Resolving noise-control conflict by gene duplication

Authors: Michal Chapal[1], Sefi Mintzer[1], Sagie Brodsky[1], Miri Carmi[1] and Naama Barkai[1]†

Affiliations:

[1]Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel.

†Correspondence to: naama.barkai@weizmann.ac.il

Abstract:

Gene duplication promotes adaptive evolution in two principle ways: allowing one duplicate to evolve a new function and resolving adaptive conflicts by splitting ancestral functions between the duplicates. In an apparent departure from both scenarios, low-expressing transcription factor (TF) duplicates commonly regulate similar sets of genes and act in overlapping conditions. To examine for possible benefits of such apparently redundant duplicates, we examined the budding yeast duplicated stress regulators Msn2 and Msn4. We show that Msn2,4 indeed function as one unit, inducing the same set of target genes in overlapping conditions, yet this two-factor composition allows its expression to be both environmental-responsive and with low-noise, thereby resolving an adaptive conflict that inherently limits expression of single genes. Our study exemplified a new model for evolution by gene duplication whereby duplicates provide adaptive benefit through cooperation, rather than functional divergence: attaining two-factor dynamics with beneficial properties that cannot be achieved by a single gene.

Introduction:

The number of transcription factor (TFs) expressed in eukaryotes increases rapidly with increasing genome size and organism complexity, ranging from ~50 in obligate parasites to >1000 in high eukaryotes[1]. Gene duplication played a major role in this evolutionary expansion[2,3], as is evident from the fact that the number of DNA Binding Domains (DBD) remained practically constant with the increasing genome size. In fact, the majority of TFs belong to just a few DBD families[2,3], the content of which increased rapidly with genome size. Understanding the adaptive forces that promote this duplication-dependent expansion is therefore of a great interest.

Gene duplication can promote evolution by allowing one of the duplicates to adapt a novel function while the second duplicate maintains the ancestral function. More often, however, the two duplicates do not gain a new function but rather lose complementary subsets of ancestral functions[4,5]. Sub-functionalization not only explains duplicate maintenance, but can also promote adaptive evolution by enabling further optimization of each individual function and resolving adaptive conflicts of the ancestral gene[6,7]. Indeed, optimizing a dual-function

protein is often constrained by conflicting requirements imposed by the different functions: a mutation that favor one function can perturb the other function, presenting an adaptive conflict that is only resolved upon duplication.

In the context of TFs, duplication may allow one factor to acquire a new set of target genes (neo-functionalization). Alternatively, the ancestral targets could split between the duplicates (sub-functionalization). In both scenarios, duplicate divergence would increase and refine the regulatory logic. Previous studies exemplified both scenarios[8–10], but whether they are relevant for the majority of TF duplicates remained unclear. In fact, present-day genomes express a large number of TF duplicates that regulate highly similar, or even redundant, set of targets.

Budding yeast provide a convenient platform for studying the possible adaptive role of apparently redundant duplicates. The yeast lineage underwent a Whole Genome Duplication (WGD) event about one hundred million years ago[11], and while most duplicates generated in this event were lost, about 10% were retained, amongst which TF are over-represented. Many of the retained TF duplicates show little signs of divergence but rather retained a highly conserved DBD, and, accordingly, bind the same DNA binding motif and regulate a highly similar set of genes. We reasoned that studying such duplicates might help shed light on possible benefits provided by TF duplication.

As a case in point, we investigated Msn2 and Msn4, a well-studied TF duplicates in budding yeast, which activate a large number of targets of the environmental stress response[12,13]. Previous studies established the tight similarity between these factors[14], but we still began our study by increasing experimental resolution in an attempt to detect target divergence. Our results, however, re-enforced the conclusion that the two factors regulate the same set of target genes, translocate to the nucleus with the precise same dynamics, and expressed in an overlapping set of conditions.

Our search for differences between the duplicates pointed us in a different direction: the challenge which cells face when attempting to minimize noise in gene expression. Transcription is a stochastic process and is therefore characterized by random variations (noise) between genetically identical cells[15,16]. Cell to cell expression variability is deleterious for genes that require precise tuning of expression, such as dosage sensitive genes[17,18], but beneficial when enabling processes not possible by deterministic dynamics, such as bet-hedging strategies[19–21]. Accordingly, noise levels vary greatly between genes[22]. Yet, the ability to tune expression noise through changes in gene promoter is limited by mechanistic constraints, and in particular by the well-documented conflict between regulatory control and noise: genes that are regulated over a wide dynamic range also show a high level of expression noise[15,23–25]. Thus, while coding for low-noise expression is possible, it comes at the cost of lowering the dynamic range over which expression can be changed by regulatory signals.

Our study shows that through duplication, Msn2,4 resolved this interplay between environmental responsive and noise. Following duplication, Msn2 expression became highly stable, showing limited responsiveness to environmental conditions and low expression noise.

By contrast, Msn4 expression accentuated the environmental-responsive expression of the unduplicated homologue. This resulted in an overall expression of the Msn2,4 unit that is responsive to the environment but is also low-noise at the low expressing conditions. We provide evidence that this expression tuning is phenotypically adaptive, and define the genetic changes that correlates with the change in gene responsiveness and noise. Our results suggest that duplicates can promote adaptive evolution not only through functional divergence, as suggested by the neo- or sub-functionalization models, but also through effective cooperation, by attaining two-factor dynamics with emergence beneficial properties that cannot be achieved using a single gene.

Results:

Low-noise (Poisson) distribution of MSN2 expression in individual cells

Msn2 and Msn4 are TF duplicates that regulate the stress response in budding yeast[12,26]. Stress genes show a noisy expression, and we were therefore surprised to observe that Msn2 is expressed at very similar amounts across individual cells. In fact, Msn2-GFP was the least noisy of all proteins expressed at its level, as quantified in a study surveying >2500 GFP-fused proteins[22] (Figure 1A). To examine whether this low noise is also seen at the *MSN2* transcript level, we used single-molecule Fluorescent In-Situ Hybridization (smFISH)[27] technique (Figure 1B). The number of *MSN2* transcripts in individual cells was well described by a Poisson distribution, as expected when individual mRNA transcripts are produced and degraded at constant rates[28,29]. This distribution presents the lower limit of gene expression noise, obtained in the absence of regulation, and other noise-amplifying processes[28].

Increasing Msn2 expression promotes stress protection but reduces cell growth rate:

Low expression noise characterizes genes coding for essential functions or components of large complexes[29,30], for which expression tuning is beneficial[30–32]. By contrast, Msn2 is not essential, does not participate in large complexes, and is mostly inactive in rich media. To examine whether, and how, Msn2 expression level impacts cell fitness, we engineered a library of strains expressing Msn2 at gradually increasing amounts using synthetic promoters[33]. Measuring growth rates of the library strains using a sensitive competition assay (Figure 1C), we found that decreasing Msn2 expression to below its wild-type levels, and down to a complete deletion, had no detectable effect on growth rate within the resolution of our assay (0.5%). By contrast, increasing Msn2 abundance gradually decreased growth rate (Figure 1D,S1). Next, we tested the effect of Msn2 levels on the ability to proliferate in harsh stress, by incubating the library cells with high H_2O_2 concentrations (Figure 1C,E,S1). Here, increasing Msn2 levels was beneficial: cells that expressed high levels of Msn2 resumed growth faster than low-expressing ones. Therefore, increasing Msn2 expression better protects cells against stress, but reduces their growth rate. An optimal Msn2 level is therefore desirable to balance the need for rapid growth and stress protection, explaining the requirement for low-noise tuning of its gene expression.

