombe*

A: Relation between CAGE tag cluster width and CpG content. X-axes show CAGE tag cluster width as defined in Fig.S1A. Left Y-axes and black histogram show the distribution of widths in five pooled *M. musculus* lung CAGE libraries (left panel) and this study (*S. pombe,* right panel). Red trend-lines and right Y-axes show the average number of CpG dinuclotides per bp (200 bp centered on TSS peaks). 95% confidence intervals are shown as dotted lines. Only CAGE-defined TSSs corresponding to annotated TSSs are analyzed.

B: Relation between TATA-box occurrence and tag cluster width. X-axes show the -50 to -10 region relative to TSS peaks. Y-axes show the average TATA-box motif match score (0-100% of maximal score). TSSs are stratified by their tag cluster width, indicated by line color. CAGE-defined TSSs from *M. musculus* and *S. pombe* are shown in left and right panel, respectively. Only CAGE-defined TSSs corresponding to annotated TSSs are analyzed.

C: Analysis of bidirectional transcription initiation by NET-seq.

X-axes show distance relative to *S. pombe* CAGE-define TSSs in bp. Y-axis shows average Net-seq signal, in forward (blue) or reverse (red) direction relative to the TSS. Only TSSs corresponding to annotated PomBase TSSs are analyzed and are selected to have low forward NET-seq signal upstream of the TSS: for analysis of remaining TSS, see **Fig.S3C**.

D: MNase-seq signal in bidirectional promoters. X-axes show distance relative to *S. pombe* CAGE-define TSSs in bp. Y-axis shows average MNase-seq signal (similar to Fig.2F) in TSSs corresponding to panel C.

Figure 4. Differential expression of TSS in response to stress and media change

A: Multi-dimensional scaling (MDS) plot of CAGE libraries. X-and Y-axes show the first two MDS dimensions. Each point corresponds to a CAGE library, colored by condition (as in Fig.1B). Axes are scaled as leading log fold changes; the root-mean-squared average of the log fold changes of the top 1000 genes best separating each sample.

B: Differential expression analysis on TSS level. Bar plot shows the number of up-/down-regulated in each set of differentially expressed TSSs, corresponding to effects relative to the reference group (Ctrl-EMM) for the four main effects in the linear model. TSSs were considered significant if $FDR < 0.05$ and $|\log_2 \text{fold change}| > 0.5$, using limma-voom.

C: Expression patterns of differentially expressed TSSs. Heat map rows correspond to all differentially expressed TSSs in panel B. Columns correspond to all CAGE libraries, as indicated by color. Heat map shows the relative CAGE expression ($\log_2$(TPM) scaled to unit mean/variance across rows) for a given TSS and library.

D: Overlap of differentially expressed TSS between treatments. Venn diagram shows the overlap between sets of differentially expressed TSSs as defined in panel B. Number pairs

within Venn diagram areas show the number of up- and down regulated TSSs (upper and lower number, respectively). Number in lower right corner are non-differentially expressed TSSs.

Figure 5. Gene Ontology analysis of differentially expressed genes
A: Enrichment of key Gene Ontology (GO) terms in the stress response. Each panel shows the 10 top over-represented GO-terms for each up- or down-regulated set of genes (as in Fig.S4) as bar charts, sorted by $-\log_{10}$ over-representation *P*-values (X-axis). GO-terms shared between two or more stress effects (H_2O_2, Heat or Nitro) or nutritional changes (Nitro and YES) are indicated by grey bars.
B: Expression of genes with the "ribosome biogenesis" GO term. Heat map rows show CAGE libraries, columns show genes. Cell color indicates relative gene expression as in Fig.4C. Genes are annotated with Cellular Component GO-terms at the top (black square indicates that a certain gene has a given GO-term).

Figure 6. De-novo DNA motif enrichment in promoters of differentially expressed TSSs
A: Enriched motifs in promoters of upregulated TSSs. Columns show the top 5 enriched *de-novo* motifs in the promoter regions of each set of differentially expressed TSSs (as in Fig.4B), where the top motif is the most enriched. Motifs are shown as sequence logos, where Y-axis shows bit score and X-axis shows nucleotide positons in bp. Text above logos show consensus sequence, foreground / background frequency and closest matching known *S. cerevisae* motif (score in parenthesis indicated quality of match).
B: Enriched motifs in promoters of downregulated TSSs. Panels are arranged as in panel A, but show enriched motifs for promoters of downregulated TSSs.

Figure 7. Single Nucleotide Variation (SNV) frequency around CAGE-defined TSSs
A: SNVs around TSSs compared to CG-content and MNase-seq signal. X-axis in all panels shows distance from the CAGE-defined TSS peak in bp. Top panel shows average CG di-nucleotide frequency (grey lines show bp-level averages and black lines show a trend line, as in Fig.2E). Middle panel shows average MNase-seq signal, where larger values correspond to positioned nucleosomes (similar to Fig.2F). The lower panel shows average number of SNVs, with zoom-in panel showing the -300 to +100 promoter region (grey lines show bp-level averages and black lines show a loess regression trend line). Vertical dashed teal lines indicate regions with high nucleosome occupancy. For the zoom-in panel, horizontal dot-dashed line indicates the average number of SNVs in coding exon regions as a reference, vertical dashed lines indicates the -30 position (typical TATA box position) and the 0 position (the TSS peak).
B: Model of number of SNVs in promoters across annotation categories and differential expression sets. Y-axis shows the different modeled effects, composed of different annotation categories (as in Fig.2A) and the differential expression sets (as in Fig.4B, also indicated by background colors). X-axis shows the incidence rate ratio: the estimated fold-change increase in the number of SNVs, compared to a non-differentially expressed TSS near a PomBase-annotated TSS. Points indicate estimates for the incident rate ratio for every modeled effect, whiskers indicate 95% confidence intervals. Whisker color indicates whether the P-value associated with an incident rate ratio is <0.05.

Figure 8. Differential TSS usage within genes.
A: Number of TSSs per gene. X-axis indicates the number of TSSs per gene and Y-axis the number of genes. Colors indicate single-TSS genes (white) and multi-TSS genes (black).
B: TSS structure compared to annotation and expression. X-axis shows the annotation categories of intragenic TSSs (as in Fig.2A), bar plots show the number of TSSs in each category, sub-divided into dominant (most highly expressed, black) and non-dominant TSSs within genes having >1 TSS (grey) and TSSs within single-TSS genes (white).

C: Number of genes showing differential TSS usage compared to differential gene expression. The top panel shows the total number of differentially expressed genes for each effect; genes are divided into those having one or more TSSs. The bottom panel shows the number of genes having at least one TSS with differential expression compared to the rest of TSSs within the gene. X-axes indicate effects and Y-axis indicates the number of genes.
D: Differential TSS usage across different annotation categories. Bar plots show the number of TSS having differential expression compared to the rest of TSSs within the same gene, split by effect (rows, indicated by color), and PomBase annotation overlap (columns).
E: Genome browser example of differential TSS usage in the *sod1* locus. Plot is organized as in Fig.1C. See main text for details.
F: Genome browser example of differential TSS usage in the *gpa1* locus. Plot is organized as in Fig.1C. See main text for details.

Fig9: Candidates for alternative TSSs affecting domain composition.
A: Expression levels of domain composition-disruptive TSSs (TSSs in or upstream of protein domains): Stacked histogram of TSS expression; X axis show $\log_2$(TPM), Y axis show number of TSSs. TSSs are split depending on whether the TSS disrupts domain structure and whether it contributes more than 10% of total gene expression in at least 3 libraries, as indicated by legend.
B: Number of differentially used alternative TSSs potentially affecting domain composition:
X-axis shows the different conditions. Y-axis shows the number of differentially used TSSs with potential to disrupt domain composition. Color indicates whether the TSS contributes more than 10% of total gene expression in at least 3 libraries.
C: Genome browser example of differential TSS usage in the *arg11* locus. Plot is organized as in Fig.1C, but also shows PFAM domains. TSSs downstream of, or overlapping, PFAM domains are highlighted with light red background. See main text for details.
D: Genome browser example of differential TSS usage in the *fus1* locus. Plot is organized as in panel C. See main text for details.
E: Genome browser example of differential TSS usage in the *cds1* locus. Plot is organized as in panel C. See main text for details.

References

1. Russell,P. and Nurse,P. (1986) Schizosaccharomyces pombe and Saccharomyces cerevisiae: a look at yeasts divided. *Cell*, **45**, 781–782.

2. Allshire,R.C. and Ekwall,K. (2015) Epigenetic Regulation of Chromatin States in Schizosaccharomyces pombe. *Cold Spring Harbor Perspectives in Biology*, **7**, a018770.

3. Wilkinson,M.G. and Millar,J.B. (1998) SAPKs and transcription factors do the nucleocytoplasmic tango. *Genes & Development*, **12**, 1391–1397.

4. Chen,D., Toone,W.M., Mata,J., Lyne,R., Burns,G., Kivinen,K., Brazma,A., Jones,N. and Bähler,J. (2003) Global transcriptional responses of fission yeast to environmental stress. *Molecular Biology of the Cell*, **14**, 214–229.

5. Chen,D., Wilkinson,C.R.M., Watt,S., Penkett,C.J., Toone,W.M., Jones,N. and Bähler,J.

(2008) Multiple pathways differentially regulate global oxidative stress responses in fission yeast. *Molecular Biology of the Cell*, **19**, 308–317.

6. Papadakis,M.A. and Workman,C.T. (2015) Oxidative stress response pathways: Fission yeast as archetype. *Crit. Rev. Microbiol.*, **41**, 520–535.

7. Sánchez-Mir,L., Salat-Canela,C., Paulo,E., Carmona,M., Ayté,J., Oliva,B. and Hidalgo,E. (2018) Phospho-mimicking Atf1 mutants bypass the transcription activating function of the MAP kinase Sty1 of fission yeast. *Curr. Genet.*, **64**, 97–102.

8. Wilkinson,M.G., Samuels,M., Takeda,T., Toone,W.M., Shieh,J.C., Toda,T., Millar,J.B. and Jones,N. (1996) The Atf1 transcription factor is a target for the Sty1 stress-activated MAP kinase pathway in fission yeast. *Genes & Development*, **10**, 2289–2301.

