System models in Yeast

Introduction

Yeasts have a long history as eukaryotic model organism due to their easy cultivation and their simple and accessible biology, which nevertheless allows to study conserved eukaryotic cellular and molecular processes. An advantage is the possibility to perform high throughput experiments with Yeast on a scale, hard to reach with other eukaryotic model organisms. This allows the systematic study of environmental conditions and their resulting responses like autophagy or the Heat Shock Response (HSR). Furthermore, the availability of loss of function mutants (knockouts) and double knockouts of its approximately 6000 genes allows to probe the molecular interactions in a systematic manner. The widespread use of Yeast strains like *saccharomyces cerevisiae* led to the creation of many databases and tools by the scientific community, which can be used for systematic studies. The properties and applications of these databases and tools should be outlined on the example of the Yeast Heat Shock Response and its central transcription factors HSF1 and MSN2/MSN4.

Cells need to maintain protein homeostasis under a variety of environmental conditions like increased heat, osmotic stress or oxidative stress. Under these conditions, the structural and functional integrity of protein homeostasis (proteostasis) needs to be maintained (Hartl et al., 2011). Therefore, protein folding needs to be assisted further and misfolded proteins, which can lead to toxic aggregates, need to be refolded or degraded. Without processes to respond to these conditions, the tolerance to environmental stress decreases significantly. The production of heat shock proteins (HSP) like chaperones and the synthesis of osmoprotectants like Trehalose are central parts of the cellular response to such stimuli. Chaperone proteins can assist in the folding of proteins by interaction with their non-native states, inhibit aggregation and favor disaggregation. This process can be divided into the general, constitutive expression of housekeeping chaperones and the stress response due to environmental stimuli. This stress response needs to be tightly controlled, as it is expensive to maintain for the cell and comes with a decrease in translational activity and proliferation (Durchschlag et al., 2004). The control of the response is facilitated by the transcription factor (TF) MSN2 and its paralogue MSN4 (Schmitt & McEntee, 1996) and a heat shock response activated by the transcription factor HSF1. The degree of functional separation regarding the type of stress and contribution to the constitutive chaperone response is still under debate (Solís et al., 2016). Even though there is a consensus binding site and high redundancy between Msn2p and Msn4p.

The S288C reference genome and the Saccharomyces Genome Databank

The genetic foundation for the systematic study of yeast is formed by the Saccharomyces Genome Databank (SGD) and the S288C reference Genome (Cherry et al., 2012). The database maintains the reference genome for the study of *saccharomyces cerevisiae* and allows to map standard genes with their names and aliases to their systematic name and a unique SGD identifier. This information is extended by common variants and mapping to homologues in the reference genomes of other strains. Studies into the phenotype of knockouts, overexpression or conditional expression are manually curated and given for the corresponding genes. Furthermore, manually curated information from low- and high throughput experiments and automatic classifications coming from the Gene Ontology Annotation project are given as Gene Ontology (GO) terms (Hong et al., 2008). This allows to classify genes and their products in *saccharomyces cerevisiae* based on the biological process (BP), molecular function (MF) and cellular compartment (CC). Based on this knowledge, gene sets produced by high throughput methods can be analyzed and enriched for certain GO terms.

HSF1 (`YGL073W`, `SGD:S000003041`) is located on chromosome VII and codes for an 833 amino acid protein of 93.25 kDa. The molecular function of the transcription factor is described as `MF GO:0003700` which stands for DNA binding transcription factor activity, as it binds to heat shock elements (HSE). Being a transcription factor in the regulation of transcription, the protein is localized in the nucleus (`CC GO:0005634`). The biological processes associated with the gene correspond to the function in the heat shock response (`BP GO:0009408`) and come from stimulating transcription as transcription factor (`BP GO:0045944`). Additionally, negative regulation of TORC signaling (`BP GO:1904262`) by Hsf1p is associated with increase in proteasomes (Rousseau & Bertolotti, 2016). The role of HSF1 becomes clear from curated phenotype data in the SGD. A reduction in function of HSF1 leads to increased heat sensitivity and decreased chemical tolerance of the cell. Overexpression and activation however confer decreased heat

sensitivity, increased thermotolerance but also a decreased budding index. This hints at the cost to maintain the protective heat shock response.

MSN2 (YMR037C, SGD:S000004640) and MSN4 (YKL062W, SGD:S000001545) are paralogues with an amino acid identity of 41 percent. Even though they share the same functions and a shared DNA binding motif (STRE stress response element), they only complement each other partially and may have different roles in the stress response (Schmitt & McEntee, 1996). Both factors are present in the cytosol (CC GO:0005829) and in the nucleus (CC GO:0005634) as their regulation is partially facilitated by import end export from the nucleus and phosphorylation by protein kinase A (PKA) (Görner et al., 1998). The molecular function is defined as DNA binding in cis-regulatory regions (MF GO:0000987) and with RNA-Polymerase II specific transcription factor activity (MF GO:0000981). MSN2 and MSN4 are both associated with a variety of biological processes. As central regulators in a the general stress response they activate transcription in response of acidic and alkaline pH (BP GO:0061402, GO:0061422), heat stress (BP GO:0061408) or oxidative stress coming from hydrogen peroxide (BP GO:0061407). Both knockouts are viable and come with a decreased chronological lifespan, resistance to chemicals or oxidative stress.

YEASTRACT – a curated database of transcription factor associations

The expression of genes is mediated by a delicate network of transcription factors and specific DNA binding sides which can result in enhancement and silencing of gene expression. YEASTRACT (Yeast Search for Transcriptional Regulators And Consensus Tracking) aims to aggregate this data from scientific publications and high throughput methods to make it accessible for systematic investigations (Monteiro et al., 2020). The database consists of 183 transcription factors and more than 170.000 associations. Each association stems from a scientific publication and gives evidence for the interaction between on of the transcription factors and a target gene. The interaction is furthermore characterized by the type of the regulation, negative or positive, and the environmental conditions like stressed, unstressed or starvation conditions like nitrogen depletion.

A simple way how this data can be used is to reason about to regulatory task of a transcription factor, as one TF is expected to have downstream targets with common function. Because of this, pathway memberships or functions will be enriched in the set of target genes compared to a random selection of genes. This enrichment can be shown with a functional enrichment analysis and enriched functions in the HSF1 and MSN2/4 associated genes will tell us about their function and mode of action. The analysis will probe for the enrichment of a variety of predefined gene sets, for certain GO terms for example, and will give us a p-value for the significance of enrichment for each gene set, corrected for multiple testing. For HSF1 there are 1785 positive regulatory associations for 1323 different targets in the Yeastract database. On the same hand there are 503 negative associations for 490 different targets. If we focus on data coming from experiments under stressed conditions, we get 544 different positively regulated Genes and 4 negatively regulated Genes, which hints at an enhancing mode of action in stressed cells. Similar data are available for the targets of MSN2 and MSN4 and can be seen in Table 1. The scripts used in the analysis of the Yeastract database and generation of plot can be found in the project repository under the name yeastract_analysis.ipynb.

The target genes affected by these transcription factors are not exclusive and are shared to some degree between the different stress responses. This can be seen in Figure 1 which shows the overlap of target associations between HS1, MSN2 and MSN4. All three TFs share a core of 338 positively associated genes under stress conditions, which is unsurprising, as different types of stress response employ common tools, like chaperones, to maintain proteostasis. The target genes affected by the paralogues MSN2 and MSN4 are shared between the two TF, as all 702 MSN4 target genes are included in the 854 targets. In contrast to the common positive response, the negatively regulated association are mostly shared by MSN2 and MSN4 but not by HSF1.

Nevertheless, these findings need to be evaluated with care as the data only shows qualitative associations from high-throughput experiments and the mode of regulation can differ. To imagine a switch like activation of these genes, mediated by a single factor (TF) would be naïve, and multiple layers of further transcriptional, translational and posttranslational regulation are involved. Furthermore, not all these targets will be directly regulated by the transcription factor and subsequent waves of transcription factor activation can give rise to a higher fidelity. The data is also grouped due to a general stress response. The type of stressed, as discussed previously, has an impact on the stress response and the factors involved.

Table 1, associated genes for the TFs HSF1, MSN2 and MSN4 according to the data curated in the YEASTRACT database.

	all conditions		stress conditions	
	positive targets	negative targets	positive targets	negative targets
HSF1	1323	490	544	4
MSN2	1502	1049	854	382
MSN4	1133	615	702	231

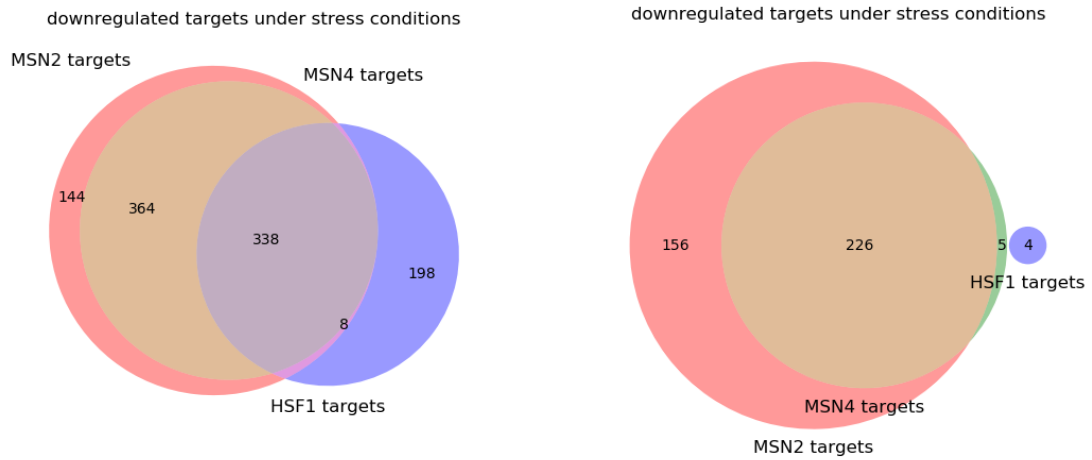

Figure 1, Targets of the transcription factors HSF1, MSN2 and MSN4 under stressed conditions and their overlap.