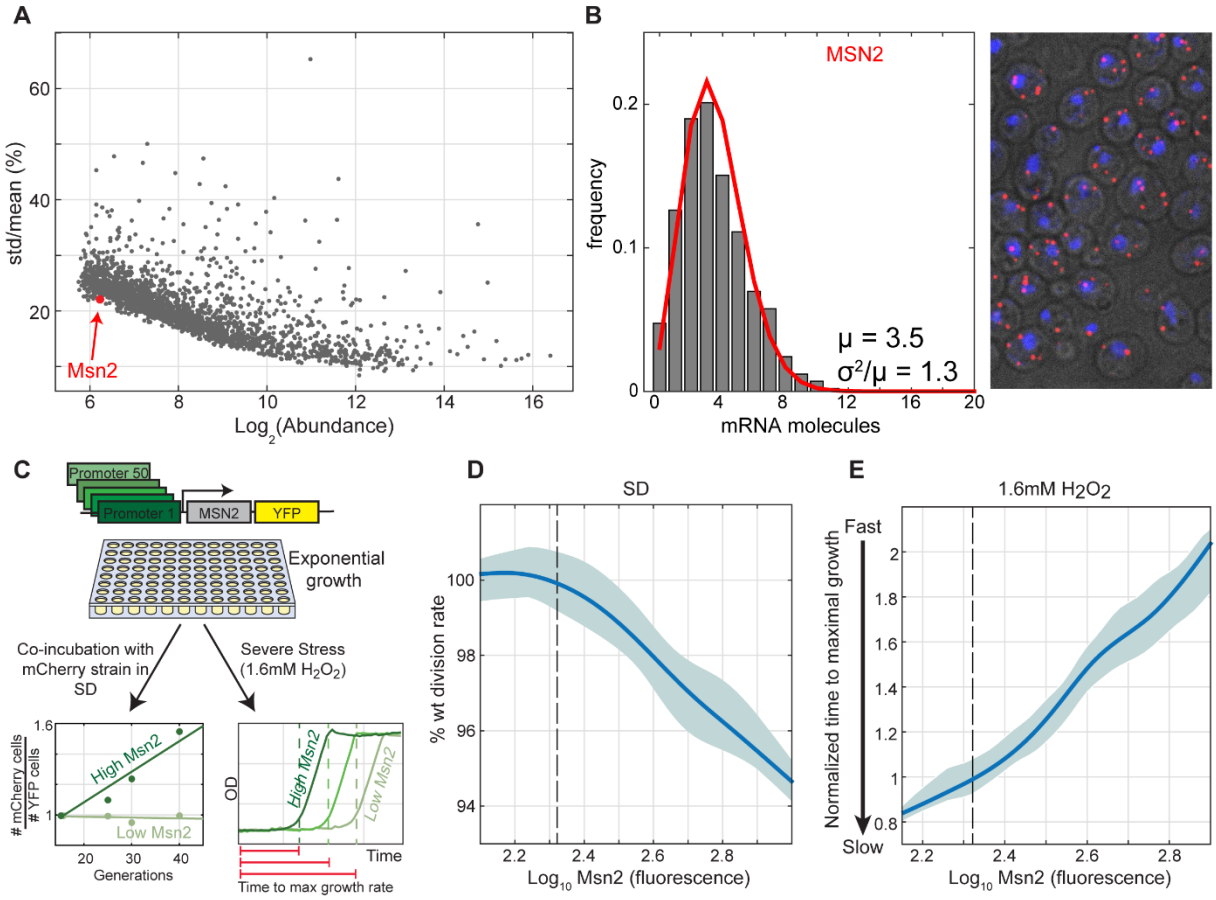

Figure 1. ***Tuning of Msn2 expression in rapidly growing cells:***

(**A**) *Cell-to-cell variability of Msn2-GFP is the lowest of all equally abundant proteins:* Shown is the noise vs. abundance data of ~2,500 GFP-fused proteins, data from: Newman et al.[22] Msn2 is shown as a red dot. Msn4-GFP was not detected.

(**B**) *Low-noise (Poisson) distribution of MSN2 in individual cells: MSN2* expression levels were measured using smFISH. (Left) *MSN2* mRNA counts distribution, quantified in >650 single cells. Red line represents Poisson fit to the data. (Right) Fixed cells labeled with *MSN2* mRNA in red and DAPI staining in blue, in a maximal z-projection image.

(**C-E**) *Msn2 expression increase stress protection by compensate growth rate in the absence of stress*: A library of 50 strains expressing Msn2-YFP under different synthetic promoters (from Keren et al.[33]), defining a range of expression values was generated (**C,** Methods). This library was used to define the effect of Msn2 expression level on growth rate and stress protection. Growth rates were measured using a sensitive competition assay, and is shown in (**D).** Stress protection was measured by subjecting exponentially growing cells to H_2O_2 (1.6mM), and identifying the time at which growth was first detected by continuous OD measurements (**E).** Shown are the median of all strains and repeats in solid line and 25-75 percentiles in the shaded areas. Dashed lines indicate wild-type Msn2 level.

The tradeoff between rapid growth and stress preparation depends on the contribution of these two parameters to the overall population fitness, as defined by the evolutionary history. This relative contribution, in turn, depends on growth conditions. For example, when cells encounter optimal growth conditions, maximizing division rate dominates, but when nutrients become limiting, or respiration is triggered, the importance of stress protection increases. Consistent with this, wild-type cells were better protected against H_2O_2 exposure at higher cell densities, as they approached stationary phase, resuming growth faster after stress induction (Figure 2A). We therefore expected Msn2 expression to increase with cell density. This, however, was not the case. Although Msn2 contributed to stress protection at all densities, its expression remained constant, throughout the growth curve (Figure 2B).

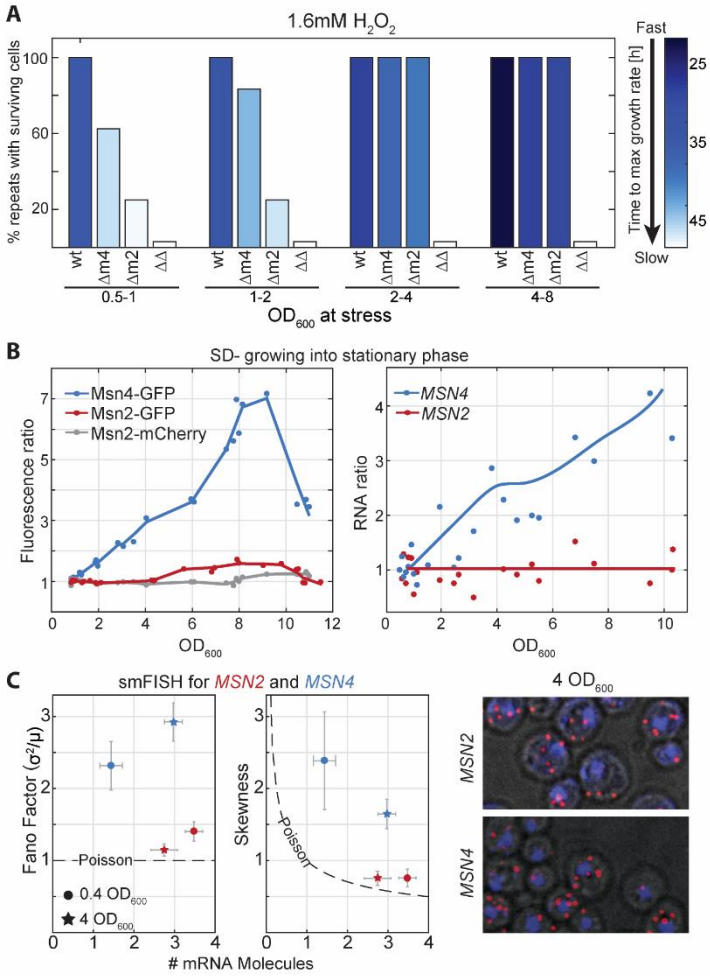

Figure. 2. ***Msn4 expression and its contribution to stress preparation increases as cells exit exponential growth:***