9. Toone,W.M., Kuge,S., Samuels,M., Morgan,B.A., Toda,T. and Jones,N. (1998) Regulation of the fission yeast transcription factor Pap1 by oxidative stress: requirement for the nuclear export factor Crm1 (Exportin) and the stress-activated MAP kinase Sty1/Spc1. *Genes & Development*, **12**, 1453–1463.

10. Sakurai,H. and Takemori,Y. (2007) Interaction between heat shock transcription factors (HSFs) and divergent binding sequences: binding specificities of yeast HSFs and human HSF1. *The Journal of biological chemistry*, **282**, 13334–13341.

11. Jia,X., He,W., Murchie,A.I.H. and Chen,D. (2011) The global transcriptional response of fission yeast to hydrogen sulfide. *PLoS ONE*, **6**, e28275.

12. Fantes,P. and Nurse,P. (1977) Control of cell size at division in fission yeast by a growth-modulated size control over nuclear division. *Experimental Cell Research*, **107**, 377–386.

13. Shiozaki,K. and Russell,P. (1995) Cell-cycle control linked to extracellular environment by MAP kinase pathway in fission yeast. *Nature*, **378**, 739–743.

14. Hartmuth,S. and Petersen,J. (2009) Fission yeast Tor1 functions as part of TORC1 to control mitotic entry through the stress MAPK pathway following nutrient stress. *J Cell Sci*, **122**, 1737–1746.

15. Mata,J., Lyne,R., Burns,G. and Bähler,J. (2002) The transcriptional program of meiosis and sporulation in fission yeast. *Nat Genet*, **32**, 143–147.

16. Lizio,M., Deviatiiarov,R., Nagai,H., Galan,L., Arner,E., Itoh,M., Lassmann,T., Kasukawa,T., Hasegawa,A., Ros,M.A., *et al.* (2017) Systematic analysis of transcription start sites in avian development. *Plos Biol*, **15**, e2002887.

17. Rach,E.A., Yuan,H.Y., Majoros,W.H., Tomancak,P. and Ohler,U. (2009) Motif composition, conservation and condition-specificity of single and alternative transcription start sites in the Drosophila genome. *Genome Biol.*, **10**, R73.

18. Haberle,V., Li,N., Hadzhiev,Y., Plessy,C., Previti,C., Nepal,C., Gehrig,J., Dong,X., Akalin,A., Suzuki,A.M., *et al.* (2014) Two independent transcription initiation codes overlap on vertebrate core promoters. *Nature*, 10.1038/nature12974.

19. Schor,I.E., Degner,J.F., Harnett,D., Cannavò,E., Casale,F.P., Shim,H., Garfield,D.A., Birney,E., Stephens,M., Stegle,O., *et al.* (2017) Promoter shape varies across populations and affects promoter evolution and expression noise. *Nat Genet*, **49**, 550–558.

20. Morton,T., Petricka,J., Corcoran,D.L., Li,S., Winter,C.M., Carda,A., Benfey,P.N., Ohler,U. and Megraw,M. (2014) Paired-end analysis of transcription start sites in Arabidopsis reveals plant-specific promoter signatures. *Plant Cell*, **26**, 2746–2760.

21. 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.

22. Rach,E.A., Winter,D.R., Benjamin,A.M., Corcoran,D.L., Ni,T., Zhu,J. and Ohler,U. (2011) Transcription Initiation Patterns Indicate Divergent Strategies for Gene Regulation at the Chromatin Level. *PLoS Genetics*, **7**, e1001274.

23. Forrest,A.R.R., Kawaji,H., Rehli,M., Baillie,J.K., 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.

24. Chen,Y., Pai,A.A., Herudek,J., Lubas,M., Meola,N., Järvelin,A.I., Andersson,R., Pelechano,V., Steinmetz,L.M., Jensen,T.H., *et al.* (2016) Principles for RNA metabolism and alternative transcription initiation within closely spaced promoters. *Nat Genet*, **48**, 984–994.

25. Haberle,V., Li,N., Hadzhiev,Y., Plessy,C., Previti,C., Nepal,C., Gehrig,J., Dong,X., Akalin,A., Suzuki,A.M., *et al.* (2014) Two independent transcription initiation codes overlap on vertebrate core promoters. *Nature*, **507**, 381–385.

26. Takahashi,H., Lassmann,T., Murata,M. and Carninci,P. (2012) 5' end-centered expression profiling using cap-analysis gene expression and next-generation sequencing. *Nat Protoc*, **7**, 542–561.

27. Kawaji,H., Lizio,M., Itoh,M., Kanamori-Katayama,M., Kaiho,A., Nishiyori-Sueki,H., Shin,J.W., Kojima-Ishiyama,M., Kawano,M., Murata,M., *et al.* (2014) Comparison of CAGE and RNA-seq transcriptome profiling using clonally amplified and single-molecule next-generation sequencing. *Genome Research*, **24**, 708–717.

28. 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 Research*, **21**, 1260–1272.

29. Valen,E., Pascarella,G., Chalk,A., Maeda,N., Kojima,M., Kawazu,C., Murata,M., Nishiyori,H., Lazarevic,D., Motti,D., *et al.* (2009) Genome-wide detection and analysis of hippocampus core promoters using DeepCAGE. *Genome Research*, **19**, 255–265.

30. Arner,E., Daub,C.O., Vitting-Seerup,K., Andersson,R., Lilje,B., Drablos,F., Lennartsson,A., Rönnerblad,M., Hrydziuszko,O., Vitezic,M., *et al.* (2015) Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science*, **347**, 1010–1014.

31. Petersen,J. and Russell,P. (2016) Growth and the Environment ofSchizosaccharomyces

pombe. *Cold Spring Harbor Protocols*, **2016**, pdb.top079764.

32. Ben Langmead, Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25–R25.

33. Wood,V., Harris,M.A., McDowall,M.D., Rutherford,K., Vaughan,B.W., Staines,D.M., Aslett,M., Lock,A., Bähler,J., Kersey,P.J., *et al.* (2011) PomBase: a comprehensive online resource for fission yeast. *Nucleic Acids Research*, **40**, D695–D699.

34. McDowall,M.D., Harris,M.A., Lock,A., Rutherford,K., Staines,D.M., Bähler,J., Kersey,P.J., Oliver,S.G. and Wood,V. (2015) PomBase 2015: updates to the fission yeast database. *Nucleic Acids Research*, **43**, D656–61.

35. Lawrence,M., Huber,W., Pagès,H., Aboyoun,P., Carlson,M., Gentleman,R., Morgan,M.T. and Carey,V.J. (2013) Software for computing and annotating genomic ranges. *PLoS Comput. Biol.*, **9**, e1003118.

36. Hahne,F. and Ivanek,R. (2016) Visualizing Genomic Data Using Gviz and Bioconductor. In *Statistical Genomics Methods and Protocols*. New York.

37. Wagih,O. (2017) ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics*, **33**, 3645–3647.

38. Khan,A., Fornes,O., Stigliani,A., Gheorghe,M., Castro-Mondragon,J.A., van der Lee,R., Bessy,A., Chèneby,J., Kulkarni,S.R., Tan,G., *et al.* (2018) JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Research*, **46**, D260–D266.

39. Mathelier,A., Fornes,O., Arenillas,D.J., Chen,C.-Y., Denay,G., Lee,J., Shi,W., Shyr,C., Tan,G., Worsley-Hunt,R., *et al.* (2016) JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Research*, **44**, D110–5.

40. Persson,J., Steglich,B., Smialowska,A., Boyd,M., Bornholdt,J., Andersson,R., Schurra,C., Arcangioli,B., Sandelin,A., Nielsen,O., *et al.* (2016) Regulating retrotransposon activity through the use of alternative transcription start sites. *EMBO reports*, **17**, 753–768.

41. DeGennaro,C.M., Alver,B.H., Marguerat,S., Stepanova,E., Davis,C.P., Bähler,J., Park,P.J. and Winston,F. (2013) Spt6 regulates intragenic and antisense transcription, nucleosome positioning, and histone modifications genome-wide in fission yeast. *Mol Cell Biol*, **33**, 4779–4792.

42. Wery,M., Gautier,C., Descrimes,M., Yoda,M., Vennin-Rendos,H., Migeot,V., Gautheret,D., Hermand,D. and Morillon,A. (2017) Native Elongating Transcript Sequencing reveals global anti-correlation between sense and antisense nascent transcription in fission yeast. *RNA*, 10.1261/rna.063446.117.

43. Bornholdt,J., Saber,A.T., Lilje,B., Boyd,M., Jørgensen,M., Chen,Y., Vitezic,M., Jacobsen,N.R., Poulsen,S.S., Berthing,T., *et al.* (2017) Identification of Gene Transcription Start Sites and Enhancers Responding to Pulmonary Carbon Nanotube Exposure in Vivo.

ACS Nano, **11**, 3597–3613.

44. Robinson,M.D., McCarthy,D.J. and Smyth,G.K. (2009) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Nat genet*, **26**, 139–140.

45. Ritchie,M.E., Phipson,B., Di Wu, Hu,Y., Law,C.W., Shi,W. and Smyth,G.K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, **43**, e47–e47.

46. Wilhelm,B.T., Marguerat,S., Watt,S., Schubert,F., Wood,V., Goodhead,I., Penkett,C.J., Rogers,J. and Bähler,J. (2008) Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature*, **453**, 1239–1243.

47. Heinz,S., Benner,C., Spann,N., Bertolino,E., Lin,Y.C., Laslo,P., Cheng,J.X., Murre,C., Singh,H. and Glass,C.K. (2010) Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell*, **38**, 576–589.

48. Jeffares,D.C., Rallis,C., Rieux,A., Speed,D., Převorovský,M., Mourier,T., Marsellach,F.X., Iqbal,Z., Lau,W., Cheng,T.M.K., *et al.* (2015) The genomic and phenotypic diversity of Schizosaccharomyces pombe. *Nat Genet*, **47**, 235–241.

49. Venables,W.N. and Ripley,B.D. (2002) Modern Applied Statistics with S Springer New York, New York, NY.

50. Finn,R.D., Coggill,P., Eberhardt,R.Y., Eddy,S.R., Mistry,J., Mitchell,A.L., Potter,S.C., Punta,M., Qureshi,M., Sangrador-Vegas,A., *et al.* (2015) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research*, **44**, D279–D285.

51. Boyd,M., Coskun,M., Lilje,B., Andersson,R., Hoof,I., Bornholdt,J., Dahlgaard,K., Olsen,J., Vitezic,M., Bjerrum,J.T., *et al.* (2014) Identification of TNF-α-responsive promoters and enhancers in the intestinal epithelial cell model Caco-2. *DNA Res*, **21**, 569–583.