The bioinformatics tool DAVID can be used to enrich these positively regulated genes based on their GO terms to identify common features. An enrichment for the 544 upregulated targets of HSF1 under stressed conditions is in agreement with the countermeasures taken by the cell in the heat shock response. The GO BP terms are tested for their significant enrichment and corrected for multiple testing via Benjamini-Hochberg.

The most enriched BP terms for HSF1 consist of the chaperone response and the processes for protein folding (BP GO:0006457, $p = 1.1 \times 10^{-6}$) and protein refolding (BP GO:0042026, $p = 2.8 \times 10^{-6}$). On the other hand processes for the degradation of unfolded protein by the Ubiquitin Proteasome System are enriched like the ER-associated ubiquitin-dependent protein catabolic process which corresponds to the unfolded protein response in the endoplasmic reticulum (ER) and the ERAD pathway (BP GO:0030433, $p = 8.8 \times 10^{-4}$) or the proteasome assembly (BP GO:0043248, $p = 6.4 \times 10^{-2}$). Most interesting, genes associated with sporulation (BP GO:0030435, $p = 4.2 \times 10^{-3}$) are activated likewise, which constitutes the ultimate survival mechanism under sustained stress. An overview of the significantly enriched GO BP terms can be seen in Table 2.

Table 2, Enriched GO BP terms in the 544 positively regulated targets in response to stress by HSF1 in Yeast. The highest six GO terms are ranked based on their Benjamini-Hochberg score.

Term		Fold Enrichment	Benjamini Hochberg
GO:0006457	protein folding	3.5	1.1×10^{-6}
GO:0042026	protein refolding	8.5	2.8×10^{-6}
GO:0034605	cellular response to heat	5.1	7.1×10^{-4}
GO:0030433	ER-associated ubiquitin-dependent protein catabolic process	3.8	8.8×10^{-4}

GO:0055114	oxidation-reduction process	1.8	2.9×10^{-3}
GO:0030435	sporulation resulting in formation of a cellular spore	2.5	4.2×10^{-3}

Even though the highest enriched GO terms differ for MSN2 and MSN4 as seen in Table 3 and Table 4, all these terms occur as highly enriched terms for both TFs. The fact, that the stress response by MSN2/4 is needed to counteract oxidative stress, can be seen from the enriched GO BP terms. The GO terms oxidation-reduction process (GO:0055114, $p < 2.5 \times 10^{-12}$), cellular oxidant detoxification (GO:0098869, $p < 5.2 \times 10^{-5}$), and cellular response to oxidative stress (GO:0034599, $p < 1.7 \times 10^{-4}$) are the most enriched GO BP terms for both MSN2 and MSN4. Nevertheless, genes in the folding of proteins (BP GO:0006457, $p = 1.7 \times 10^{-2}$ for MSN4) and the cellular response to heat (GO:0055114, $p < 5.6 \times 10^{-3}$) are enriched too, which may be part of the core set of stress genes discussed earlier. Genes associated with the ubiquitin proteasome system can be found as well, though not as highly enriched as in the case of HSF1. On the same hand, genes associated with sporulation (GO:0030435, $p < 4.8 \times 10^{-3}$) are enriched in the MSN2/4 data, as this response is independent of the type of stress as well.

MSN2 is associated with 854 positively regulated targets under stress and has 9 highly enriched ($Benjamini\ Hochberg < 0.05$) GO terms. MSN4 has 152 less associated genes but 18, more diverse, highly enriched GO terms, which may seem counterintuitive. The fact, that more gene sets in the MSN4 analysis are significantly enriched cannot be interpreted as a broader mode of action but is a weakness of the underlying high throughput data and analysis. The associations in the YEASTRACT database are only qualitative and give no quantitative estimate regarding the strength of the association other than the numbers of evidence for an association. This makes the enrichment sensitive for minor gene-target associations, which can decrease the enrichment of other gene sets. An additional bias can come from the different experiments giving evidence for different TF associations, which can't be considered equal and equally distributed over the TFs. Even if mathematically correct, the enrichment score and significance of the gene sets should not be compared between TFs and evaluated in an absolute manner.

Table 3, Enriched GO BP terms in the 854 positively regulated targets in response to stress by MSN2 in Yeast. The highest six GO terms are ranked based on their Benjamini-Hochberg score.

Term		**Fold Enrichment**	**Benjamini Hochberg**
GO:0055114	oxidation-reduction process	2.3	3.3×10^{-15}
GO:0098869	cellular oxidant detoxification	4.5	5.2×10^{-5}
GO:0034599	cellular response to oxidative stress	2.8	1.7×10^{-4}
GO:0008152	metabolic process	1.7	3.3×10^{-3}
GO:0006099	tricarboxylic acid cycle	3.7	3.0×10^{-3}
GO:0055114	cellular response to heat	3.7	5.6×10^{-3}

Table 4, Enriched GO BP terms in the 702 positively regulated targets in response to stress by MSN4 in Yeast. The highest six GO terms are ranked based on their Benjamini-Hochberg score.

Term		**Fold Enrichment**	**Benjamini Hochberg**
GO:0055114	oxidation-reduction process	2.4	2.5×10^{-12}
GO:0034599	cellular response to oxidative stress	3.3	3.7×10^{-6}
GO:0098869	cellular oxidant detoxification	5.1	1.9×10^{-5}
GO:0005975	carbohydrate metabolic process	2.6	5.8×10^{-4}

GO:0055114	cellular response to heat	3.7	7.2×10^{-4}
GO:0030435	sporulation resulting in formation of a cellular spore	3.7	9.3×10^{-4}

Protein-protein interactions and ConsensusPathDB

Cellular processes are based on large networks of proteins with enzymatical, structural or signaling tasks. All these proteins interact via noncovalent, specific protein-protein interactions (PPI) to pursue their function. These data are a valuable component of systematic models and therefore aggregated in a variety of databases. Most frequently, the data consists of qualitative entries characterizing the interactions and can be used to infer interaction networks. ConsensusPathDB aggregates these PPI databases in yeast as a meta database and has collected data about 155,225 Interactions from eleven different sources (Herwig et al., 2016). Next to this PPI data, ConsensusPathDB has also data about signaling and metabolic reactions, genetic interactions and gene regulation, drug target interactions and biochemical pathways.

To show the properties of these data, the PPI interactions for the TFs Hsf1p, Msn2p and Msn4p were extracted. In the next step, the interaction partners were matched with their genes in the yeast genome annotation by using Uniprot (Bateman, 2019). The associated genes were then probed for enrichment of GO and KEGG gene sets with the g:profiler functional enrichment software (Raudvere et al., 2019). This allows to get a general overview of the proteins interacting with the TFs and the function of the interaction. The scripts used in this analysis and for the generation of plots can be found in the project repository under consensus_path_analysis.ipynb.

Table 5, protein-protein interactions for the TFs HSF1, MSN2 and MSN4 according to ConsensusPathDB.

	Number of distinct protein-protein interactions		
	HSF1	MSN2	MSN4
Confidence > 0.9	2	7	4
Confidence > 0.5	16	13	11
Confidence > 0.0	35	37	34
All	396	381	748

The interaction partners of the TFs are enriched for functions and locations expected for the regulatory function in the nucleus. The interaction partners are enriched in biological function involved with chromatin remodeling for transcriptional activation like histone acetylation (Figure 5) and membership in the histone acetylation complexes like the SAGA (Spt-Ada-Gcn5 acetyltransferase) complex (Figure 3). Furthermore, the partners are also involved in the transcriptional machinery, which is recruited by the interaction with basal transcription factors like the TFIID subunit TAF5 (Hsf1p, confidence = 0.89).