(**A**) *The contribution of Msn2 and Msn4 to stress preparation changes along the growth curve:* Cells at different stages along the growth curve (ODs) were diluted into media containing 1.6mM H_2O_2 and were followed by continuous OD measurements to define the time at which growth was first detected. Shown is the percent of repeats with surviving cells of each strain in different cell densities, and the time to resume growth (color code).

(**B**) *Msn4 expression increases along the growth curve in protein and transcript levels, while Msn2 expression remains stable:* expression was measured using fluorescent protein fusion (**B**, left), and transcription profiles (**B,** right).

(**C**). *MSN4 expression is noisy while MSN2 expression follows the Poissonian variance:* mRNA molecules of *MSN2,4* were counted in > 4000 single cells with smFISH in exponentially growing cells (circles) and at OD_{600} = 4 (stars). Shown are the mean number of molecules at the x-axis and the Fano-factor (Left) and skewness (Middle) of the mRNA distribution at the y-axis. Dashed line represents the Poisson distribution parameters. (Right) smFISH imaging examples.

Msn4 expression is environmental-sensitive and high noise:

Msn4, the Msn2 duplicate, is also a stress genes activator[13,26]. Msn4-GFP was undetectable in reported measurements[34,35], suggesting that its expression level is low during rapid growth. We reasoned that Msn4 expression increases along the growth curve to promote stress protection. This was indeed the case: Msn4 expression increased with cell density, both at the transcript and the protein levels (Figure 2B,C). This higher expression was accompanied by increased contribution to stress protection, as was measured by introducing H_2O_2 to strains deleted of msn2 in different cell densities (Figure 2A). Consistent with the control-noise tradeoff described above, this dynamic regulation of *MSN4* was accompanied by a high level of expression noise, which significantly exceeded the Poissonian variance (Figure 2C).

Msn2 and Msn4 co-localize to the nucleus with the same dynamics in individual cells:

Msn4 could collaborate with Msn2 in promoting stress protection by regulating the same set of genes or by inducing a distinct set of targets. Similarly, it could respond to the same, or to different sets of post-translational factors. Since activation of Msn2,4 culminates in nuclear localization[36,37], we first followed the nuclear translocation dynamics of fluorescent-tagged Msn2 and Msn4 (Figure 3A). In response to stress, the two factors translocated to the nucleus within minutes, showing precisely the same kinetics within individual cells (Figure 3B,C- yellow shade, S2). Similarly, during the stochastic pulsing following stress[37,38], translocation of the two factors was tightly synchronized within individual cells, but not between different cells (Figure 3C-pink shade). Deletion of one factor did not affect the dynamics of its duplicate (Figure S3).

Msn2 and Msn4 induce the same set of target genes:

Next, we examined for differences in Msn2,4 target genes. In rapidly growing cells, deletion of *MSN2* strongly reduced stress gene induction, while deletion of *MSN4* had little, if any, effect (Figure 3D, S5). Swapping the *MSN2,4* promoters completely reversed the target induction capacity of these factors (Figure 3E, S6B). The identity of the targets remained the same: Msn4 driven by the *MSN2* promoter induced precisely the same targets normally induced by Msn2. When tested in conditions where both factors are expressed to equivalent amounts, the two factors induced the same set of genes (Figure 3F, S6). Since a previous study[39] which followed stress gene induction using fluorescence reporters, indicated some differences in individual targets dependence on Msn2,4, we examined specifically the genes reported to be differently regulated. However, none of these genes showed any difference in their Msn2,4 dependency in any of the 6 conditions for which we performed tight time-course measurements (Figure S7). To further corroborate these results, we also measured the genome-wide binding profiles of the two factors, using the sensitive ChEC-seq method[40]. The two factors bound to the precise same promoters, occupied the precise same positions within individual promoters, and showed an identical preference for their common (known) DNA binding motif (Figure S8). This identity of Msn2,4 targets is consistent with the high conservation of their DNA binding domains (Figure 3G), and identity of their *in-vitro* DNA binding preferences[41] (Figure S9). We conclude that Msn2,4 proteins are co-regulated by the same signals and at the same kinetics, and activate the same set of target genes with the same kinetics, essentially functioning as one TF.