52. Smale,S.T. and Kadonaga,J.T. (2003) The RNA polymerase II core promoter. *Annual review of biochemistry*, **72**, 449–479.

53. Ponjavic,J., Lenhard,B., Kai,C., Kawai,J., Carninci,P., Hayashizaki,Y. and Sandelin,A. (2006) Transcriptional and structural impact of TATA-initiation site spacing in mammalian core promoters. *Genome Biol.*, **7**, R78.

54. Schneider,T.D. and Stephens,R.M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Research*, **18**, 6097–6100.

55. Lantermann,A.B., Straub,T., Strålfors,A., Yuan,G.-C., Ekwall,K. and Korber,P. (2010) Schizosaccharomyces pombe genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of Saccharomyces cerevisiae. *Nature Structural & Molecular Biology*, **17**, 251–257.

56. Moyle-Heyrman,G., Zaichuk,T., Xi,L., Zhang,Q., Uhlenbeck,O.C., Holmgren,R., Widom,J. and Wang,J.-P. (2013) Chemical map of Schizosaccharomyces pombe reveals species-specific features in nucleosome positioning. *Proceedings of the National Academy of*

Sciences, **110**, 20158–20163.

57. González,S., García,A., Vázquez,E., Serrano,R., Sánchez,M., Quintales,L. and Antequera,F. (2016) Nucleosomal signatures impose nucleosome positioning in coding and noncoding sequences in the genome. *Genome Research*, **26**, 1532–1543.

58. Lenhard,B., Sandelin,A. and Carninci,P. (2012) Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat Rev Genet*, **13**, 233–245.

59. Seila,A.C., Calabrese,J.M., Levine,S.S., Yeo,G.W., Rahl,P.B., Flynn,R.A., Young,R.A. and Sharp,P.A. (2008) Divergent transcription from active promoters. *Science*, **322**, 1849–1851.

60. Preker,P., Nielsen,J., Kammler,S., Lykke-Andersen,S., Christensen,M.S., Mapendano,C.K., Schierup,M.H. and Jensen,T.H. (2008) RNA exosome depletion reveals transcription upstream of active human promoters. *Science*, **322**, 1851–1854.

61. Jensen,T.H., Jacquier,A. and Libri,D. (2013) Dealing with Pervasive Transcription. *Molecular Cell*, **52**, 473–484.

62. Thomson,E., Ferreira-Cerca,S. and Hurt,E. (2013) Eukaryotic ribosome biogenesis at a glance. *J Cell Sci*, **126**, 4815–4821.

63. Marguerat,S., Lawler,K., Brazma,A. and Bähler,J. (2014) Contributions of transcription and mRNA decay to gene expression dynamics of fission yeast in response to oxidative stress. *RNA Biol*, **11**, 702–714.

64. Reja,R., Vinayachandran,V., Ghosh,S. and Pugh,B.F. (2015) Molecular mechanisms of ribosomal protein gene coregulation. *Genes Dev.*, **29**, 1942–1954.

65. Huang,H., Kawamata,T., Horie,T., Tsugawa,H., Nakayama,Y., Ohsumi,Y. and Fukusaki,E. (2015) Bulk RNA degradation by nitrogen starvation-induced autophagy in yeast. *EMBO J*, **34**, 154–168.

66. Sugimoto,A., Iino,Y., Maeda,T., Watanabe,Y. and Yamamoto,M. (1991) Schizosaccharomyces pombe ste11+ encodes a transcription factor with an HMG motif that is a critical regulator of sexual development. *Genes & Development*, **5**, 1990–1999.

67. Xue-Franzén,Y., Kjærulff,S., Holmberg,C., Wright,A. and Nielsen,O. (2006) Genomewide identification of pheromone-targeted transcription in fission yeast. *BMC Genomics*, **7**, 303.

68. Leong,H.S., Dawson,K., Wirth,C., Li,Y., Connolly,Y., Smith,D.L., Wilkinson,C.R.M. and Miller,C.J. (2014) A global non-coding RNA system modulates fission yeast protein levels in response to stress. *Nat Comms*, **5**.

69. Taricani,L., Feilotter,H.E., Weaver,C. and Young,P.G. (2001) Expression of hsp16 in response to nucleotide depletion is regulated via the spc1 MAPK pathway in Schizosaccharomyces pombe. *Nucleic Acids Research*, **29**, 3030–3040.

70. Maeta,K., Nomura,W., Takatsume,Y., Izawa,S. and Inoue,Y. (2007) Green tea polyphenols function as prooxidants to activate oxidative-stress-responsive transcription factors in yeasts. *Appl. Environ. Microbiol.*, **73**, 572–580.

71. Marion,R.M., Regev,A., Segal,E., Barash,Y., Koller,D., Friedman,N. and O'Shea,E.K. (2004) Sfp1 is a stress- and nutrient-sensitive regulator of ribosomal protein gene expression. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 14315–14322.

72. Kim,L., Hoe,K.-L., Yu,Y.M., Yeon,J.-H. and Maeng,P.J. (2011) The fission yeast GATA factor, Gaf1, modulates sexual development via direct down-regulation of ste11+ expression in response to nitrogen starvation. *PLoS ONE*, **7**, e42409–e42409.

73. Rustici,G., Mata,J., Kivinen,K., Lió,P., Penkett,C.J., Burns,G., Hayles,J., Brazma,A., Nurse,P. and Bähler,J. (2004) Periodic gene expression program of the fission yeast cell cycle. *Nat Genet*, **36**, 809–817.

74. Petersen,J. and Hagan,I.M. (2005) Polo kinase links the stress pathway to cell cycle control and tip growth in fission yeast. *Nature*, **435**, 507–512.

75. Kainou,T., Shinzato,T., Sasaki,K., Mitsui,Y., Giga-Hama,Y., Kumagai,H. and Uemura,H. (2006) Spsgt1, a new essential gene of Schizosaccharomyces pombe, is involved in carbohydrate metabolism. *Yeast*, **23**, 35–53.

76. Yazdi,P.G., Pedersen,B.A., Taylor,J.F., Khattab,O.S., Chen,Y.-H., Chen,Y., Jacobsen,S.E. and Wang,P.H. (2015) Increasing Nucleosome Occupancy Is Correlated with an Increasing Mutation Rate so Long as DNA Repair Machinery Is Intact. *PLoS ONE*, **10**, e0136574.

77. Davuluri,R.V., Suzuki,Y., Sugano,S., Plass,C. and Huang,T.H.M. (2008) The functional consequences of alternative promoter use in mammalian genomes. *Trends Genet.*, **24**, 167–177.

78. Lee,J., Kwon,E.-S., Kim,D.-W., Cha,J. and Roe,J.-H. (2002) Regulation and the role of Cu,Zn-containing superoxide dismutase in cell cycle progression of Schizosaccharomyces pombe. *Biochem. Biophys. Res. Commun.*, **297**, 854–862.

79. Obara,T., Nakafuku,M. and Yamamoto,M. (1991) Isolation and characterization of a gene encoding a G-protein alpha subunit from Schizosaccharomyces pombe: involvement in mating and sporulation pathways. *Proc. Natl Acad Sci. USA*, **88**, 5877–5881.

80. Kjærulff,S., Lautrup-Larsen,I., Truelsen,S., Pedersen,M. and Nielsen,O. (2005) Constitutive activation of the fission yeast pheromone-responsive pathway induces ectopic meiosis and reveals ste11 as a mitogen-activated protein kinase target. *Mol Cell Biol*, **25**, 2045–2059.

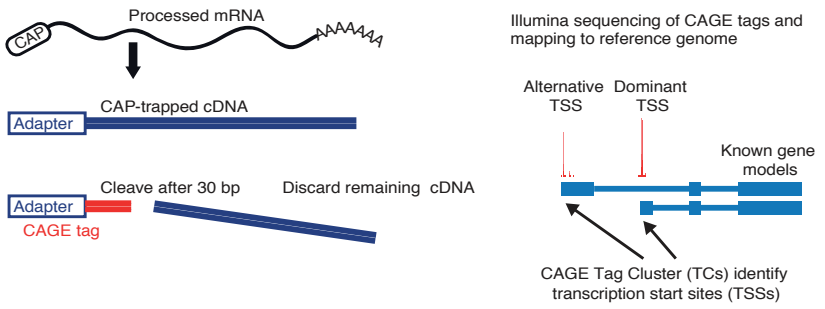
81. Tanaka,K. and Russell,P. (2004) Cds1 phosphorylation by Rad3-Rad26 kinase is mediated by forkhead-associated domain interaction with Mrc1. *The Journal of biological chemistry*, **279**, 32079–32086.

82. Li,H., Hou,J., Bai,L., Hu,C., Tong,P., Kang,Y., Zhao,X. and Shao,Z. (2015) Genome-wide analysis of core promoter structures in Schizosaccharomyces pombe with DeepCAGE. *RNA Biol*, **12**, 525–537.

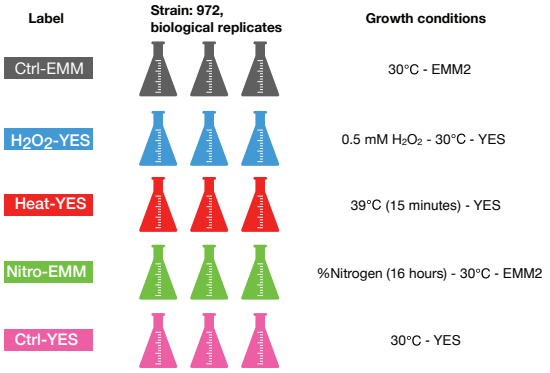
83. Vitting-Seerup,K. and Sandelin,A. (2017) The Landscape of Isoform Switches in Human Cancers. *Mol Cancer Res*, **15**, 1206–1220.