The interaction partners also reflect the specificities in activation for the TFs. Hsf1p activation is known to be driven by a combined mechanism of Serine and Threonine phosphorylation and chaperone titration (Anckar & Sistonen, 2011). This mechanism is based on the sequestration of Hsf1p by chaperones in the absence of stress and unfolded proteins. In case of increased stress, unfolded proteins bind to the Hsf1p bound HSP90 and HSP70 type chaperones and lead to a release of bound Hsf1p. Subsequent phosphorylation and trimerization is then needed to gain full transcriptional activation, which leads to the specific activation of the HSR and a feed-back regulation. The PPIs of Hsf1p agree with this mechanism and reflect some of this behavior. Proteins associated with Threonine and Serine kinase activity are enriched in the set of the 35 interacting proteins (Figure 3) and central chaperones of the HSP70 family can be found (Figure 2). Interactions, between Msn2p and Msn4p and these chaperones are found likewise (Figure 2), although the role of these interactions is not as clear. Some of the PPI partners of Msn4p can although be attributed to the activation process. As stated earlier, Msn2p and Msn4p activation is antagonized by PKA activity and subunits of the PKA are found to be interacting with Msn4p (KAPR confidence = 0.14, KAPA confidence = 0.99, KAPB confidence = 0.99). Next to inhibitory phosphorylation by PKA, phosphorylation by the negatively PKA regulated kinase Yak1 is known to be a part of activation of Msn2p and Hsf1p (Lee et al., 2008). PPIs are documented for both TFs (Msn2p confidence = 0.12, Hsf1p confidence = 0.58)

Most interesting, binding of Msn2p with proteins associated with the MAPK (mitogen activated protein kinase) pathway is documented in the database (Figure 4). Interaction is shown for the MAP kinase Hog1 (confidence = 0.99) and the Hog1 target Hot1 (confidence = 0.98). This is in concordance with the suspected of interplay between the osmotic stress MAPK pathway and the MSN2/4 stress response (Capaldi et al., 2008).

Even though these interactions can give interesting information about the mode of action and are in agreement with the literature, the same caution should be applied, as for all qualitative high throughput data. A single PPI is a complicated structural phenomenon and subject to allosteric effects, covalent modification and complex formation. Reducing these interactions to a binary qualitative model deprives most of this fidelity and data should be interpreted accordingly.

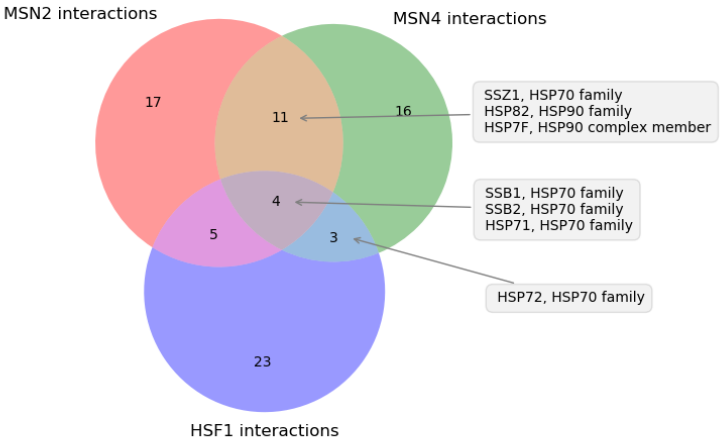

Figure 2, Protein-Protein Interactions for Msn2p, Msn4p and Hsf1p with confidence > 0 from the ConsensusPathwayDB. Shared interactions are shown as venn diagram and chaperone interactions are highlighted.

GO:BP		HSF1
Term Name	Term ID	p_adj
peptidyl-amino acid modification	GO:0018193	1.075×10^{-10}
peptidyl-serine phosphorylation	GO:0018105	9.842×10^{-9}
regulation of transcription, DNA-templated	GO:0006355	2.225×10^{-8}
regulation of nucleic acid-templated transcription	GO:1903506	2.289×10^{-8}
nucleic acid-templated transcription	GO:0097659	1.208×10^{-7}
transcription, DNA-templated	GO:0006351	1.208×10^{-7}
peptidyl-threonine phosphorylation	GO:0018107	1.842×10^{-6}

GO:CC		HSF1
Term Name	Term ID	p_adj
SAGA complex	GO:0000124	3.757×10^{-11}
SAGA-type complex	GO:0070461	1.340×10^{-10}
histone acetyltransferase complex	GO:0000123	1.212×10^{-7}
protein acetyltransferase complex	GO:0031248	4.560×10^{-7}

Figure 3, Functional enrichment of GO BP and CC terms for the 35 HSF1 interaction partners with confidence > 0 from ConsensusPathDB. The protein names were matched to their genes with Uniprot. The enrichment was performed with g:Profiler version e100_eg47_p14_3fdcd05, database updated on (07.07.2020).

GO:BP		MSN2
Term Name	Term ID	p_adj
biological regulation	GO:0065007	1.366×10^{-9}
response to stress	GO:0006950	3.818×10^{-8}
regulation of transcription by RNA polymerase II	GO:0006357	1.483×10^{-7}
regulation of biological process	GO:0050789	5.063×10^{-7}
transcription by RNA polymerase II	GO:0006366	1.426×10^{-6}
regulation of transcription, DNA-templated	GO:0006355	8.241×10^{-6}
protein folding	GO:0006457	6.677×10^{-4}
protein refolding	GO:0042026	3.373×10^{-3}

GO:CC		MSN2
Term Name	Term ID	p_adj
nucleus	GO:0005634	1.312×10^{-3}
H4/H2A histone acetyltransferase complex	GO:0043189	4.811×10^{-2}

KEGG		MSN2
Term Name	Term ID	p_adj
MAPK signaling pathway - yeast	KEGG:04011	2.244×10^{-2}

Figure 4, Functional enrichment of GO BP and CC terms and KEGG pathway membership for the 37 MSN2 interaction partners with confidence > 0 from ConsensusPathDB. The protein names were matched to their genes with Uniprot. The enrichment was performed with g:Profiler version e100_eg47_p14_3fdcd05, database updated on (07.07.2020).

GO:BP		MSN4
Term Name	Term ID	p_adj
histone acetylation	GO:0016573	4.112×10^{-5}
peptidyl-lysine acetylation	GO:0018394	4.688×10^{-5}
response to stress	GO:0006950	2.235×10^{-4}
protein folding	GO:0006457	3.784×10^{-4}

GO:CC		MSN4
Term Name	Term ID	p_adj
H4 histone acetyltransferase complex	GO:1902562	7.957×10^{-5}
histone acetyltransferase complex	GO:0000123	1.200×10^{-4}
SAGA complex	GO:0000124	1.308×10^{-3}
SAGA-type complex	GO:0070461	2.291×10^{-3}

Figure 5, Functional enrichment of GO BP and CC terms for the 34 MSN4 interaction partners with confidence > 0 from ConsensusPathDB. The protein names were matched to their genes with Uniprot. The enrichment was performed with g:Profiler version e100_eg47_p14_3fdcd05, database updated on (07.07.2020).

Investigating the Transcriptome of the Yeast HSR

Introduction

A central dogma in biology is the flow of information from DNA and Genes, being transcribed to RNA and then translated into proteins. For this reason, methods to analyze these layers are a central part of the biologic toolbox and separated into Genetics, Proteomics or Transcriptomics. Each of these layers has distinct capabilities and advantages but also limitations and are best combined. Transcriptomic methods analyze the whole universe of the transcribed genome which consists of rRNA, mRNAs and further regulatory and structural RNAs, therefore it can be used to monitor the state of the cell by analyzing its current mode of transcription. This also sets the limitations for transcriptomic analysis, as only processes resulting in a change in transcription can be studied. Compared to post transcriptional or post translational regulation, these changes happen on a longer timescale.

These transcriptomic methods will be applied in the systematic study of the Heat Shock Response in Yeast, discussed earlier. The focus on the analysis will lie on the transcriptional regulation of this response and the interplay of the transcription factors HSF1 and MSN2/4. *saccharomyces cerevisiae* was grown under regular conditions of 25°C and RNA sequencing was used to sequence the transcriptome as a reference. If nothing else is specified, fold-changes are always relative to this control of the respective strain. Then a mild (37°C) and severe (42°C) heat shock was induced, and the transcriptome was analyzed likewise after 10 and 40 minutes of exposure. All these experiments were performed as 3 biologically replicates. To investigate the role of the HSF1 and MSN2/4, two single knockouts and a double knockout were generated, and the analyses was repeated. The sequencing datasets were preprocessed and converted into fold change values relative to the control experiment.

Methods used in the analysis of the transcriptomic data

A custom lists of gene sets was prepared from the Yeastract database mentioned previously. Gene sets for positive and negative associated genes were generated and further separated into stress and non-stress/unknown associations. This resulted in 525 gene sets associated with 220 transcription factors. Furthermore, genes identified with shared influence by MSN2/4 and HSF1 were used as a "core stress" gene set and genes with MSN2/4 but no HSF1 association were included as "MSN2/4 exclusive" gene set. Further gene sets for the MSN2 exclusive and HSF1 exclusive associations were generated too. The script used to generate the basic gene sets can be found in the project repository under the name yeastract_to_GMT.ipynb. The script yeastract_analysis.ipynb can be used to append the MSN2/MSN4/HSF1 specific gene sets from the initial yeastract analysis. The scripts for the analysis of the transcriptomic data and generation of plots gen be found in the project repository under the name transcriptome_analysis.ipynb.

As described previously, enrichment analysis of gene sets was performed . Custom gene sets and sets coming from Gene Ontology (Ashburner et al., 2000; Hong et al., 2008), Wikipathways (Slenter et al., 2018) and KEGG (Kanehisa, 2000) annotations were used with the g:profiler functional enrichment software and ranked gene lists (Raudvere et al., 2019).