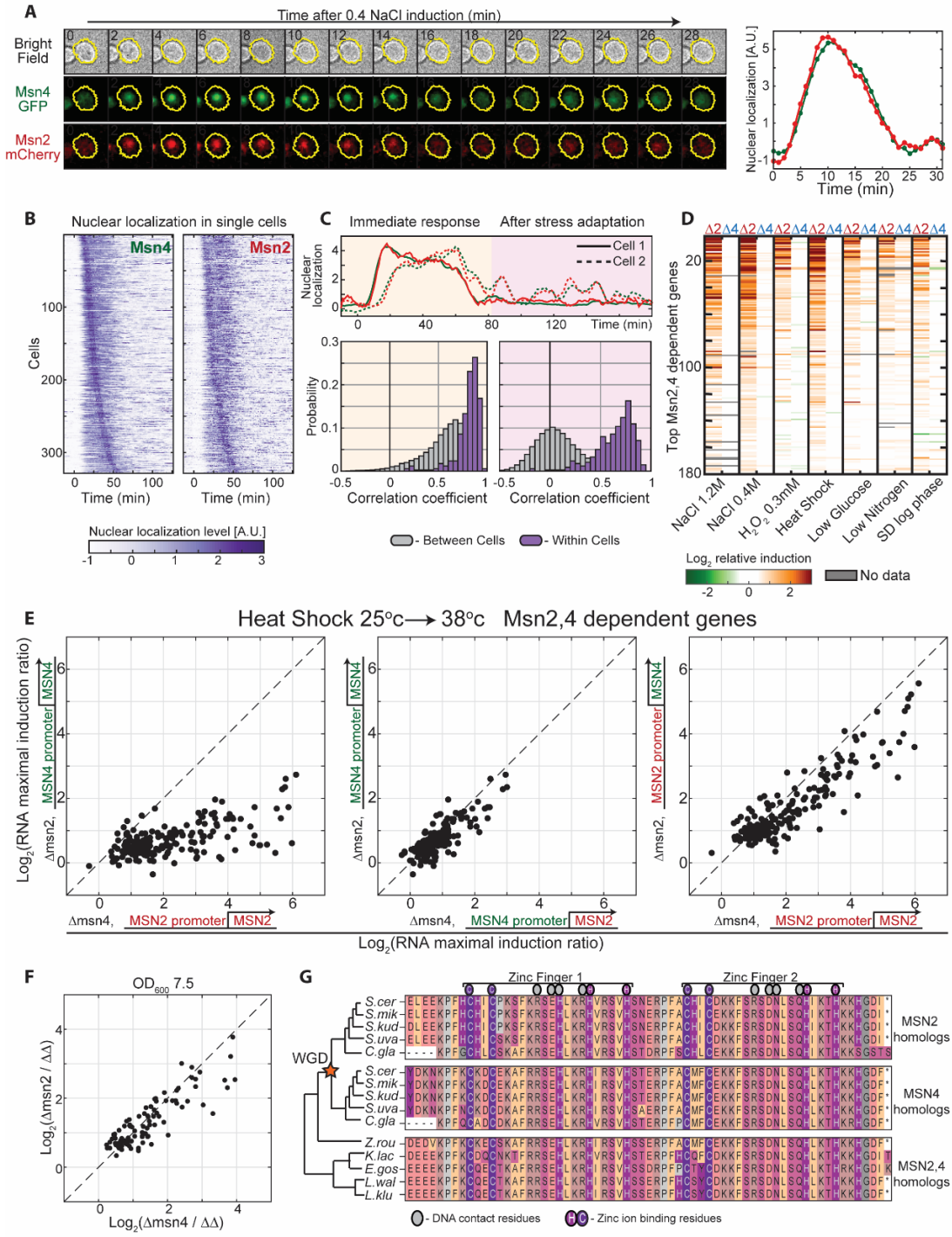

Figure. 3. Redundancy in Msn2 and Msn4 activity*:*

(A-C) Single cells expressing Msn4-GFP and Msn2-mChrerry were visualized using microfluidics-coupled live microscopy. Both proteins were readily visualized when cells were first cultured at intermediate or high OD (as MSN4 is undetectable in low ODs when cells grow exponentially). Cells

were tracked as they were exposed to 0.4M, 1.2M or 1.4M NaCl. Cells were segmented and the nuclear localization of both proteins was quantified. A representative cell in time in 3 channels and a quantification of nuclear localization levels of Msn2 (red) and Msn4 (green) are shown in **A**. Temporal traces of 328 single cells in 1.2M NaCl, ordered in both columns by the time of Msn4-GFP nuclear localization are shown in **B** (0.4M and 1.4M NaCl in Figure S2). Correlations between the individual traces of Msn2 vs. Msn4 nuclear localization levels in single cells were calculated. Distributions of the correlation coefficients within the same (purple) or in different (gray) cells are shown in **C**, separately comparing the immediate response (left) and the longer-time dynamics (right).

(**D**) *Stress response in rapidly growing cells depends on Msn2, but not Msn4:* Exponentially growing cells were exposed to the indicated stresses. Genome-wide transcription profiles were measured at 3' time resolution following stress induction, for the first 60 minutes, and 10' for the next 30 minutes. The stress response of each gene was summarized by its integrated (log2) change over the time course. The experiment was repeated in wild-type cells, in single deleted cells (Δmsn2, Δmsn4) and double deleted cells (Δmsn2Δmsn4). Shown are the differences between gene induction of the wild-type vs. the single deletion strains (Δ*msn2* or Δmsn4 at the left/right column, respectively). 180 genes are shown, selected and ordered by the fold-change difference in their response in the wild-type vs. single deletion cells, these genes contain stress induced modules defined by other studies (Figure S4).

(E) *Msn2 and Msn4 induce the same set of target genes:* during exponential growth, when Msn2 expression is higher than Msn4, deletion of Msn2 results in a significantly stronger effect on stress gene expression **(**left**)**, but this effect was fully reversed by swapping the Msn2 and Msn4 promoters **(**middle and right**)**. Each dot represents a target gene and its induction ratio between the indicated strain and the double msn2 and msn4 deletion strain.

(**F**) In high OD (7.5), when both factors are expressed, stress genes are induced equally. Each dot is an induced target gene.

(**G**) *Msn2 and Msn4 bind DNA through a highly conserved DNA binding domain:* Alignment of Msn2 and Msn4 DNA binding domains and their homologues in 10 species of the *Ascomycota* phylum that diverged before or after WGD (star). Colors indicate amino acid residue types.

Differential design of the Msn2,4 promoters explains the differences in their expression flexibility and noise:

Msn2 expression is stable along the growth curve, while Msn4 is strongly induced. To examine whether this differential regulation in expression is specific to these conditions, or is rather a more general property of the two genes, we surveyed a dataset composed of thousands of transcription profiles[12,42,43]. Expression of Msn2 showed little variability under all conditions tested, while Msn4 was highly variable (Figure 4A,B). Expression of Msn2 and Msn4 therefore conforms to the general tradeoff between expression noise and regulatory control: Msn2 is stable across conditions and shows low cell to cell variability (noise), while Msn4 expression readily responds to environmental signals and is noisy.

Previous studies have defined promoter designs that encode for flexible and noisy, or stable and low-noise expression[44–46]. Flexible promoters tend to contain a TATA box and bind nucleosomes immediately upstream to their Transcription Start Site (TSS), while stable promoters lack a TATA box and display a Nucleosome Free Region (NFR) upstream of their TSS. Consistent with their differential flexibility, we find that the *MSN4* promoter contains a TATA box, binds nucleosomes around its TSS and contains a large number of TF binding sites. By contrast, the *MSN2* promoter does not contain a TATA box, displays an NFR immediately upstream of the TSS and is largely devoid of TF binding sites (Figure 4C; data from[47–49]).

When aligned by their coding frames, the nucleosome patterns along the upstream regions of *MSN2* and *MSN4* promoters are highly similar. However, the location of the TSS is different: in *MSN4*, the TSS is positioned ~105 bp away in a region that is nucleosome occupied, while in *MSN2* the TSS is significantly further upstream and located on the border of an NFR. The resulting 5'UTR of *MSN2* is exceptionally long (~430 bp of length, found in only 2% of *S. cerevisiae* genes). Deleting this NFR region from the *MSN4* promoter, practically eliminated Msn4 induction along the growth curve. By contrast, deleting regions close to the ORF had little, if any further effect (Figure S10). Furthermore, replacing this region in the *MSN4* promoter by the corresponding region from *MSN2* promoter, including its new TSS and NFR, increased *MSN4* expression and reduced its noise (Figure S11). Therefore, this promoter region accounts for the differential expression characteristics of *MSN2* and *MSN4*.

A shift in the TSS following WGD event modified MSN2 promoter design:

Msn2,4 were duplicated in the whole-genome duplication (WGD) event, ~100 million years ago[11], and were retained in all WGD species tracing to this event. To examine whether the differential promoter structure of *MSN2,4* is conserved in other WGD species, we used available 5' RNA data[50] and further profiled TSS positioning in these species. The TSS positions of the *MSN2* and *MSN4* homologues were conserved in all post-WGD species (Figure 4D). Sequence analysis further indicated that the TATA box was conserved in all *MSN4* homologues but absent from all *MSN2* homologues (Figure 4D). We next profiled 5' RNA in two non-WGD species. The transcript of the single *MSN* homologue has a short 5'UTR, similar to that of *MSN4*. This pattern of conservation is consistent with a scenario in which the stable *MSN2* promoter evolved from an ancestral flexible promoter through a shift in the TSS to a distant, TATA-lacking position, at the boundary of a nearby NFR.