Figure 1

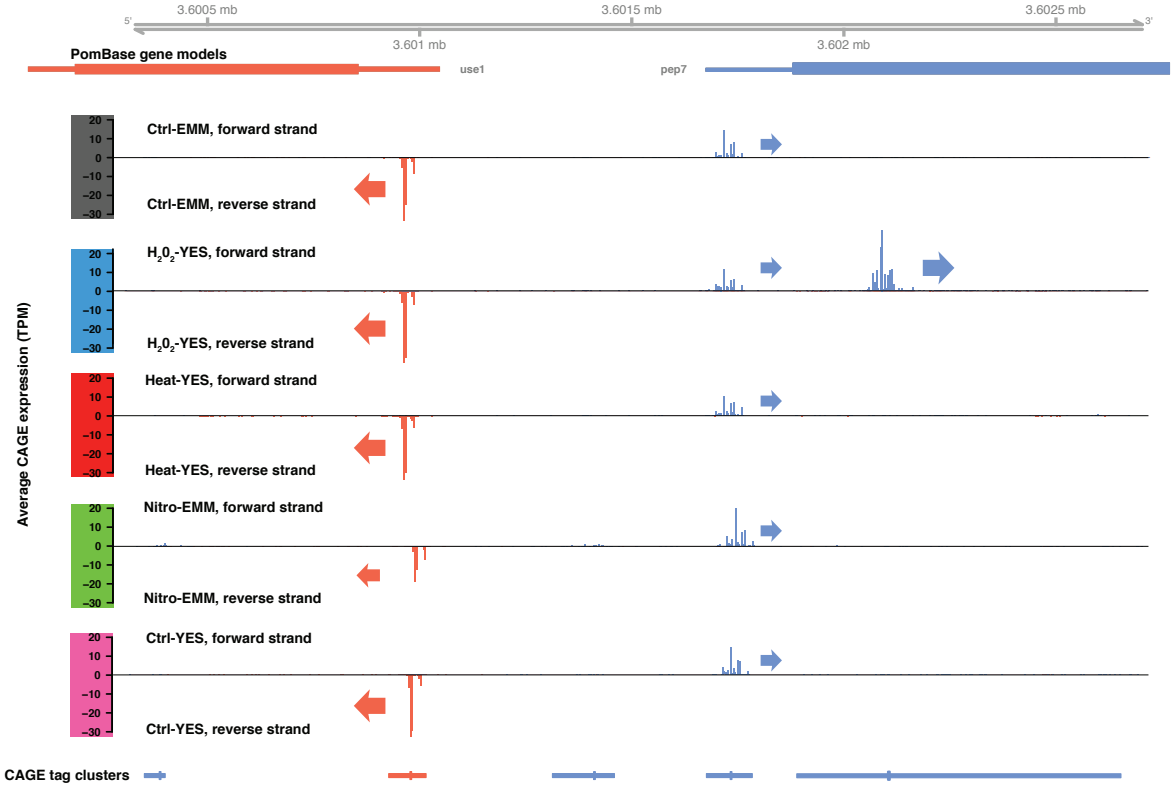
A



B



C



D

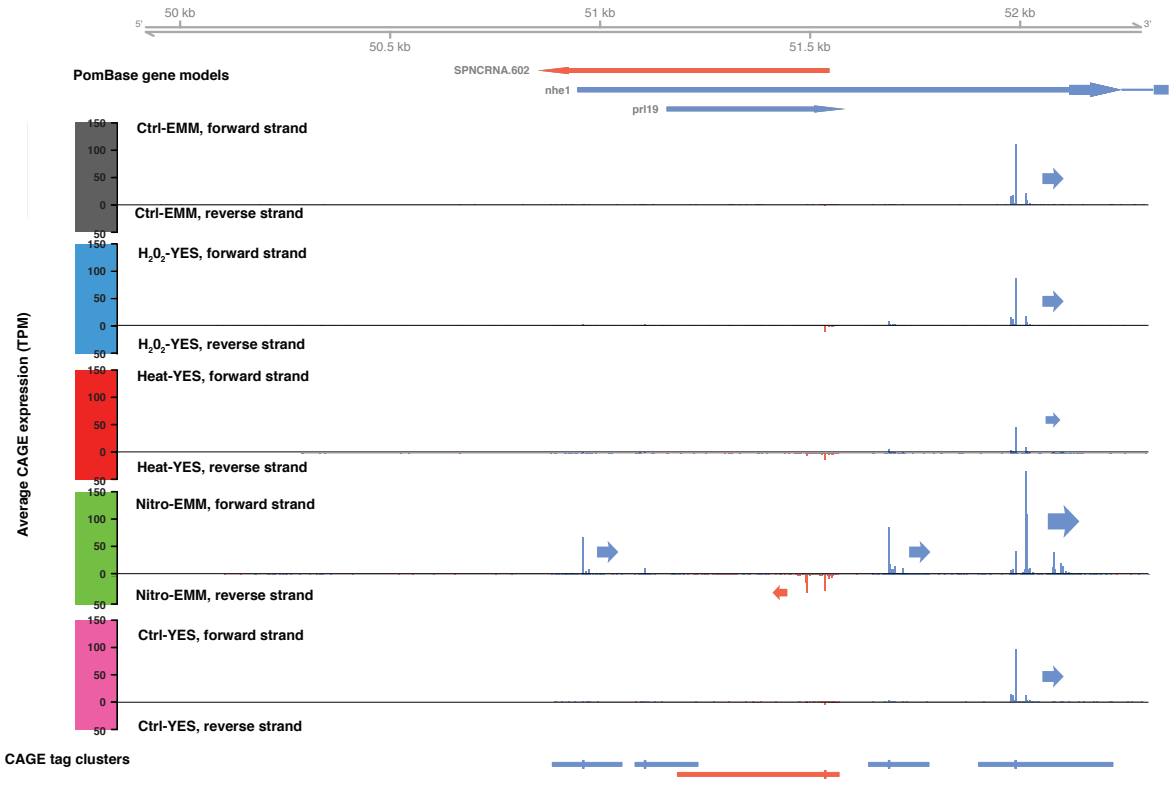


Figure 2

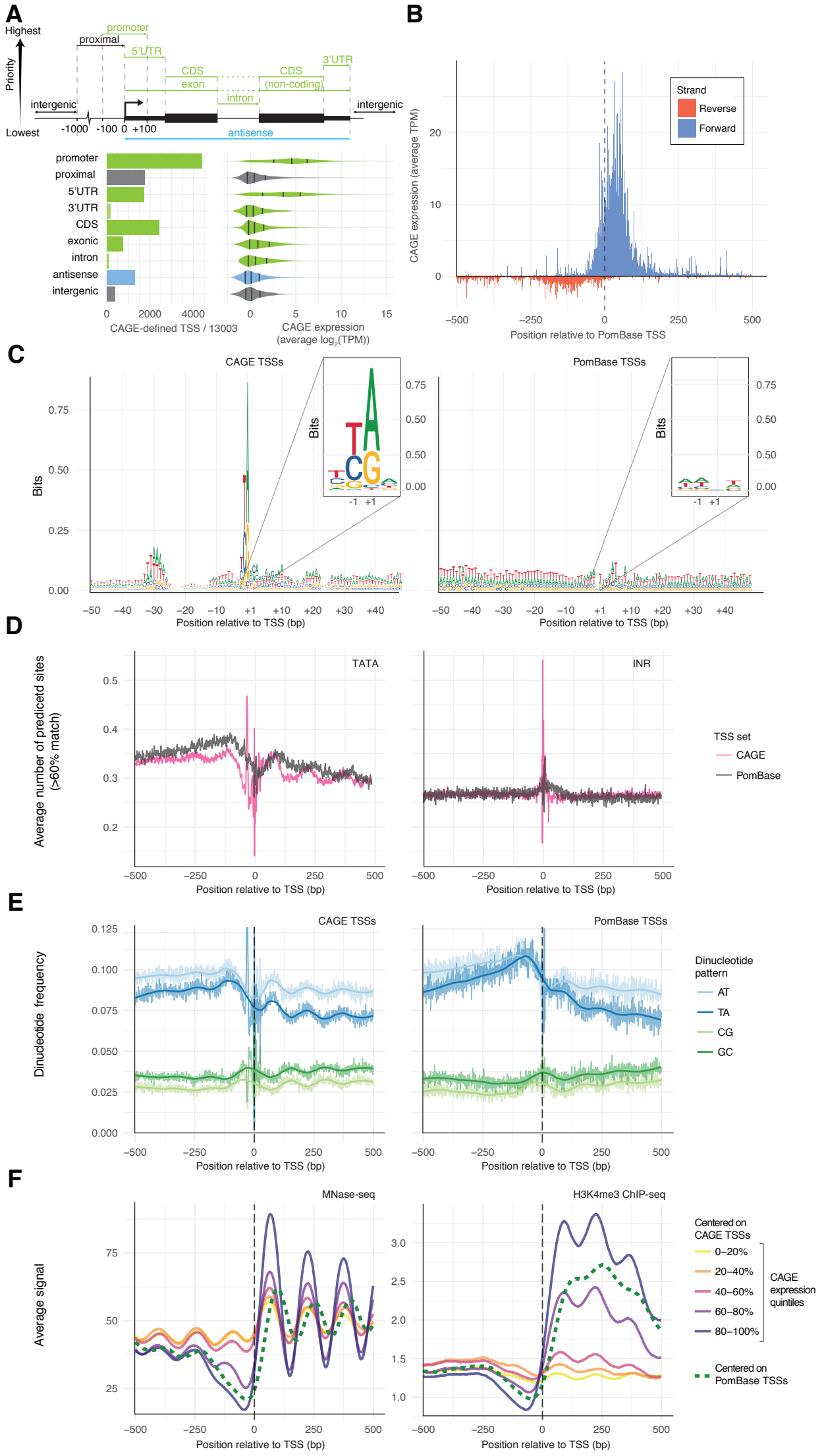

A
Highest
Priority
Lowest
promoter
proximal
5'UTR
CDS
exon
CDS
(non-coding)
3'UTR
intron
intergenic
intergenic
-1000
-100
0
+100
antisense
promoter
proximal
5'UTR
3'UTR
CDS
exonic
intron
antisense
intergenic
CAGE-defined TSS / 13003
CAGE expression
(average log2(TPM))
B
Strand
Reverse
Forward
CAGE expression (average TPM)
Position relative to PomBase TSS
C
CAGE TSSs
PomBase TSSs
Bits
Position relative to TSS (bp)
D
TATA
INR
Average number of predicetd sites
(>60% match)
Position relative to TSS (bp)
TSS set
CAGE
PomBase
E
CAGE TSSs
PomBase TSSs
Dinucleotide frequency
Position relative to TSS (bp)
Dinucleotide
pattern
AT
TA
CG
GC
F
MNase-seq
H3K4me3 ChIP-seq
Average signal
Position relative to TSS (bp)
Centered on
CAGE TSSs
0–20%
20–40%
40–60%
60–80%
80–100%
CAGE
expression
quintiles
Centered on
PomBase TSSs