Jupyter Notebook (Kluyver et al., 2016) was used to implement the analysis of the transcriptomic data in the python programming language (Rossum, 1995). Pandas (McKinney, 2004) and Numpy (Oliphant, 2007) were used for datastructures and calculations and Matplotlib (Hunter, 2007) was used for plotting.

The heat shock transcriptome in wild type yeast

The temperature conditions and sample timepoints were chosen to investigate two modes of the HSR. The low temperature (37°C) environments results in a mild heat shock, which can be countered by the cell by reprogramming the transcriptome (Mühlhofer et al., 2019). Under mild conditions, the HSR becomes apparent as an initial burst, which declines over time. The effect on the transcriptome becomes apparent, when the highly up- and downregulated genes (at least 4-fold increase or decrease) are analyzed. After 10 minutes of exposure to the 37°C heat shock in the wild type (WT), 148 genes were highly upregulated, and 319 genes were highly downregulated (Figure 6**Fehler! Verweisquelle konnte nicht gefunden werden.**). This set of Genes was reduced to 78 upregulated genes and 83 downregulated genes after 40 minutes of exposure. The set of highly downregulated genes after 40 minutes of was nearly a complete subset of the initial downregulated genes . This was also the case for the upregulated genes, although not as clear as 16

new genes were highly upregulated after 40 minutes. This expression of late genes could hint at some shift in transcriptome, after accommodating to the new steady state, which are not associated with the immediate HSR (Figure 6).

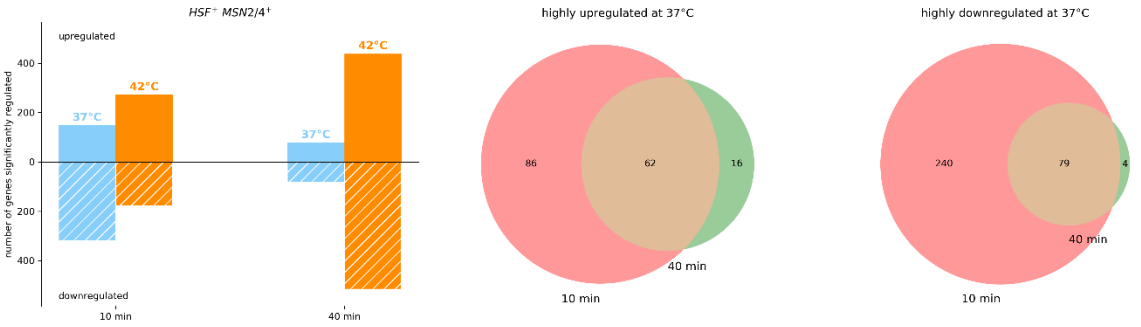

Figure 6, Overview of highly up- and downregulated genes in the WT strain. Left: number of genes more than 4-fold up or downregulated in response to the heat shock. Right: Venn diagrams for the gene sets of highly up- and downregulated genes in the HSR at 37°C.

The early transcriptome of the mild heat shock is highly enriched in genes for metabolic processes and glycogen metabolism (Figure 8). This also includes the upregulation of Trehalose metabolic processes ($p = 3.7 \times 10^{-3}$) which produce Trehalose as chemical chaperone. In case of the severe heat shock, these changes in the transcriptome of the metabolism are not as strong as in the mild heat shock. The increase of these processes is coupled to the burst phase of the HSR and is decreasing after 40 min of exposure. In case of the severe heat shock, no significant increase in these processes is found, as the cell is shifting to dormancy.

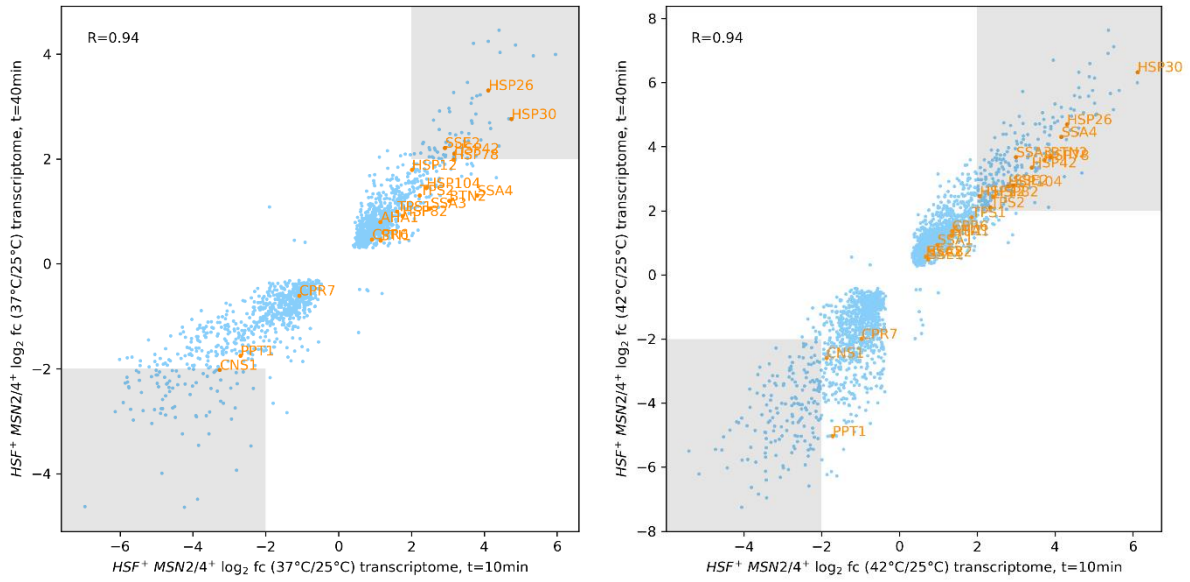

Figure 7, significant fold changes ($p < 0.05$) in expression for the wildtype cell lines after 10 and 40 minutes of exposure. Regions of high up- and downregulation ($|log_2\, fc| > 2$) Left: moderate stress at 37° Right: severe stress at 42°C. Selected HSPs are highlighted.

Even if some of the upregulated genes in the mild heat shock are HSPs and chaperones like HSP30 (26.7-fold increased, $p < 2.25 \times 10^{-4}$), HSP26 (17.25-fold increase, $p < 2.25 \times 10^{-4}$) or SSA4 (14.0-fold increased, $p < 2.25 \times 10^{-4}$), only these three could be identified under the 40-highest upregulated genes (Figure 7). After 40 minutes of exposure, this trend aggravates and only HSP30 (6.8-fold increase, $p < 1.0 \times 10^{-3}$) and HSP26 (9.8-fold increase, $p < 2.29 \times 10^{-4}$) are found under the 40-highest upregulated genes, as the transcriptome accommodates to the stimulus. No gene sets with annotations connected to chaperone activity or protein folding were found to be significantly enriched by the chosen method. This confirms previous observations (Mühlhofer et al., 2019), as heat shock proteins and chaperones only constitute a minor part of the HSR under mild conditions and become more important at severe conditions. This can be seen in the immediate response to the severe heat shock, where chaperones are highly

upregulated (Figure 7). Transcripts of the chaperone HSP30 are increased 71.5-fold ($p < 1.49 \times 10^{-4}$) after 10 minutes of severe heat shock compared to non-stress condition.

As the mild heat shock is not close enough to be lethal for the cell, it should not result in spore formation, which is its ultimate measure of survival. This is confirmed by enrichment analysis of the highly upregulated genes for the CC and BP GO annotations, which show no significant enrichment of spore components as a result of the stress (Figure 9). When the more severe heat shock is applied, spore formation is initiated and pursued over the course of the stress stimulus. This becomes apparent by significant enrichment of genes annotated with spore formation and spore compartments (Figure 9).

The highly downregulated transcripts in response to the heat shock with a more than 4-fold decrease in abundancy were analyzed likewise. For the mild heat shock, enrichment analysis of the 319 early downregulated genes shows association with ribosome biogenesis, location in the nucleolus and transcriptional regulation (Figure 10). This early response attenuates as the cell adapts to the heat shock and the enrichment of genes associated with ribosome biogenesis is decreased by magnitudes after 40 minutes of exposure. Most interestingly, no significant enrichment of transcripts coding for structural constituents of ribosomes was found. This mode of regulation agrees with previous studies of the HSR transcriptome where rRNA and ribosome levels stay relatively constant and only ribosome biogenesis is immediately affected. The early response to the severe heat shock is similar to the mild heat shock but won't attenuate under prolonged exposure. Additionally, structural components of the ribosome and protein subunits are found to be enriched under the highly downregulated genes after 40 minutes of severe heat shock. This initial decrease in ribosome biogenesis and ribosomes at later stage is indicating a regulated shutdown, as the general protein production decreases under these conditions (Mühlhofer et al., 2019).