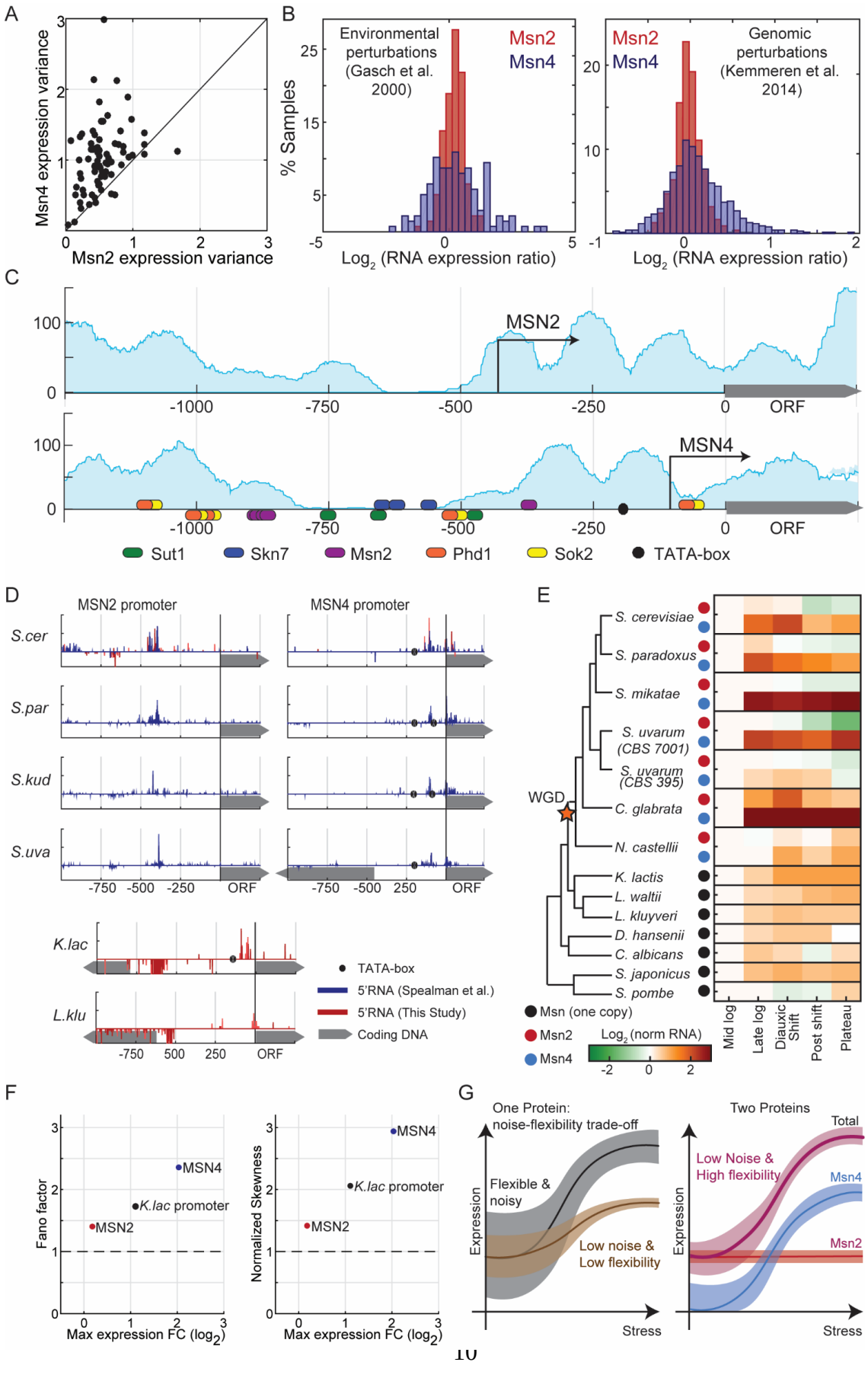
A
Msn4 expression variance
Msn2 expression variance
B
Environmental perturbations (Gasch et al. 2000)
Msn2
Msn4
% Samples
Log2 (RNA expression ratio)
Genomic perturbations (Kemmeren et al. 2014)
C
MSN2
MSN4
ORF
Sut1
Skn7
Msn2
Phd1
Sok2
TATA-box
D
MSN2 promoter
MSN4 promoter
S.cer
S.par
S.kud
S.uva
K.lac
L.klu
TATA-box
5'RNA (Spealman et al.)
5'RNA (This Study)
Coding DNA
E
S. cerevisiae
S. paradoxus
S. mikatae
S. uvarum (CBS 7001)
S. uvarum (CBS 395)
C. glabrata
N. castellii
K. lactis
L. waltii
L. kluyveri
D. hansenii
C. albicans
S. japonicus
S. pombe
WGD
Msn (one copy)
Msn2
Msn4
Log2 (norm RNA)
Mid log
Late log
Diauxic Shift
Post shift
Plateau
F
Fano factor
Max expression FC (log2)
MSN4
K.lac promoter
MSN2
Normalized Skewness
G
One Protein: noise-flexibility trade-off
Flexible & noisy
Low noise & Low flexibility
Expression
Stress
Two Proteins
Total
Low Noise & High flexibility
Msn4
Msn2

Figure. 4. Msn2 shifted its TSS and gained a stable expression pattern in species that diverged from *S. cerevisiae* following WGD.

(**A-B**) Three data types were considered. First, we downloaded >230 mRNA expression datasets available in SPELL[42], and compared the variance of *MSN2* and *MSN4* expression in each dataset with more than 20 samples (A, each dataset is a dot). Second, we compared the distribution of *MSN2* and *MSN4* expression levels in two large datasets, representing multiple stress conditions[12] (B, left) or gene deletions[43](B, right).

(**C**) *The MSN2 promoter displays properties of the stable, low noise type, while MSN4 promoter conforms to the flexible noisy type:* The pattern of nucleosome occupancy along the two promoters as defined by Weiner et al.[47], is shown in blue shade. Arrows represent TSS positions, as defined by Park et al.[48]. Ellipses denote TFs binding sites as defined by MacIsaac et al.[49] TATA-box (black circles) is defined as: TATA[AT]A[AT].

(**D**) *MSN2 promoter displays an uncharacteristically long 5'UTR that is conserved in all species that diverged after the WGD event:* Shown are the promoter maps of *MSN2,4* homologues in the indicated species. mRNA 5' end mapping data from Spealman et al.[50] is shown in blue, mRNA 5' end from this study in red. TATA-box is defined as in C.

(**E**) *MSN2 homologues are stably expressed along the growth curve, while MSN4 homologues show the flexible expression of the single MSN* homologues *found in species that diverged from S. cerevisiae prior to the WGD event:* Shown are expression levels of the Msn2,4 homologues in all indicated species, in 5 time-points along the growth curve. Data from Thompson et al.[51]

(**F**) *Expression of the K.lactis Msn2,4 homologue shows intermediate flexibility and noise.* On the x axis the maximal fold change expression of Msn2, Msn4 and the *K.lactis* homologue (data from: Thompson et al.[51]). Y axes show attributes of the expression distribution measured by smFISH, in Msn2, Msn4 and Msn2 in *S.cerevisiae* driven by the promoter of the *K.lactis* homologue. Shown are the fano factor (left) and the skewness of the distribution normalized to the skewness of a Poisson distribution with the same mean as the data (right).