Figure 3

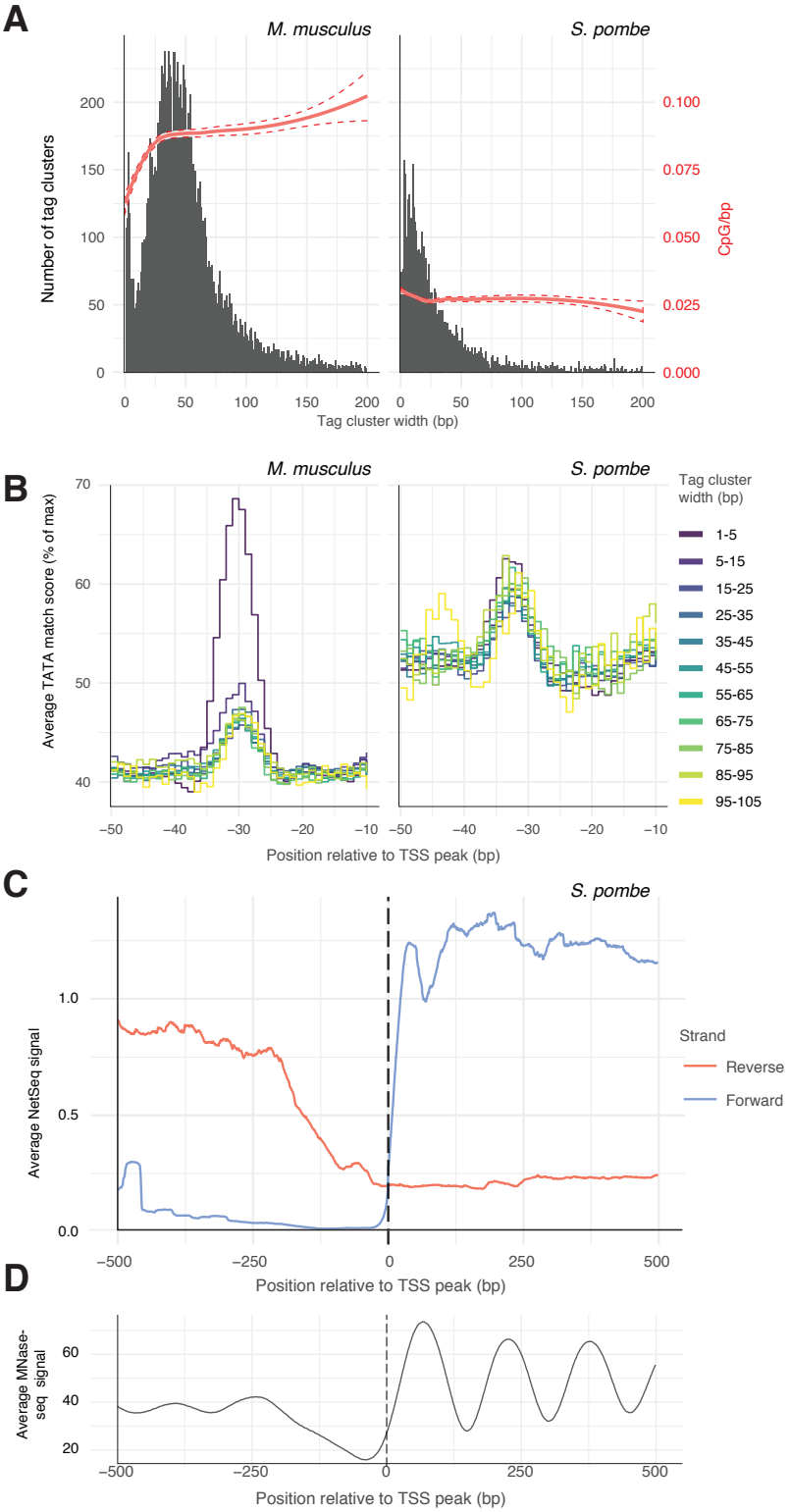

A
M. musculus
S. pombe
Number of tag clusters
CpG/bp
Tag cluster width (bp)
B
M. musculus
S. pombe
Average TATA match score (% of max)
Tag cluster width (bp)
1-5
5-15
15-25
25-35
35-45
45-55
55-65
65-75
75-85
85-95
95-105
Position relative to TSS peak (bp)
C
S. pombe
Average NetSeq signal
Strand
Reverse
Forward
Position relative to TSS peak (bp)
D
Average MNase-seq signal
Position relative to TSS peak (bp)

Figure 4

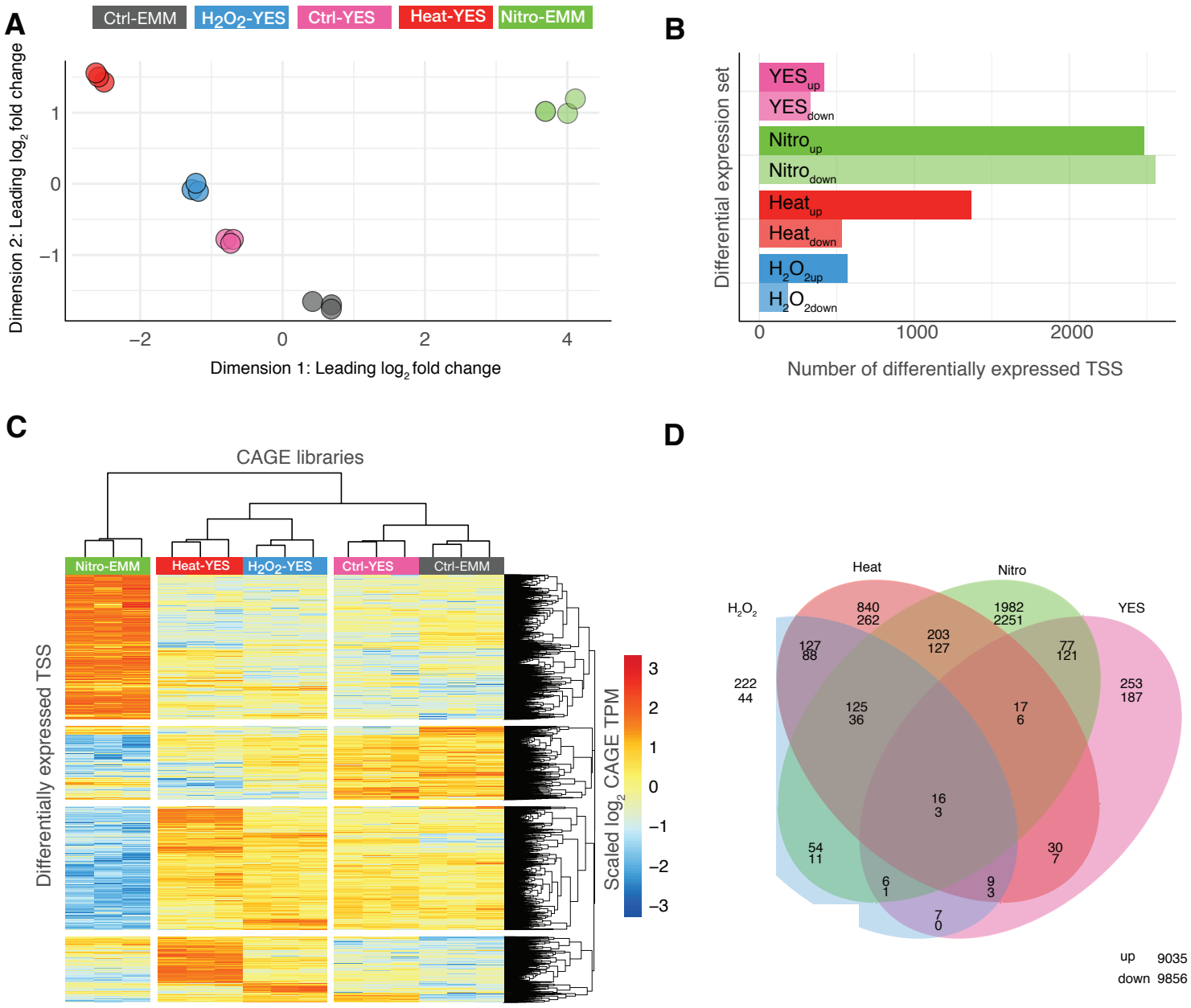

A
Ctrl-EMM
H2O2-YES
Ctrl-YES
Heat-YES
Nitro-EMM
Dimension 2: Leading log2 fold change
Dimension 1: Leading log2 fold change
B
Differential expression set
YESup
YESdown
Nitroup
Nitrodown
Heatup
Heatdown
H2O2up
H2O2down
Number of differentially expressed TSS
C
CAGE libraries
Nitro-EMM
Heat-YES
H2O2-YES
Ctrl-YES
Ctrl-EMM
Differentially expressed TSS
Scaled log2 CAGE TPM
D
Heat
Nitro
H2O2
YES
up 9035
down 9856