GO:BP

Term Name	Term ID	H+_M+_37_10 p_adj	H+_M+_37_40 p_adj	H+_M+_42_10 p_adj	H+_M+_42_40 p_adj
energy reserve metabolic process	GO:0006112	4.001×10^{-7}	1.956×10^{-4}	3.426×10^{-5}	5.892×10^{-2}
carbohydrate metabolic process	GO:0005975	3.912×10^{-6}	2.597×10^{-1}	1.117×10^{-2}	1.388×10^{-1}
glycogen metabolic process	GO:0005977	5.882×10^{-6}	1.320×10^{-4}	2.703×10^{-4}	2.354×10^{-1}
generation of precursor metabolites and energy	GO:0006091	1.437×10^{-5}	2.628×10^{-1}	4.243×10^{-2}	1.000
response to temperature stimulus	GO:0009266	9.820×10^{-5}	2.130×10^{-2}	5.598×10^{-3}	1.000
response to heat	GO:0009408	2.536×10^{-4}	1.423×10^{-2}	4.642×10^{-2}	1.000
cellular carbohydrate metabolic process	GO:0044262	3.068×10^{-4}	9.326×10^{-2}	2.145×10^{-2}	5.030×10^{-1}
glycogen biosynthetic process	GO:0005978	4.950×10^{-4}	6.797×10^{-2}	1.423×10^{-2}	1.000
glucan metabolic process	GO:0044042	8.468×10^{-4}	8.257×10^{-3}	4.643×10^{-2}	1.000
trehalose metabolic process	GO:0005991	3.685×10^{-3}	1.000	6.080×10^{-2}	1.000

1 to 10 of 10 Page 1 of 1

KEGG

Term Name	Term ID	H+_M+_37_10 p_adj	H+_M+_37_40 p_adj	H+_M+_42_10 p_adj	H+_M+_42_40 p_adj
Starch and sucrose metabolism	KEGG:00500	2.669×10^{-4}	1.496×10^{-1}	3.989×10^{-1}	1.000

1 to 1 of 1 Page 1 of 1

WP

Term Name	Term ID	H+_M+_37_10 p_adj	H+_M+_37_40 p_adj	H+_M+_42_10 p_adj	H+_M+_42_40 p_adj
Glycogen Catabolism	WP:WP478	9.588×10^{-3}	9.561×10^{-2}	6.448×10^{-1}	1.000

Figure 8, Enrichment of the GO BP and Wikipathways and KEGG gene sets for the highly upregulated genes in response to the heat shock for the wildtype strain.

GO:BP		H+_M+_37_10	H+_M+_37_40	H+_M+_42_10	H+_M+_42_40
Term Name	Term ID	p_adj	p_adj	p_adj	p_adj ↑
spore wall assembly	GO:0042244	1.000	1.000	1.553×10^{-1}	6.221×10^{-6}
ascospore wall biogenesis	GO:0070591	1.000	1.000	1.553×10^{-1}	6.221×10^{-6}
spore wall biogenesis	GO:0070590	1.000	1.000	1.553×10^{-1}	6.221×10^{-6}
ascospore wall assembly	GO:0030476	1.000	1.000	1.553×10^{-1}	6.221×10^{-6}
fungal-type cell wall assembly	GO:0071940	1.000	1.000	1.851×10^{-1}	8.663×10^{-6}
cell wall assembly	GO:0070726	1.000	1.000	2.196×10^{-1}	1.195×10^{-5}
cellular component assembly involved in morphogenesis	GO:0010927	1.000	1.000	5.775×10^{-1}	3.040×10^{-5}
ascospore formation	GO:0030437	1.000	1.000	1.179×10^{-1}	4.543×10^{-5}
cell development	GO:0048468	1.000	1.000	1.179×10^{-1}	4.543×10^{-5}
cellular component morphogenesis	GO:0032989	1.000	1.000	7.065×10^{-1}	4.926×10^{-5}

GO:CC		H+_M+_37_10	H+_M+_37_40	H+_M+_42_10	H+_M+_42_40
Term Name	Term ID	p_adj ↑	p_adj	p_adj	p_adj
spore wall	GO:0031160	1.000	1.000	4.355×10^{-4}	6.632×10^{-9}
ascospore wall	GO:0005619	1.000	1.000	9.536×10^{-3}	3.619×10^{-7}
external encapsulating structure	GO:0030312	1.000	1.000	2.894×10^{-2}	2.921×10^{-5}
cell wall	GO:0005618	1.000	1.000	2.894×10^{-2}	2.921×10^{-5}
fungal-type cell wall	GO:0009277	1.000	1.000	7.312×10^{-2}	1.948×10^{-4}
ascospore-type prospore	GO:0042764	1.000	1.000	1.519×10^{-1}	6.649×10^{-4}
prospore membrane	GO:0005628	1.000	1.000	1.519×10^{-1}	6.649×10^{-4}
intracellular immature spore	GO:0042763	1.000	1.000	1.519×10^{-1}	6.649×10^{-4}
cell periphery	GO:0071944	1.000	1.000	1.330×10^{-2}	7.749×10^{-4}
intrinsic component of membrane	GO:0031224	1.000	1.000	3.178×10^{-2}	1.574×10^{-3}

Figure 9, enrichment of the GO CC and BP terms associated with sporulation for the highly upregulated genes in response to the heat shock for the wildtype strain.

GO:MF		H+_M+_37_10	H+_M+_37_40	H+_M+_42_10	H+_M+_42_40
Term Name	Term ID	p_adj ↑	p_adj	p_adj	p_adj
structural constituent of ribosome	GO:0003735	1.000	1.000	1.000	9.338×10^{-10}

GO:BP		H+_M+_37_10	H+_M+_37_40	H+_M+_42_10	H+_M+_42_40
Term Name	Term ID	p_adj ↑	p_adj	p_adj	p_adj
ribosome biogenesis	GO:0042254	8.881×10^{-121}	9.855×10^{-4}	3.471×10^{-92}	7.021×10^{-129}
ncRNA processing	GO:0034470	1.091×10^{-118}	7.432×10^{-1}	8.036×10^{-83}	1.631×10^{-103}
ribonucleoprotein complex biogenesis	GO:0022613	7.370×10^{-110}	4.081×10^{-3}	4.942×10^{-87}	3.316×10^{-120}
ncRNA metabolic process	GO:0034660	9.115×10^{-109}	1.000	2.379×10^{-74}	4.656×10^{-100}
rRNA processing	GO:0006364	4.898×10^{-106}	1.223×10^{-1}	8.676×10^{-76}	1.781×10^{-98}
rRNA metabolic process	GO:0016072	7.642×10^{-106}	1.682×10^{-1}	4.840×10^{-73}	2.226×10^{-95}
RNA processing	GO:0006396	3.258×10^{-89}	1.000	1.739×10^{-66}	1.376×10^{-74}
RNA metabolic process	GO:0016070	6.948×10^{-56}	1.000	2.412×10^{-40}	1.059×10^{-32}
cytoplasmic translation	GO:0002181	1.000	1.000	1.000	2.513×10^{-21}
translation	GO:0006412	1.000	1.000	1.000	9.618×10^{-14}
peptide biosynthetic process	GO:0043043	1.000	1.000	1.000	2.117×10^{-13}

GO:CC		H+_M+_37_10	H+_M+_37_40	H+_M+_42_10	H+_M+_42_40
Term Name	Term ID	p_adj ↑	p_adj	p_adj	p_adj
nucleolus	GO:0005730	6.008×10^{-133}	4.437×10^{-4}	9.653×10^{-94}	4.761×10^{-111}
cytosolic ribosome	GO:0022626	1.000	1.000	1.000	3.435×10^{-21}
ribosome	GO:0005840	1.000	1.000	1.000	1.358×10^{-13}
ribosomal subunit	GO:0044391	1.000	1.000	1.000	2.254×10^{-13}

KEGG		H+_M+_37_10	H+_M+_37_40	H+_M+_42_10	H+_M+_42_40
Term Name	Term ID	p_adj ↑	p_adj	p_adj	p_adj
Ribosome biogenesis in eukaryotes	KEGG:03008	1.105×10^{-28}	1.000	4.825×10^{-20}	1.605×10^{-16}
RNA polymerase	KEGG:03020	2.548×10^{-12}	1.000	5.594×10^{-6}	5.256×10^{-8}
Ribosome	KEGG:03010	1.000	1.000	1.000	1.133×10^{-10}

WP		H+_M+_37_10	H+_M+_37_40	H+_M+_42_10	H+_M+_42_40
Term Name	Term ID	p_adj ↑	p_adj	p_adj	p_adj
Eukaryotic Transcription Initiation	WP:WP425	8.332×10^{-4}	1.000	2.361×10^{-1}	1.000
Cytoplasmic Ribosomal Proteins	WP:WP210	1.000	1.000	1.000	2.570×10^{-9}
Translation Factors	WP:WP32	1.000	1.000	3.314×10^{-1}	2.011×10^{-2}

Figure 10, enrichment of the GO MF, CC, BP, and Wikipathways and KEGG gene sets for the highly downregulated genes in response to the heat shock for the wildtype strain.

Transcriptional regulation of the HSR

The expression of genes is depending on specific TFs, which enhance or silence the transcription machinery. Even though this regulation can happen in a simple, switch like manner, the transcription of genes in eukaryotes is most often regulated by a delicate network of TFs, interaction partners, regulatory modifications and translocations. The role of specific factors can be probed by the production of knockout cell lines, which miss only a single gene and the subsequent study of the transcriptome. In this case, the production of the HSF1 and MSN2/4 single and double knockout cell lines was used to investigate the function of the two TFs and ways, how these TFs interact.