(**G**) *Model*: duplication of *Msn2,4 resolved conflict between environmental responsiveness and noise:* Single genes whose expression is sensitive to environmental conditions but will suffer from high noise in non-stressed conditions, limiting the ability to precisely tune intermediate expression levels while maintaining environmental-responsive expression. Gene duplication can resolve this conflict. See text for details.

MSN4 accentuated the environmental-responsive but noisy expression of the non-WGD homologue, while MSN2 gained a stable, low-noise expression:

To examine whether the differential expression flexibility of *MSN2,4* is also conserved in the other post-WGD species, we used available expression data[51] of 13 yeast species along their growth curves. In all post-WGD species, *MSN4* expression increased along the growth curve, while *MSN2* expression remained stable (Figure 4E). The same dataset also profiled non-WGD species, allowing us to also examine the expression of the *MSN* single homologue in these species. The single *MSN* homologue in the non-WGD species showed a moderate induction along the growth curve, with dynamic range that was larger than that of *MSN2*, but lower than that of *MSN4* (Figure 4E). To examine whether this intermediate regulation is also reflected in the expression noise, we introduced the *MSN* promoter from *K. lactis*, a non-WGD species into *S. cerevisiae*, upstream of the *MSN2* ORF, and measured expression noise using smFISH. As predicted, this promoter showed an intermediate noise level that was higher than *MSN4* but lower than *MSN2* (Figure 4F). In fact, when plotted on the noise-control curve, the three promoters all fell on the same line, consistent with same-proportion change in noise and dynamic range of regulated expression. Therefore, our analysis suggests that *MSN2* gained its stable, low-noise expression following the duplication event, likely by shifting its TSS, while *MSN4* accentuated the regulated expression of the ancestral factor, likely through the acquisition of new binding sites for transcription factors, increasing its dynamic range and expression noise.

Discussion:

Taken together, our study defined a novel role for the Msn2,4 duplication. We were initially surprised to find that these two duplicates regulate the same set of target genes, translocate to the nucleus with the precise same dynamics, and are expressed in an overlapping set of conditions. What limits their replacement, in at least some species, by a single factor of a more refined transcriptional control? Our data shows that Msn2,4 function as one unit whose expression is both environmentally-responsive and low-noise (Figure 4G), thereby resolving an inherent conflict that limits the tuning of individual gene expression. Msn2 provides the low-noise basal expression, whereas Msn4 is induced when additional amounts are needed.

What could be the evolutionary force promoting this new evolution? Since the MSN duplication traces to the WGD event, it is tempting to propose that its new expression characteristics were driven by the shift in metabolism: Rapidly growing non-WGD species respire, while WGD species ferment. Following this metabolic change, genes needed in respiring cells may shift from being constitutively expressed, to being Msn-dependent, as was indeed reported[52]. We propose that changes in the identity of Msn2,4-dependent genes accentuated its phenotypic effects on growth and drove selection for increased precision of Msn2 expression.

Gene duplications is a major source of evolutionary innovation[4,5] that greatly contributes to the expansion of transcription networks[2,3]. A surprisingly large fraction of TF duplicates, however, retained a conserved DNA binding domain and bind to the same DNA motif (Figure S12), suggesting limited divergence in regulatory targets. These, and other duplicates of apparent redundant function[53–56] do not comply with the accepted models of neo- or sub-functionalization explaining duplicate advantage. Our study suggests a third model whereby duplicates with redundant biochemical properties realize dynamic properties that are not possible, or difficult to achieve using a single factor. In the case of Msn2,4, duplication resolved a conflict between regulatory control and noise. In other cases, interactions between the factors may define a circuit with dynamic properties not implementable

by a single gene[55,57,58]. Further studies will define the relative contribution of such circuit-forming mechanisms in explaining the retention of TFs or other duplicates.

References:

1. Charoensawan, V., Wilson, D. & Teichmann, S. A. Genomic repertoires of DNA-binding transcription factors across the tree of life. *Nucleic Acids Res.* **38**, 7364–7377 (2010).
2. Weirauch, M. T. & Hughes, T. R. A Catalogue of Eukaryotic Transcription Factor Types, Their Evolutionary Origin, and Species Distribution. in *Sub-cellular biochemistry* **52**, 25–73 (2011).
3. Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* **172**, 650–665 (2018).
4. Conant, G. C. & Wolfe, K. H. Turning a hobby into a job: How duplicated genes find new functions. *Nat. Rev. Genet.* **9**, 938–950 (2008).
5. Soskine, M. & Tawfik, D. S. Mutational effects and the evolution of new protein functions. *Nat. Rev. Genet.* **11**, 572–582 (2010).
6. Hittinger, C. T. & Carroll, S. B. Gene duplication and the adaptive evolution of a classic genetic switch. *Nature* **449**, 677–681 (2007).
7. Des Marais, D. L. & Rausher, M. D. Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature* **454**, 762–765 (2008).
8. Voordeckers, K., Pougach, K. & Verstrepen, K. J. How do regulatory networks evolve and expand throughout evolution? *Curr. Opin. Biotechnol.* **34**, 180–188 (2015).
9. Pérez, J. C. *et al.* How duplicated transcription regulators can diversify to govern the expression of nonoverlapping sets of genes. *Genes Dev.* **28**, 1272–7 (2014).
10. Baker, C. R., Hanson-Smith, V. & Johnson, A. D. Following gene duplication, paralog interference constrains transcriptional circuit evolution. *Science* **342**, 104–8 (2013).
11. Wolfe, K. H. & Shields, D. C. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**, 708–713 (1997).
12. Gasch, A. P. *et al.* Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes. *Mol. Biol. Cell* **11**, 4241–4257 (2000).
13. Schmitt, A. P. & McEntee, K. Msn2p, a zinc finger DNA-binding protein, is the transcriptional activator of the multistress response in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 5777–82 (1996).
14. Estruch, F. Stress-controlled transcription factors, stress-induced genes and stress tolerance in budding yeast. *FEMS Microbiol. Rev.* **24**, 469–486 (2000).
15. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183–6 (2002).
16. Raser, J. M. & O'Shea, E. K. Noise in gene expression: origins, consequences, and control. *Science* **309**, 2010–3 (2005).
17. Lehner, B. Selection to minimise noise in living systems and its implications for the evolution of gene expression. *Mol. Syst. Biol.* **4**, (2008).
18. Metzger, B. P. H., Yuan, D. C., Gruber, J. D., Duveau, F. & Wittkopp, P. J. Selection on noise constrains variation in a eukaryotic promoter. *Nature* **521**, 344–347 (2015).
19. Eldar, A. & Elowitz, M. B. Functional roles for noise in genetic circuits. *Nature* **467**, 167–173 (2010).

20. Raj, A. & van Oudenaarden, A. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell* **135**, 216–226 (2008).

21. Yaakov, G., Lerner, D., Bentele, K., Steinberger, J. & Barkai, N. Coupling phenotypic persistence to DNA damage increases genetic diversity in severe stress. *Nat. Ecol. Evol.* **1**, 0016 (2017).