Figure 5

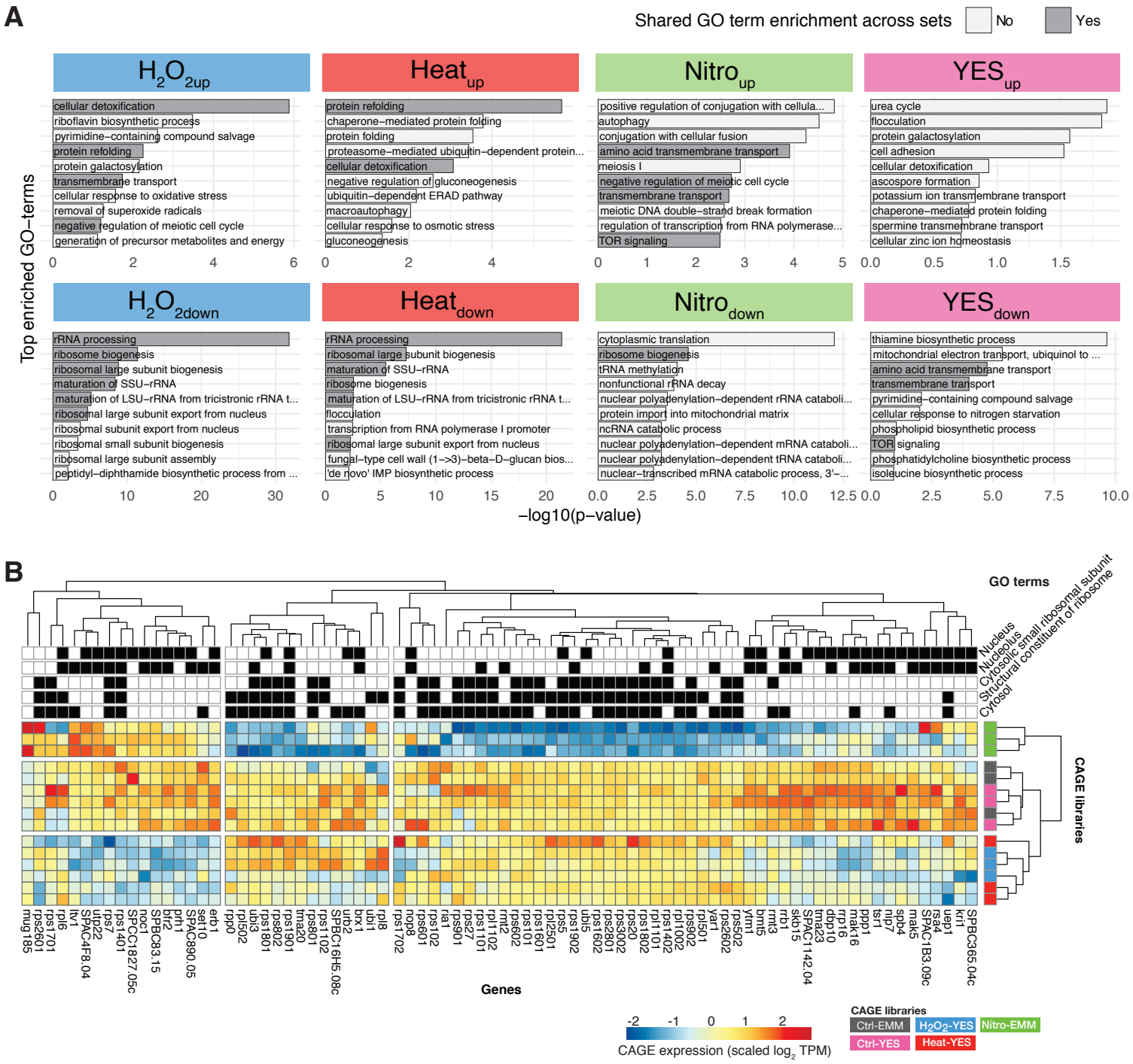

A
Shared GO term enrichment across sets
No
Yes
Top enriched GO–terms
H2O2up
cellular detoxification
riboflavin biosynthetic process
pyrimidine–containing compound salvage
protein refolding
protein galactosylation
transmembrane transport
cellular response to oxidative stress
removal of superoxide radicals
negative regulation of meiotic cell cycle
generation of precursor metabolites and energy
Heatup
protein refolding
chaperone–mediated protein folding
protein folding
proteasome–mediated ubiquitin–dependent protein...
cellular detoxification
negative regulation of gluconeogenesis
ubiquitin–dependent ERAD pathway
macroautophagy
cellular response to osmotic stress
gluconeogenesis
Nitroup
positive regulation of conjugation with cellula...
autophagy
conjugation with cellular fusion
amino acid transmembrane transport
meiosis I
negative regulation of meiotic cell cycle
transmembrane transport
meiotic DNA double–strand break formation
regulation of transcription from RNA polymerase...
TOR signaling
YESup
urea cycle
flocculation
protein galactosylation
cell adhesion
cellular detoxification
ascospore formation
potassium ion transmembrane transport
chaperone–mediated protein folding
spermine transmembrane transport
cellular zinc ion homeostasis
H2O2down
rRNA processing
ribosome biogenesis
ribosomal large subunit biogenesis
maturation of SSU–rRNA
maturation of LSU–rRNA from tricistronic rRNA t...
ribosomal large subunit export from nucleus
ribosomal subunit export from nucleus
ribosomal small subunit biogenesis
ribosomal large subunit assembly
peptidyl–diphthamide biosynthetic process from ...
Heatdown
rRNA processing
ribosomal large subunit biogenesis
maturation of SSU–rRNA
ribosome biogenesis
maturation of LSU–rRNA from tricistronic rRNA t...
flocculation
transcription from RNA polymerase I promoter
ribosomal large subunit export from nucleus
fungal–type cell wall (1–>3)–beta–D–glucan bios...
'de novo' IMP biosynthetic process
Nitrodown
cytoplasmic translation
ribosome biogenesis
tRNA methylation
nonfunctional rRNA decay
nuclear polyadenylation–dependent rRNA catabol...
protein import into mitochondrial matrix
ncRNA catabolic process
nuclear polyadenylation–dependent mRNA catabol...
nuclear polyadenylation–dependent tRNA catabol...
nuclear–transcribed mRNA catabolic process, 3'–...
YESdown
thiamine biosynthetic process
mitochondrial electron transport, ubiquinol to ...
amino acid transmembrane transport
transmembrane transport
pyrimidine–containing compound salvage
cellular response to nitrogen starvation
phospholipid biosynthetic process
TOR signaling
phosphatidylcholine biosynthetic process
isoleucine biosynthetic process
–log10(p–value)
B
GO terms
Nucleus
Nucleolus
Cytosolic small ribosomal subunit
Structural constituent of ribosome
Cytosol
CAGE libraries
Genes
CAGE expression (scaled log2 TPM)
CAGE libraries
Ctrl-EMM
H2O2-YES
Nitro-EMM
Ctrl-YES
Heat-YES