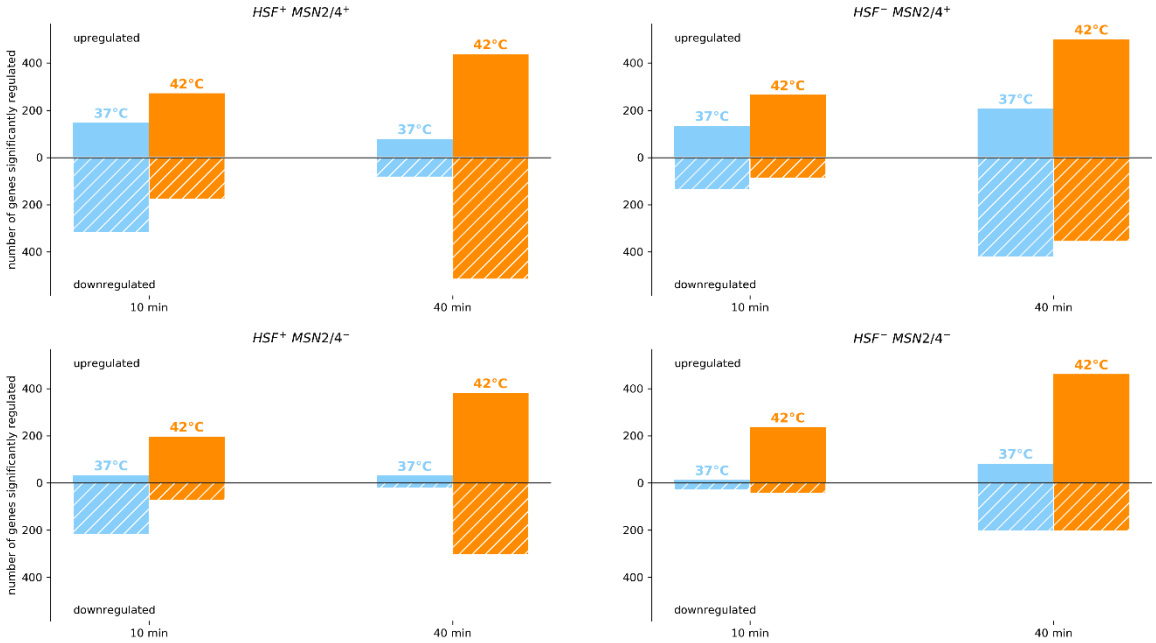

Figure 11, highly up- and downregulated genes for the different knockouts and conditions. Genes with a fold change larger than 4 were considered highly regulated.

The relevance of the TFs MSN2/4 and HSF1 in regulating the heat shock response can clearly be seen by the number of highly up and downregulated genes after exposure to the heat shock. Genes were considered highly upregulated in this analysis when their transcript changed at least 4-fold compared to the wild type strain at 25°. The result of the MSN2/4 Knockout was a decrease in highly upregulated genes after 10 minutes (33 vs 148) and after 40 minutes (31 vs 78) of exposure to the mild heat shock (Figure 11). The number of highly downregulated genes was affected likewise, although not as strong. Only 21 genes were highly downregulated after 40 minutes of exposure to the mild heat shock. This could indicate that the cell can cope with a mild heat shock even without factors coordinated by MSN2 and MSN4. This hypothesis gains further support from the transcriptome of the HSF1 knockout strain under mild heat shock conditions. Although the initial response in up and downregulated genes is smaller than in the wildtype, the response wont attenuate under prolonged stress and increases to (207 vs 78) upregulated and (421 vs 83) downregulated genes. The functional annotation of the downregulated genes seems to be similar to the observed annotation for the severe heat shock in the wildtype strain. An initial decrease in ribosome biogenesis genes is followed by further decrease of ribosome constituents (Figure 12) as the growth of the cell is inhibited by the heat stress. This underlines the crucial role of HSF1 for the successful adaption of the cell to the heat shock in contrast to MSN2 and MSN4.

Even though the highly downregulated genes seem to be comparable to the wildtype HSR under severe conditions, the response differs in the annotation of the upregulated genes as no enrichment in sporulation associated genes is found (Figure 13). This could mean that a decrease in growth potential by downregulation of the ribosomal machinery isn't enough to stimulate sporulation and hints at sporulation being a secondary independent process. Under severe conditions, genes associated with sporulation seem to be enriched in the HSF1 knockout analogous to the observations in the wildtype. Which could indicate that HSF1 is not critical in the regulation of this secondary process.

GO:BP		H-_M+_37_10	H-_M+_37_40	H-_M+_42_10	H-_M+_42_40
Term Name	Term ID	p_adj	p_adj	p_adj	p_adj
ribosome biogenesis	GO:0042254	3.642×10^{-90}	1.008×10^{-112}	3.690×10^{-59}	1.697×10^{-114}
ribonucleoprotein complex biogenesis	GO:0022613	7.945×10^{-82}	6.059×10^{-101}	7.577×10^{-54}	6.142×10^{-104}
rRNA processing	GO:0006364	2.564×10^{-72}	6.560×10^{-84}	4.952×10^{-46}	1.820×10^{-82}
rRNA metabolic process	GO:0016072	1.434×10^{-71}	5.926×10^{-82}	1.430×10^{-44}	1.091×10^{-79}
ribosomal small subunit biogenesis	GO:0042274	3.555×10^{-30}	4.830×10^{-54}	5.388×10^{-16}	7.738×10^{-44}
ribosomal large subunit biogenesis	GO:0042273	8.112×10^{-41}	7.070×10^{-47}	3.898×10^{-26}	6.810×10^{-49}
cytoplasmic translation	GO:0002181	1.000	1.451×10^{-19}	1.000	5.384×10^{-14}
translation	GO:0006412	1.000	1.708×10^{-8}	1.000	3.825×10^{-14}
GO:CC		H-_M+_37_10	H-_M+_37_40	H-_M+_42_10	H-_M+_42_40
Term Name	Term ID	p_adj	p_adj	p_adj	p_adj
nucleolus	GO:0005730	1.847×10^{-95}	1.369×10^{-94}	1.346×10^{-62}	2.476×10^{-97}
preribosome	GO:0030684	5.567×10^{-73}	3.661×10^{-82}	8.041×10^{-43}	1.045×10^{-79}
cytosolic ribosome	GO:0022626	1.000	4.620×10^{-24}	1.000	9.858×10^{-13}
ribosome	GO:0005840	1.000	1.661×10^{-15}	1.000	1.598×10^{-8}
ribosomal subunit	GO:0044391	1.000	6.683×10^{-15}	1.000	5.990×10^{-7}
KEGG		H-_M+_37_10	H-_M+_37_40	H-_M+_42_10	H-_M+_42_40
Term Name	Term ID	p_adj	p_adj	p_adj	p_adj
Ribosome biogenesis in eukaryotes	KEGG:03008	5.396×10^{-22}	2.020×10^{-19}	4.741×10^{-14}	1.069×10^{-16}
Ribosome	KEGG:03010	1.000	6.181×10^{-18}	1.000	4.313×10^{-6}
RNA polymerase	KEGG:03020	3.086×10^{-2}	9.312×10^{-8}	7.699×10^{-3}	3.904×10^{-6}
WP		H-_M+_37_10	H-_M+_37_40	H-_M+_42_10	H-_M+_42_40
Term Name	Term ID	p_adj	p_adj	p_adj	p_adj
Cytoplasmic Ribosomal Proteins	WP:WP210	1.000	7.513×10^{-15}	1.000	1.761×10^{-4}

Figure 12, enrichment of the GO BP, CC, Wikipathways and KEGG gene sets for the highly downregulated genes in response to the heat shock for the HSF1 knockout strain under different conditions.

GO:BP		H-_M+_37_10	H-_M+_37_40	H-_M+_42_10	H-_M+_42_40
Term Name	Term ID	p_adj	p_adj	p_adj	p_adj
spore wall assembly	GO:0042244	1.000	1.000	1.288×10^{-3}	4.028×10^{-8}
ascospore wall assembly	GO:0030476	1.000	1.000	1.288×10^{-3}	4.028×10^{-8}
ascospore wall biogenesis	GO:0070591	1.000	1.000	1.288×10^{-3}	4.028×10^{-8}
spore wall biogenesis	GO:0070590	1.000	1.000	1.288×10^{-3}	4.028×10^{-8}
sporulation	GO:0043934	1.000	1.000	3.645×10^{-3}	1.101×10^{-5}

Figure 13, enrichment of the GO BP gene sets associated with sporulation for the highly upregulated genes in response to the heat shock for the HSF1 knockout strain under different conditions.

The transcriptomic data was further analyzed for the enrichment of Yeastract association gene sets, described previously. The rationale behind this analysis is the idea, that secondary activation of transcription factors will increase the abundance of transcripts positively affected by this transcription factor. This correlation could be used to identify activated secondary TFs. But as genes are not exclusively affected by a single transcription factor, inferring the activation of a TF by its Target Genes should investigated on an individual basis and the overlap and exclusivity of the gene set should be assessed. This approach could nevertheless be useful to generate hypothesis about activation networks or confirm known interactions.