22. Newman, J. R. S. *et al.* Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* **441**, 840–846 (2006).

23. Lehner, B. Conflict between Noise and Plasticity in Yeast. *PLoS Genet.* **6**, e1001185 (2010).

24. Hornung, G. *et al.* Noise-mean relationship in mutated promoters. *Genome Res.* **22**, 2409–17 (2012).

25. Choi, J. K. & Kim, Y.-J. Intrinsic variability of gene expression encoded in nucleosome positioning sequences. *Nat. Genet.* **41**, 498–503 (2009).

26. Martínez-Pastor, M. T. *et al.* The Saccharomyces cerevisiae zinc finger proteins Msn2p and Msn4p are required for transcriptional induction through the stress response element (STRE). *EMBO J.* **15**, 2227–2235 (1996).

27. Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877–879 (2008).

28. Munsky, B., Neuert, G. & van Oudenaarden, A. Using gene expression noise to understand gene regulation. *Science* **336**, 183–7 (2012).

29. Bar-Even, A. *et al.* Noise in protein expression scales with natural protein abundance. *Nat. Genet.* **38**, 636–643 (2006).

30. Fraser, H. B., Hirsh, A. E., Giaever, G., Kumm, J. & Eisen, M. B. Noise Minimization in Eukaryotic Gene Expression. *PLoS Biol.* **2**, e137 (2004).

31. Wang, Z. & Zhang, J. Impact of gene expression noise on organismal fitness and the efficacy of natural selection. *Proc. Natl. Acad. Sci. U. S. A.* **108**, E67-76 (2011).

32. Richard, M. & Yvert, G. How does evolution tune biological noise? *Front. Genet.* **5**, 374 (2014).

33. Keren, L. *et al.* Massively Parallel Interrogation of the Effects of Gene Expression Levels on Fitness. *Cell* **166**, 1282–1294.e18 (2016).

34. Huh, W.-K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).

35. Breker, M., Gymrek, M., Moldavski, O. & Schuldiner, M. LoQAtE—Localization and Quantitation ATlas of the yeast proteomE. A new tool for multiparametric dissection of single-protein behavior in response to biological perturbations in yeast. *Nucleic Acids Res.* **42**, D726–D730 (2014).

36. Petrenko, N., Chereji, R. V., McClean, M. N., Morozov, A. V. & Broach, J. R. Noise and interlocking signaling pathways promote distinct transcription factor dynamics in response to different stresses. *Mol. Biol. Cell* **24**, 2045–2057 (2013).

37. Hao, N., Budnik, B. A., Gunawardena, J. & O'Shea, E. K. Tunable signal processing through modular control of transcription factor translocation. *Science* **339**, 460–4 (2013).

38. Lin, Y., Sohn, C. H., Dalal, C. K., Cai, L. & Elowitz, M. B. Combinatorial gene regulation by

modulation of relative pulse timing. *Nature* **527**, 54–58 (2015).

39. AkhavanAghdam, Z., Sinha, J., Tabbaa, O. P. & Hao, N. Dynamic control of gene regulatory logic by seemingly redundant transcription factors. *Elife* **5**, (2016).

40. Zentner, G. E., Kasinathan, S., Xin, B., Rohs, R. & Henikoff, S. ChEC-seq kinetics discriminates transcription factor binding sites by DNA sequence and shape in vivo. *Nat. Commun.* **6**, 8733 (2015).

41. Siggers, T., Reddy, J., Barron, B. & Bulyk, M. L. Diversification of Transcription Factor Paralogs via Noncanonical Modularity in C2H2 Zinc Finger DNA Binding. *Mol. Cell* **55**, 640–648 (2014).

42. Hibbs, M. A. *et al.* Exploring the functional landscape of gene expression: directed search of large microarray compendia. *Bioinformatics* **23**, 2692–2699 (2007).

43. Kemmeren, P. *et al.* Large-Scale Genetic Perturbations Reveal Regulatory Networks and an Abundance of Gene-Specific Repressors. *Cell* **157**, 740–752 (2014).

44. Field, Y. *et al.* Distinct Modes of Regulation by Chromatin Encoded through Nucleosome Positioning Signals. *PLoS Comput. Biol.* **4**, e1000216 (2008).

45. Tirosh, I. & Barkai, N. Two strategies for gene regulation by promoter nucleosomes. *Genome Res.* **18**, 1084–1091 (2008).

46. Nicolas, D., Phillips, N. E. & Naef, F. What shapes eukaryotic transcriptional bursting? *Mol. Biosyst.* **13**, 1280–1290 (2017).

47. Weiner, A. *et al.* High-Resolution Chromatin Dynamics during a Yeast Stress Response. *Mol. Cell* **58**, 371–386 (2015).

48. Park, D., Morris, A. R., Battenhouse, A. & Iyer, V. R. Simultaneous mapping of transcript ends at single-nucleotide resolution and identification of widespread promoter-associated non-coding RNA governed by TATA elements. *Nucleic Acids Res.* **42**, 3736–3749 (2014).

49. MacIsaac, K. D. *et al.* An improved map of conserved regulatory sites for Saccharomyces cerevisiae. *BMC Bioinformatics* **7**, 113 (2006).

50. Spealman, P. *et al.* Conserved non-AUG uORFs revealed by a novel regression analysis of ribosome profiling data. *Genome Res.* **28**, 214–222 (2018).

51. Thompson, D. A. *et al.* Evolutionary principles of modular gene regulation in yeasts. **2**, 603 (2013).

52. Tirosh, I., Wong, K. H., Barkai, N. & Struhl, K. Extensive divergence of yeast stress responses through transitions between induced and constitutive activation. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 16693–8 (2011).

53. DeLuna, A. *et al.* Exposing the fitness contribution of duplicated genes. *Nat. Genet.* **40**, 676–681 (2008).

54. Dean, E. J., Davis, J. C., Davis, R. W. & Petrov, D. A. Pervasive and Persistent Redundancy among Duplicated Genes in Yeast. *PLoS Genet.* **4**, e1000113 (2008).

55. Kafri, R., Springer, M. & Pilpel, Y. Genetic Redundancy: New Tricks for Old Genes. *Cell* **136**, 389–392 (2009).

56. Diss, G. *et al.* Gene duplication can impart fragility, not robustness, in the yeast protein interaction network. *Science* **355**, 630–634 (2017).

57. Alon, U. Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* **8**, 450–461

(2007).

58. Teichmann, S. A. & Babu, M. M. Gene regulatory network growth by duplication. *Nat. Genet.* **36**, 492–496 (2004).

59. Rahman, S. & Zenklusen, D. Single-molecule resolution fluorescent in situ hybridization (smFISH) in the yeast S. cerevisiae. *Methods Mol. Biol.* **1042**, 33–46 (2013).

60. Avraham, N., Soifer, I., Carmi, M. & Barkai, N. Increasing population growth by asymmetric segregation of a limiting resource during cell division. *Mol. Syst. Biol.* **9**, (2013).