Figure 6

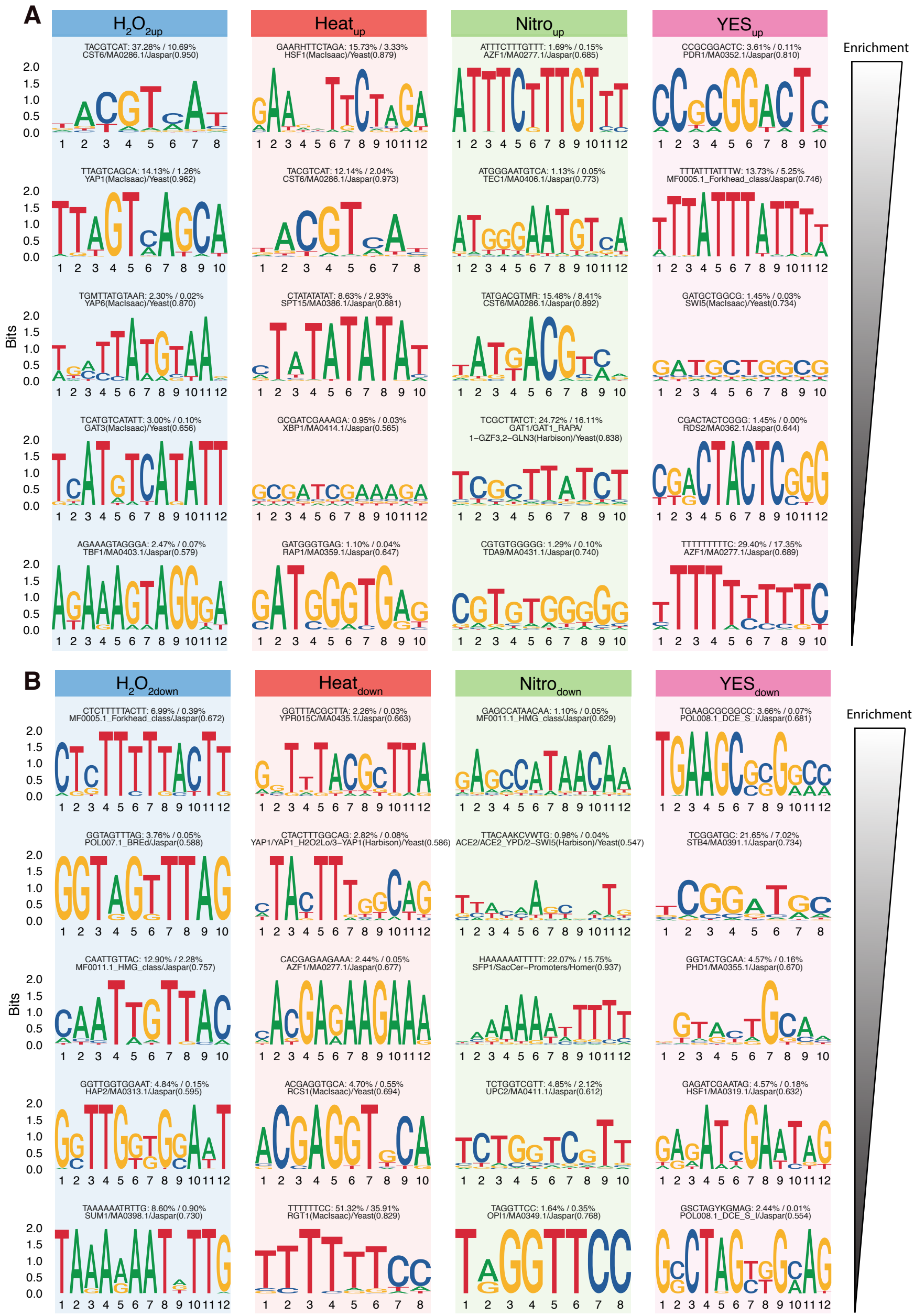

A
H2O2up
TACGTCAT: 37.28% / 10.69%
CST6/MA0286.1/Jaspar(0.950)
TTAGTCAGCA: 14.13% / 1.26%
YAP1(MacIsaac)/Yeast(0.962)
TGMTTATGTAAR: 2.30% / 0.02%
YAP6(MacIsaac)/Yeast(0.870)
TCATGTCATATT: 3.00% / 0.10%
GAT3(MacIsaac)/Yeast(0.656)
AGAAAGTAGGGA: 2.47% / 0.07%
TBF1/MA0403.1/Jaspar(0.579)
Heatup
GAARHTTCTAGA: 15.73% / 3.33%
HSF1(MacIsaac)/Yeast(0.879)
TACGTCAT: 12.14% / 2.04%
CST6/MA0286.1/Jaspar(0.973)
CTATATATAT: 8.63% / 2.93%
SPT15/MA0386.1/Jaspar(0.881)
GCGATCGAAAGA: 0.95% / 0.03%
XBP1/MA0414.1/Jaspar(0.565)
GATGGGTGAG: 1.10% / 0.04%
RAP1/MA0359.1/Jaspar(0.647)
Nitroup
ATTTCTTTGTTT: 1.69% / 0.15%
AZF1/MA0277.1/Jaspar(0.685)
ATGGGAATGTCA: 1.13% / 0.05%
TEC1/MA0406.1/Jaspar(0.773)
TATGACGTMR: 15.48% / 8.41%
CST6/MA0286.1/Jaspar(0.892)
TCGCTTATCT: 24.72% / 16.11%
GAT1/GAT1_RAPA/
1−GZF3,2−GLN3(Harbison)/Yeast(0.838)
CGTGTGGGGG: 1.29% / 0.10%
TDA9/MA0431.1/Jaspar(0.740)
YESup
CCGCGGACTC: 3.61% / 0.11%
PDR1/MA0352.1/Jaspar(0.810)
TTTATTTATTTW: 13.73% / 5.25%
MF0005.1_Forkhead_class/Jaspar(0.746)
GATGCTGGCG: 1.45% / 0.03%
SWI5(MacIsaac)/Yeast(0.734)
CGACTACTCGGG: 1.45% / 0.00%
RDS2/MA0362.1/Jaspar(0.644)
TTTTTTTTTC: 29.40% / 17.35%
AZF1/MA0277.1/Jaspar(0.689)
Enrichment
Bits
B
H2O2down
CTCTTTTTACTT: 6.99% / 0.39%
MF0005.1_Forkhead_class/Jaspar(0.672)
GGTAGTTTAG: 3.76% / 0.05%
POL007.1_BREd/Jaspar(0.588)
CAATTGTTAC: 12.90% / 2.28%
MF0011.1_HMG_class/Jaspar(0.757)
GGTTGGTGGAAT: 4.84% / 0.15%
HAP2/MA0313.1/Jaspar(0.595)
TAAAAAATRTTG: 8.60% / 0.90%
SUM1/MA0398.1/Jaspar(0.730)
Heatdown
GGTTTACGCTTA: 2.26% / 0.03%
YPR015C/MA0435.1/Jaspar(0.663)
CTACTTTGGCAG: 2.82% / 0.08%
YAP1/YAP1_H2O2Lo/3−YAP1(Harbison)/Yeast(0.586)
CACGAGAAGAAA: 2.44% / 0.05%
AZF1/MA0277.1/Jaspar(0.677)
ACGAGGTGCA: 4.70% / 0.55%
RCS1(MacIsaac)/Yeast(0.694)
TTTTTTCC: 51.32% / 35.91%
RGT1(MacIsaac)/Yeast(0.829)
Nitrodown
GAGCCATAACAA: 1.10% / 0.05%
MF0011.1_HMG_class/Jaspar(0.629)
TTACAAKCVWTG: 0.98% / 0.04%
ACE2/ACE2_YPD/2−SWI5(Harbison)/Yeast(0.547)
HAAAAAATTTTT: 22.07% / 15.75%
SFP1/SacCer−Promoters/Homer(0.937)
TCTGGTCGTT: 4.85% / 2.12%
UPC2/MA0411.1/Jaspar(0.612)
TAGGTTCC: 1.64% / 0.35%
OPI1/MA0349.1/Jaspar(0.768)
YESdown
TGAAGCGCGGCC: 3.66% / 0.07%
POL008.1_DCE_S_I/Jaspar(0.681)
TCGGATGC: 21.65% / 7.02%
STB4/MA0391.1/Jaspar(0.734)
GGTACTGCAA: 4.57% / 0.16%
PHD1/MA0355.1/Jaspar(0.670)
GAGATCGAATAG: 4.57% / 0.18%
HSF1/MA0319.1/Jaspar(0.632)
GSCTAGYKGMAG: 2.44% / 0.01%
POL008.1_DCE_S_I/Jaspar(0.554)
Enrichment
Bits

Figure 7

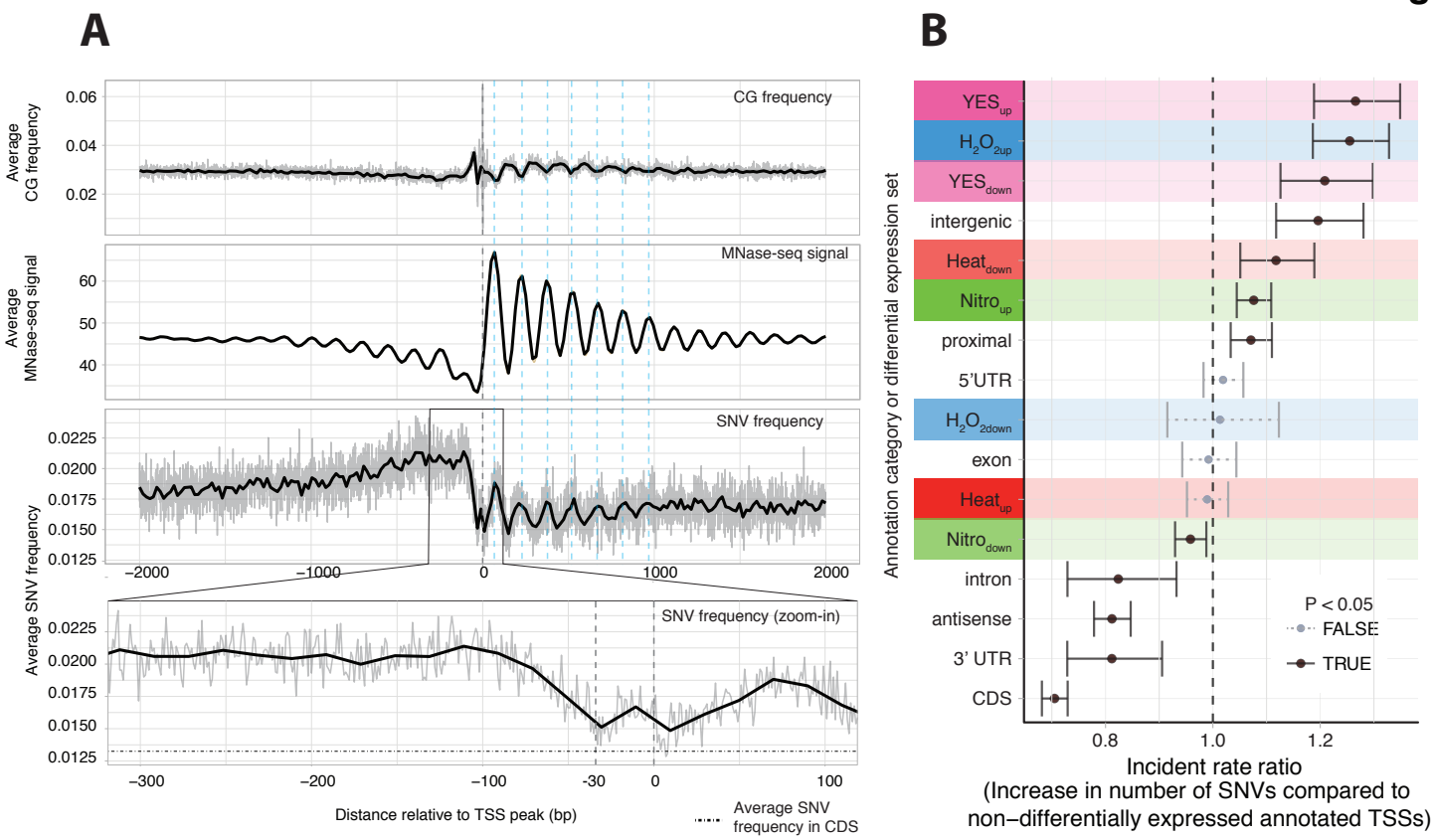

A
B
CG frequency
Average CG frequency
MNase-seq signal
Average MNase-seq signal
SNV frequency
Average SNV frequency
SNV frequency (zoom-in)
Distance relative to TSS peak (bp)
Average SNV frequency in CDS
Annotation category or differential expression set
YES up
H2O2up
YESdown
intergenic
Heatdown
Nitroup
proximal
5'UTR
H2O2down
exon
Heatup
Nitrodown
intron
antisense
3' UTR
CDS
P < 0.05
FALSE
TRUE
Incident rate ratio
(Increase in number of SNVs compared to non−differentially expressed annotated TSSs)