In order to focus on the primary HSR without secondary effects due to metabolic remodeling, disturbed proteostasis or other effects, the early transcriptome under mild heat shock conditions was analyzed. As expected, genes positively associated with MSN2 and MSN4 activation are highly enriched in the early wildtype HSR (Figure 14). This Effect is highly attenuated by the knockout of MSN2/MSN4 and HSF1 but also slightly affected by the knockout of HSF1. Most interestingly, the knockout of MSN2/MSN4 doesn't lead to a complete decrease in enrichment comparable to the double knockout. This can be explained by the previous analysis of the gene set overlap of MSN2/MSN4 and HSF1. The MSN2/MSN4 associations were accordingly divided into a core component, shared by HSF1 and a MSN2/MSN4 exclusive component. This enrichment analysis supports this separation, as the core stress response is enriched in the single knockout of both TFs and is only attenuated by the double knockout. The MSN2/MSN4 exclusive gene set is though only enriched in presence of MSN2/MSN4 and no significant enrichment can be observed for the single and double knockouts. The HSF1 exclusive associations are not significantly enriched. Enrichment of HSF1 targets is only slightly affected by the singly MSN2/MSN4 knockout and

only the double knockout leads to a clear decrease. It should be noted that this behavior can be explained by the subset of core components, which are behaving similar.

Most of the other Transcription factors associations seem to be linked to the presence of MSN2/MSN4 and the enrichment is close to the wildtype in the HSF1 knockout. The knockout of MSN2/MSN4 leads to a clear decrease in enrichment of genes associated with RLM1, a TF involved in MAPK pathway, which is known to be interacting with the MSN2/MSN4 stress response. The enrichment of positive RLM1 associations is found with high significance ($p < 1.02 \times 10^{-50}$) in the wildtype and low or no significance in the single and double knockout. Most of the TFs with enriched associations in the wildtype are found to behave in this MSN2/MSN4 dependent way, independent of HSF1. Some of these TFs like CIN5, HOT1 or SKO1 are known to be associated with the MAP kinase HOG1. Another interesting genetic interaction is found for AFT1. Genes associated with negative regulation by AFT1 are found to be enriched in the wildtype HSR and the HSF1 knockout respectively in the presence of MSN2/MSN4. The single or double knockout of MSN2/MSN4 leads to no significant enrichment of AFT1 negatively associated genes, which could be interpreted as inhibitory interaction between ATF1 and MSN2/MSN4. Only very few TF associations with low significance of enrichment are found to be enhanced by the MSN2/MSN4 knockout, compared to the wildtype and HSF1 knockout. These TFs, HAA1, SKN7 and ABF1 may be negatively regulated by MSN2/4.

Analyzing the transcriptome by methods of enrichment analysis can only give a broad overview of the regulatory interactions and the inference of causal relations should be used for hypothesis building at most. Dividing the transcriptome into highly up- and downregulated genes is based on an arbitrary threshold, which is applied uniformly to all genes. This oversimplifies biological reality, as the dynamic range in transcript abundancies is not uniform and varies for each protein class. Effects observed in in the functional enrichment analysis may appear more prominent, even if they only take a minor role in the HSR. Likewise, more delicate effects of the HSR will be missed out by this analysis as they won't lead to significant overrepresentation of genes. The same may happen with highly upregulated genes, which act only in small compact pathway. This can be seen with the few heat shock proteins, which are highly abundant in the heat shock response, but are not identified by the enrichment analysis. More previous knowledge needs to be integrated into the analysis to identify and characterize these parts of the HSR.

yeastract+		H+_M+_37_10	H+_M-_37_10	H-_M+_37_10	H-_M-_37_10
Term Name	Term ID	p_adj ↑	p_adj	p_adj	p_adj
MSN4_Yeastract_positive_stress_association	MSN4_up	1.179×10^{-91}	3.217×10^{-22}	1.111×10^{-70}	8.124×10^{-5}
MSN2_Yeastract_positive_stress_association	MSN2_up	5.985×10^{-81}	5.434×10^{-20}	6.481×10^{-62}	3.077×10^{-4}
RLM1_Yeastract_positive_stress_association	RLM1_up	1.020×10^{-50}	3.048×10^{-5}	4.331×10^{-49}	1.396×10^{-1}
core_stress_associations	core_exup	5.577×10^{-45}	6.653×10^{-25}	3.822×10^{-42}	3.913×10^{-4}
HSF1_Yeastract_positive_stress_association	HSF1_up	1.382×10^{-36}	2.406×10^{-23}	7.057×10^{-32}	5.661×10^{-3}
CIN5_Yeastract_positive_stress_association	CIN5_up	9.462×10^{-28}	7.883×10^{-9}	4.959×10^{-28}	1.810×10^{-2}
BAS1_Yeastract_positive_stress_association	BAS1_up	4.605×10^{-24}	1.748×10^{-2}	1.200×10^{-17}	3.831×10^{-1}
MSN2/4_exclusive_positive_stress_associations	MSN2/4_exup	6.293×10^{-24}	1.000	2.482×10^{-14}	1.000
STB5_Yeastract_negative_stress_association	STB5_down	5.056×10^{-21}	2.827×10^{-5}	8.014×10^{-14}	1.305×10^{-2}
ARR1_Yeastract_negative_stress_association	ARR1_down	1.011×10^{-20}	4.583×10^{-1}	2.346×10^{-22}	9.046×10^{-1}
GCN4_Yeastract_positive_stress_association	GCN4_up	4.194×10^{-17}	7.455×10^{-7}	2.637×10^{-12}	1.378×10^{-2}
HAC1_Yeastract_positive_stress_association	HAC1_up	2.094×10^{-16}	1.280×10^{-1}	5.929×10^{-14}	8.556×10^{-3}
PDR1_Yeastract_positive_stress_association	PDR1_up	8.202×10^{-16}	4.052×10^{-2}	6.214×10^{-13}	1.000
RPN4_Yeastract_positive_stress_association	RPN4_up	2.907×10^{-15}	1.672×10^{-2}	5.811×10^{-13}	2.510×10^{-1}
HOT1_Yeastract_positive_stress_association	HOT1_up	7.342×10^{-15}	7.690×10^{-5}	1.379×10^{-11}	1.342×10^{-2}
IXR1_Yeastract_negative_stress_association	IXR1_down	3.231×10^{-14}	5.785×10^{-6}	2.471×10^{-11}	2.805×10^{-6}
GIS1_Yeastract_positive_stress_association	GIS1_up	5.800×10^{-14}	1.081×10^{-4}	2.449×10^{-13}	1.000
MIG1_Yeastract_positive_stress_association	MIG1_up	2.607×10^{-13}	6.499×10^{-10}	1.137×10^{-12}	2.568×10^{-2}
MET4_Yeastract_negative_stress_association	MET4_down	2.185×10^{-11}	1.779×10^{-3}	2.472×10^{-11}	6.647×10^{-1}
GLN3_Yeastract_positive_stress_association	GLN3_up	1.644×10^{-10}	7.917×10^{-5}	4.049×10^{-5}	2.354×10^{-2}
AFT1_Yeastract_negative_stress_association	AFT1_down	2.374×10^{-10}	1.000	4.758×10^{-12}	1.000
YRR1_Yeastract_positive_stress_association	YRR1_up	4.018×10^{-10}	1.052×10^{-1}	3.234×10^{-5}	1.000
YAP1_Yeastract_positive_stress_association	YAP1_up	4.006×10^{-8}	8.070×10^{-3}	1.086×10^{-4}	4.299×10^{-1}
CRZ1_Yeastract_positive_stress_association	CRZ1_up	6.407×10^{-8}	3.786×10^{-2}	1.434×10^{-7}	1.000
ROX1_Yeastract_positive_stress_association	ROX1_up	2.265×10^{-5}	1.496×10^{-3}	3.613×10^{-2}	1.000
PDR3_Yeastract_positive_stress_association	PDR3_up	2.904×10^{-5}	1.000	2.513×10^{-2}	9.099×10^{-1}
YOX1_Yeastract_positive_stress_association	YOX1_up	6.488×10^{-5}	1.000	4.940×10^{-6}	1.000
YRR1_Yeastract_negative_stress_association	YRR1_down	1.262×10^{-4}	4.055×10^{-1}	4.605×10^{-3}	1.000
CAD1_Yeastract_positive_stress_association	CAD1_up	1.653×10^{-4}	1.000	5.154×10^{-4}	8.128×10^{-1}
HAA1_Yeastract_positive_stress_association	HAA1_up	2.037×10^{-4}	8.707×10^{-7}	1.305×10^{-2}	1.000
SKO1_Yeastract_positive_stress_association	SKO1_up	1.231×10^{-3}	1.000	3.705×10^{-5}	1.000
YAP5_Yeastract_positive_stress_association	YAP5_up	2.863×10^{-3}	1.410×10^{-1}	3.289×10^{-3}	1.000
SKN7_Yeastract_positive_stress_association	SKN7_up	3.663×10^{-3}	2.715×10^{-5}	3.175×10^{-2}	1.000
ZAP1_Yeastract_positive_stress_association	ZAP1_up	5.065×10^{-3}	1.000	2.192×10^{-1}	4.129×10^{-1}
AFT2_Yeastract_negative_stress_association	AFT2_down	6.779×10^{-3}	1.924×10^{-1}	2.727×10^{-2}	3.085×10^{-1}
ABF1_Yeastract_positive_stress_association	ABF1_up	1.489×10^{-2}	2.256×10^{-7}	4.596×10^{-2}	4.278×10^{-1}
MSN1_Yeastract_positive_stress_association	MSN1_up	1.846×10^{-1}	1.177×10^{-1}	1.000	4.539×10^{-2}
RIM101_Yeastract_negative_stress_association	RIM101_down	5.307×10^{-1}	3.275×10^{-2}	1.000	1.000
MIG3_Yeastract_positive_stress_association	MIG3_up	1.000	1.888×10^{-2}	1.000	1.000

Figure 14, The highly upregulated genes of the early mild HSR are probed for enrichment of gene sets with positive and negative TF associations from the Yeastract database.

Further investigations

Functional enrichment of gene sets was performed with the DAVID bioinformatics tool and the g:profiler functional enrichment software. Even though ranked gene lists were used with g:profiler when applicable, the full potential of the transcriptomic data for enrichment analysis wasn't used and results should be interpreted accordingly. Further investigations should consolidate the found enrichments by comparison of multiple enrichment methods and full use of the ranked gene sets. Methods for comparative analysis of multiple enrichment methods is available in software packages like EnrichmentBrowser (Geistlinger et al., 2016).

To get further insight into to the interplay between the two TFs, an expression component analysis could be performed for the different experimental conditions. This analysis can be performed on knockout transcriptomics datasets and was used previously to study the regulatory interaction between HOG1 and MSN2 (Capaldi et al., 2008). The analysis allows to study the regulatory contribution of the two TFs and dissects the expression into two TF components and a cooperative component. One could then draw conclusions about the interaction between the two TFs and functional segregation of the stress response due to this interaction.

References

Anckar, J., & Sistonen, L. (2011). Regulation of HSF1 function in the heat stress response: Implications in aging and disease. *Annual Review of Biochemistry*, *80*, 1089–1115. https://doi.org/10.1146/annurev-biochem-060809-095203

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., & Rubin, Gerald M. Sherlock, G. (2000). The Gene Ontology Consortium, Michael Ashburner1, Catherine A. Ball3, Judith A. Blake4, David Botstein3, Heather Butler1, J. Michael Cherry3, Allan P. Davis4, Kara Dolinski3, Selina S. Dwight3, Janan T. Eppig4, Midori A. Harris3, David P. Hill4, Laurie Is. *Nature Genetics*, *25*(1), 25–29. https://doi.org/10.1038/75556.Gene

Bateman, A. (2019). UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Research*, *47*(D1), D506–D515. https://doi.org/10.1093/nar/gky1049

Capaldi, A. P., Kaplan, T., Liu, Y., Habib, N., Regev, A., Friedman, N., & O'shea, E. K. (2008). Structure and function of a transcriptional network activated by the MAPK Hog1. *Nature Genetics*, *40*(11), 1300–1306. https://doi.org/10.1038/ng.235

Cherry, J. M., Hong, E. L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E. T., Christie, K. R., Costanzo, M. C., Dwight, S. S., Engel, S. R., Fisk, D. G., Hirschman, J. E., Hitz, B. C., Karra, K., Krieger, C. J., Miyasato, S. R., Nash, R. S., Park, J., Skrzypek, M. S., … Wong, E. D. (2012). Saccharomyces Genome Database: The genomics resource of budding yeast. *Nucleic Acids Research*, *40*(D1), 700–705. https://doi.org/10.1093/nar/gkr1029

Durchschlag, E., Reiter, W., Ammerer, G., & Schüller, C. (2004). Nuclear localization destabilizes the stress-regulated transcription factor Msn2. *Journal of Biological Chemistry*, *279*(53), 55425–55432. https://doi.org/10.1074/jbc.M407264200

Geistlinger, L., Csaba, G., & Zimmer, R. (2016). Bioconductor's EnrichmentBrowser: Seamless navigation through combined results of set- & network-based enrichment analysis. *BMC Bioinformatics*, *17*(1), 1–11. https://doi.org/10.1186/s12859-016-0884-1

Görner, W., Durchschlag, E., Martinez-Pastor, M. T., Estruch, F., Ammerer, G., Hamilton, B., Ruis, H., & Schüller, C. (1998). Nuclear localization of the C2H2 zinc finger protein Msn2p is regulated by stress and protein kinase A activity. *Genes and Development*, *12*(4), 586–597. https://doi.org/10.1101/gad.12.4.586

Hartl, F. U., Bracher, A., & Hayer-Hartl, M. (2011). Molecular chaperones in protein folding and proteostasis. *Nature*, *475*(7356), 324–332. https://doi.org/10.1038/nature10317

Herwig, R., Hardt, C., Lienhard, M., & Kamburov, A. (2016). Analyzing and interpreting genome data at the network level with ConsensusPathDB. *Nature Protocols*, *11*(10), 1889–1907. https://doi.org/10.1038/nprot.2016.117

Hong, E. L., Balakrishnan, R., Dong, Q., Christie, K. R., Park, J., Binkley, G., Costanzo, M. C., Dwight, S. S., Engel, S. R., Fisk, D. G., Hirschman, J. E., Hitz, B. C., Krieger, C. J., Livstone, M. S., Miyasato, S. R., Nash, R. S., Oughtred, R., Skrzypek, M. S., Weng, S., … Cherry, J. M. (2008). Gene Ontology annotations at SGD: New data sources and annotation methods. *Nucleic Acids Research*, *36*(SUPPL. 1), 577–581. https://doi.org/10.1093/nar/gkm909

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Kanehisa, M. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, *28*(1), 27–30. https://doi.org/10.1093/nar/28.1.27

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., & Willing, C. (2016). Jupyter Notebooks—a publishing format for reproducible computational workflows. *Positioning and Power in Academic Publishing: Players, Agents and Agendas - Proceedings of the 20th International Conference on Electronic Publishing, ELPUB 2016*, 87–90. https://doi.org/10.3233/978-1-61499-649-1-87

Lee, P., Cho, B. R., Joo, H. S., & Hahn, J. S. (2008). Yeast Yak1 kinase, a bridge between PKA and stress-responsive transcription factors, Hsf1 and Msn2/Msn4. *Molecular Microbiology*, *70*(4), 882–895. https://doi.org/10.1111/j.1365-2958.2008.06450.x

McKinney, W. (2004). Data Structures for Statistical Computing in Python. In *Proceedings of the 9th Python in Science Conference* (Vol. 41, Issue 3, pp. 62–72).

Monteiro, P. T., Oliveira, J., Pais, P., Antunes, M., Palma, M., Cavalheiro, M., Galocha, M., Godinho, C. P., Martins, L. C., Bourbon, N., Mota, M. N., Ribeiro, R. A., Viana, R., Sá-Correia, I., & Teixeira, M. C. (2020). YEASTRACT+: A portal for cross-species comparative genomics of transcription regulation in yeasts. *Nucleic Acids Research*, *48*(D1), D642–D649. https://doi.org/10.1093/nar/gkz859

Mühlhofer, M., Berchtold, E., Stratil, C. G., Csaba, G., Kunold, E., Bach, N. C., Sieber, S. A., Haslbeck, M., Zimmer, R., & Buchner, J. (2019). The Heat Shock Response in Yeast Maintains Protein Homeostasis by Chaperoning and Replenishing Proteins. *Cell Reports*, *29*(13), 4593-4607.e8. https://doi.org/10.1016/j.celrep.2019.11.109

Oliphant, T. E. (2007). Python for Scientific Computing. *Computing in Science & Engineering*, *9*(3), 10–20. https://doi.org/10.1109/MCSE.2007.58

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., & Vilo, J. (2019). G:Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research*, *47*(W1), W191–W198. https://doi.org/10.1093/nar/gkz369

Rossum, G. Van. (1995). Python tutorial. *Technical Report CS-R9526*, *Centrum vo*.

Rousseau, A., & Bertolotti, A. (2016). An evolutionarily conserved pathway controls proteasome homeostasis. *Nature*, *536*(7615), 184–189. https://doi.org/10.1038/nature18943

Schmitt, A. P., & McEntee, K. (1996). Msn2p, a zinc finger DNA-binding protein, is the transcriptional activator of the multistress response in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(12), 5777–5782. https://doi.org/10.1073/pnas.93.12.5777

Slenter, D. N., Kutmon, M., Hanspers, K., Riutta, A., Windsor, J., Nunes, N., Mélius, J., Cirillo, E., Coort, S. L., DIgles, D., Ehrhart, F., Giesbertz, P., Kalafati, M., Martens, M., Miller, R., Nishida, K., Rieswijk, L., Waagmeester, A., Eijssen, L. M. T., … Willighagen, E. L. (2018). WikiPathways: A multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Research*, *46*(D1), D661–D667. https://doi.org/10.1093/nar/gkx1064

Solís, E. J., Pandey, J. P., Zheng, X., Jin, D. X., Gupta, P. B., Airoldi, E. M., Pincus, D., & Denic, V. (2016). Defining the Essential Function of Yeast Hsf1 Reveals a Compact Transcriptional Program for Maintaining Eukaryotic Proteostasis. *Molecular Cell*, *63*(1), 60–71. https://doi.org/10.1016/j.molcel.2016.05.014