Figure 8

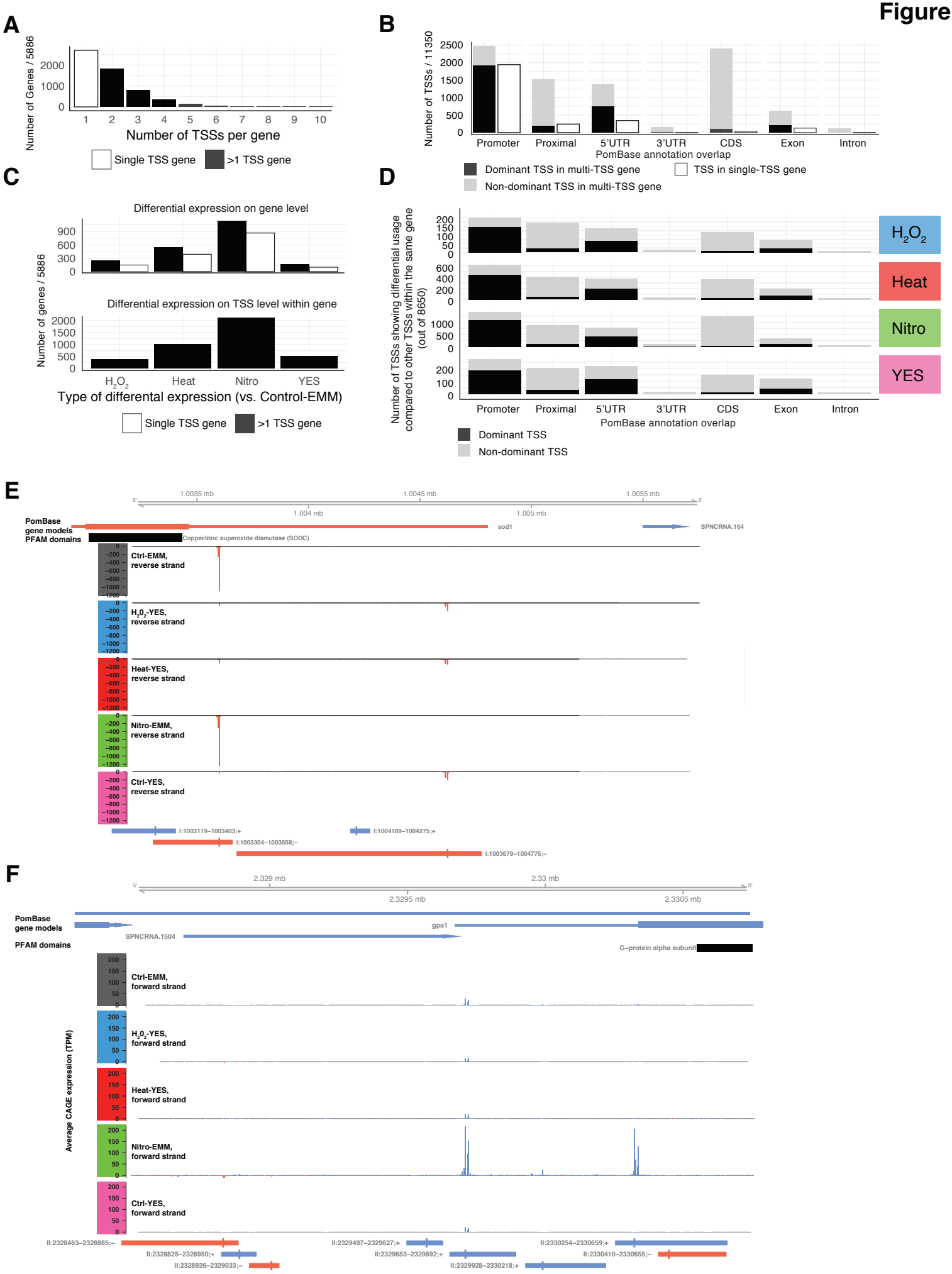
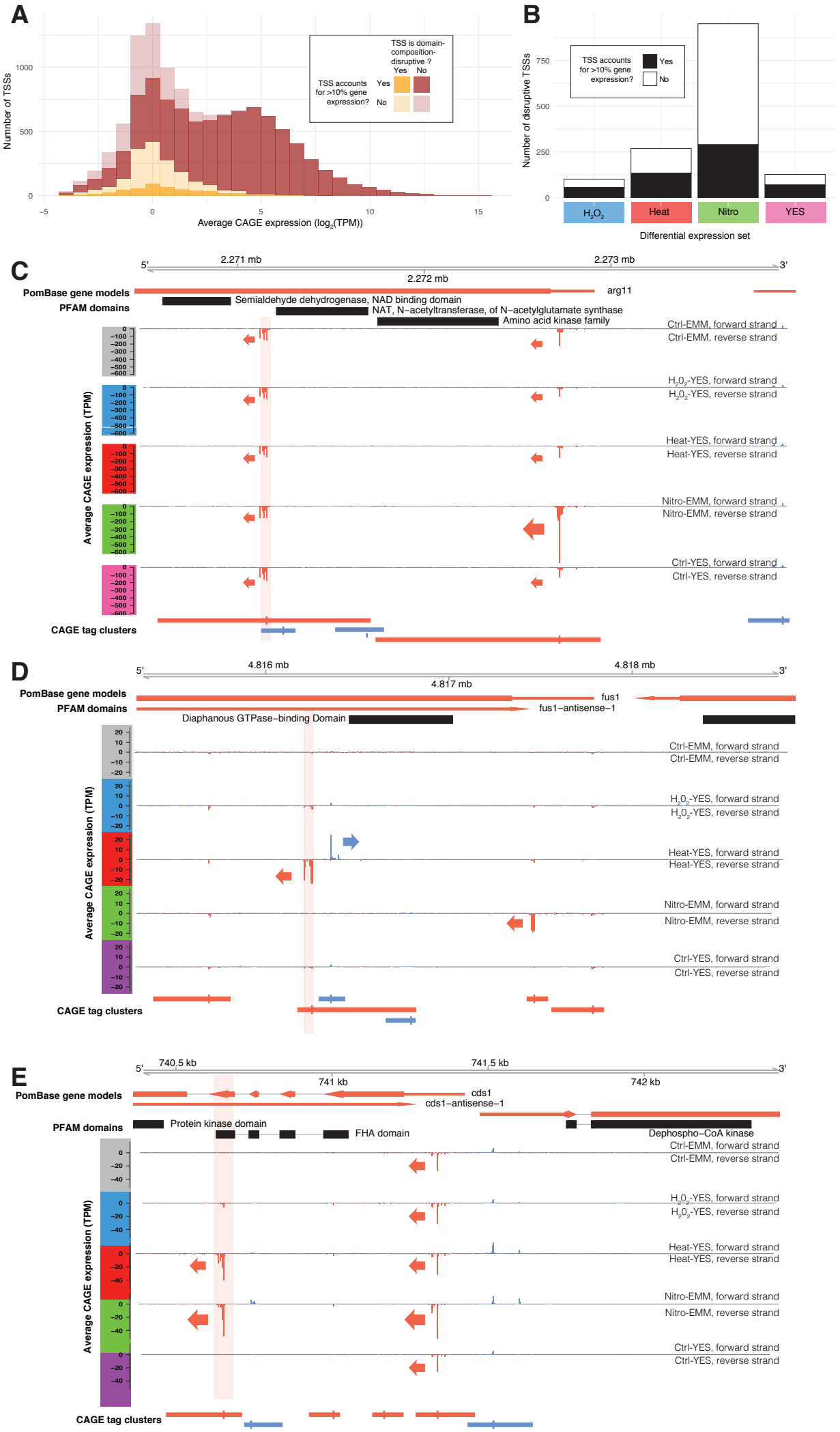

A
Number of Genes / 5886
2000
1000
0
1 2 3 4 5 6 7 8 9 10
Number of TSSs per gene
Single TSS gene
>1 TSS gene
B
Number of TSSs / 11350
2500
2000
1500
1000
500
0
Promoter
Proximal
5'UTR
3'UTR
CDS
Exon
Intron
PomBase annotation overlap
Dominant TSS in multi-TSS gene
TSS in single-TSS gene
Non-dominant TSS in multi-TSS gene
C
Differential expression on gene level
900
600
300
0
Differential expression on TSS level within gene
2000
1500
1000
500
0
Number of genes / 5886
H2O2
Heat
Nitro
YES
Type of differental expression (vs. Control-EMM)
Single TSS gene
>1 TSS gene
D
Number of TSSs showing differential usage compared to other TSSs within the same gene (out of 8650)
H2O2
Heat
Nitro
YES
Promoter
Proximal
5'UTR
3'UTR
CDS
Exon
Intron
PomBase annotation overlap
Dominant TSS
Non-dominant TSS
E
1.0035 mb
1.0045 mb
1.0055 mb
1.004 mb
1.005 mb
PomBase gene models
PFAM domains
sod1
SPNCRNA.164
Copper/zinc superoxide dismutase (SODC)
Ctrl-EMM, reverse strand
H2O2-YES, reverse strand
Heat-YES, reverse strand
Nitro-EMM, reverse strand
Ctrl-YES, reverse strand
I:1003119–1003403;+
I:1004189–1004275;+
I:1003304–1003658;–
I:1003679–1004775;–
F
2.329 mb
2.33 mb
2.3295 mb
2.3305 mb
PomBase gene models
PFAM domains
gpa1
SPNCRNA.1504
G–protein alpha subunit
Average CAGE expression (TPM)
Ctrl-EMM, forward strand
H2O2-YES, forward strand
Heat-YES, forward strand
Nitro-EMM, forward strand
Ctrl-YES, forward strand
II:2328463–2328885;–
II:2328825–2328950;+
II:2328926–2329033;–
II:2329497–2329627;+
II:2329653–2329892;+
II:2329928–2330218;+
II:2330254–2330659;+
II:2330410–2330655;–

Figure 9


A
Numnber of TSSs
1000
500
0
−5
0
5
10
15
Average CAGE expression (log2(TPM))
TSS is domain-composition-disruptive ?
Yes
No
TSS accounts for >10% gene expression?
Yes
No
B
Number of disruptive TSSs
750
500
250
0
TSS accounts for >10% gene expression?
Yes
No
H2O2
Heat
Nitro
YES
Differential expression set
C
5'
2.271 mb
2.272 mb
2.273 mb
3'
PomBase gene models
arg11
PFAM domains
Semialdehyde dehydrogenase, NAD binding domain
NAT, N−acetyltransferase, of N−acetylglutamate synthase
Amino acid kinase family
Ctrl-EMM, forward strand
Ctrl-EMM, reverse strand
H2O2-YES, forward strand
H2O2-YES, reverse strand
Heat-YES, forward strand
Heat-YES, reverse strand
Nitro-EMM, forward strand
Nitro-EMM, reverse strand
Ctrl-YES, forward strand
Ctrl-YES, reverse strand
Average CAGE expression (TPM)
CAGE tag clusters
D
5'
4.816 mb
4.817 mb
4.818 mb
3'
PomBase gene models
fus1
PFAM domains
fus1−antisense−1
Diaphanous GTPase−binding Domain
Ctrl-EMM, forward strand
Ctrl-EMM, reverse strand
H2O2-YES, forward strand
H2O2-YES, reverse strand
Heat-YES, forward strand
Heat-YES, reverse strand
Nitro-EMM, forward strand
Nitro-EMM, reverse strand
Ctrl-YES, forward strand
Ctrl-YES, reverse strand
Average CAGE expression (TPM)
CAGE tag clusters
E
5'
740,5 kb
741 kb
741,5 kb
742 kb
3'
PomBase gene models
cds1
cds1−antisense−1
PFAM domains
Protein kinase domain
FHA domain
Dephospho−CoA kinase
Ctrl-EMM, forward strand
Ctrl-EMM, reverse strand
H2O2-YES, forward strand
H2O2-YES, reverse strand
Heat-YES, forward strand
Heat-YES, reverse strand
Nitro-EMM, forward strand
Nitro-EMM, reverse strand
Ctrl-YES, forward strand
Ctrl-YES, reverse strand
Average CAGE expression (TPM)
CAGE tag clusters