61. Kafri, M., Metzl-Raz, E., Jona, G. & Barkai, N. The Cost of Protein Production. *Cell Rep.* **14**, 22–31 (2016).

62. Orsi, G. A., Kasinathan, S., Zentner, G. E., Henikoff, S. & Ahmad, K. Mapping Regulatory Factors by Immunoprecipitation from Native Chromatin. in *Current Protocols in Molecular Biology* **110**, 21.31.1-21.31.25 (John Wiley & Sons, Inc., 2015).

63. Pelechano, V., Wei, W. & Steinmetz, L. M. Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature* **497**, 127–131 (2013).

64. Ihmels, J., Bergmann, S. & Barkai, N. Defining transcription modules using large-scale gene expression data. *Bioinformatics* **20**, 1993–2003 (2004).

65. De Boer, C. G. & Hughes, T. R. YeTFaSCo: A database of evaluated yeast transcription factor sequence specificities. *Nucleic Acids Res.* **40**, D169–D179 (2012).

66. Yofe, I. *et al.* One library to make them all: streamlining the creation of yeast libraries via a SWAp-Tag strategy. *Nat. Methods* **13**, 371–378 (2016).

Acknowledgments:

We thank members of our lab for fruitful discussions and comments on the MS. We thank Nir Friedman and his group for their help and fertile discussions, especially to Daphna Joseph-Strauss. We thank Yoav Breuer for his help with constructing and performing one of the smFISH experiment. We thank Alon Appleboim for his support and suggestions. This work was supported by the ISF, and the Minerva Center.

Supplementary Materials and methods for

Resolving noise-control conflict by gene duplication

Authors: Michal Chapal[1], Sefi Mintzer[1], Sagie Brodsky[1], Miri Carmi[1] and Naama Barkai[1]†

Affiliations:

[1]Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel.

†Correspondence to: naama.barkai@weizmann.ac.il

This file includes:

Materials and Methods
Figs. S1 to S12
Tables S1 to S3
References for supplementary Materials

Materials and Methods

Strains

All strains used in this study and their genotypes are listed in Table S1. All the strains were constructed by standard genetic methods and were validated by PCR and/or sequencing of the relevant DNA.

Single molecule Fluorescence In-Situ Hybridization

For each gene (MSN2 and MSN4), a set of 48 probes was generated as described in Raj et al.[27] The probes were designed by the online program Stellaris™ Probe Designer from Biosearch Technologies, and were ordered with a fluorescent dye CAL Fluor Red® 590. Probe sequences are listed in Tables S2-S3.

Cells were grown overnight in synthetic complete (SC) medium at 30° C and constant shaking. Then diluted to reach the wanted cell densities after ~12 hours. Cells were fixated, prepared and hybridized as described in Rahman et al. [59]

Images were acquired with 100x 1.4 oil UPLSAPO objective, using Olympus IX83 based Live-Imaging system equipped with CSU-W1 spinning disc (sCMOS digital Scientific Grade Camera 4.2 MPixel). For each sample 4-6 different positions were chosen. In each position three Z-stacks images were taken with step size of 200nm for a total of >6µm: Bright-field image- 488nm laser with 100mW, DAPI image- 405nm laser with 120mW and exposure time of 250ms , mRNA image- 561nm laser with 100mW and exposure time of 1000ms. Each z-plane image was of size 2048x2048 pixels.

Single molecule quantification- Cells were segmented using a modification of a custom MATLAB software[60]. In this modification, cell centers were defined manually using the bright field images and cells borders were found automatically. mRNA counts were then performed for each cell based on the custom made MATLAB software from Raj et al.[27]

Stress experiments for RNA-seq levels

In these experiments we used the wt strains, the single msn2 or msn4 deletion strains, and the double msn2,msn4 deletion strain. Some of the experiments were also done with the strains with swapped promoters: MSN2 ORF under MSN4 promoter with a deletion of msn4, and the opposite- MSN4 ORF under MSN2 promoter, with an msn2 deletion).

Growth Conditions - Cells were grown overnight in rich medium – YPD or synthetic complete (SC) medium at 30° C (unless otherwise noted), then cells were diluted and exponentially grew for 6-8 hours before introducing the stress:

Oxidative Stress - Cells were grown continuously in 30° C. H2O2 was added to a final concentration of 0.3mM.

Heat Shock - Cells were grown continuously in 25° C, then cell culture was moved to a new flask located inside a bath orbital shaker (Cat. WBT-450, MRC) preheated to 37° C. It took less than 90 seconds for the culture to reach 37° C.

Glucose limitation - Cells were grown in SC medium with 2% glucose (Sigma-Aldrich). Then cells were washed twice and resuspended in SC with 0.1% glucose. Samples were taken before the washes (2% glucose), after every wash and for the next 100 min.

Osmotic Shock - Cells were grown continuously in 30° C. 4M NaCl solution was added to the culture to a final concentration of 0.4 M or 1.2M.

Low nitrogen - Cells were grown in SC medium with 2% glucose (Sigma-Aldrich). Then cells were washed twice and resuspended in Nitrogen-depleted medium (0.67% Yeast Nitrogen Base without amino acids and ammonium sulfate (Bacto-YNB), 2% glucose, 0.05 mM ammonium sulfate, 20 mg/l Urcail, 20 mg/l Histidine, 100 mg/l Leucin, 20 mg/l Methionine).

Growth into stationary phase – Cells were grown in SC in 30° C without changing the media.

RNA sample collection, extraction and sequencing levels

Cells were grown overnight to stationary phase and then diluted in 100ml to reach OD600 of 0.2-0.4 after 6-8 hours in constant shaking. A sample for time point zero reference was taken, then we introduced a stress perturbation as described above. For growing into stationary phase experiment, a sample of 1ml was taken every 20/30 minutes. For all other conditions a sample of 1.5ml was collected every 3 minutes for the first hour, and every 10 minutes for additional half/one hour. Samples were immediately centrifuged for 40 seconds in 13,000 rpm. The supernatant was removed, pellets were frozen in liquid nitrogen and stored at −80° C until RNA preparation.

RNA was extracted using modified protocol of nucleospin® 96 RNA kit. Specifically, cell lysis was done in a 96 deep-well plate by adding 450µl of lysis buffer containing 1M sorbitol (Sigma-Aldrich), 100mM EDTA and 0.45µl lyticase (10IU/µl). The plate was incubated at 30° C for 30 minutes in order to break the cell wall, and then centrifuged for 10 minutes at 2500 rpm, and supernatant was removed. From this stage, extraction proceeded as in the protocol of nucleospin® 96 RNA kit, only substituting β-mercaptoethanol with DTT.

For all samples sequenced by Illumina HiSeq 2500, RNA libraries were created as follows: Fragmented, poly(A)-selected RNA extracts of ~200 bp size were reverse-transcribed to cDNA using barcoded poly(T) primers. cDNA was amplified and sequenced with Illumina HiSeq 2500 using a primer complementary to the opposite adaptor to the poly(A).

For all samples sequenced by Illumina NextSeq 500, RNA libraries were created as follows: poly(A) RNA was selected by reverse transcription with a barcoded poly(T) primer. The barcoded DNA-RNA hybrids were pooled and fragmented by a hyperactive variant of the Tn5 transposase. Tn5 was stripped off the DNA by treatment with SDS 0.2% followed by SPRI cleanup and the cDNA was amplified and sequenced with Illumina NextSeq 500.

Processing and analysis of RNA-seq data

We mapped 50bp reads of the RNA-seq of every sample to the *S.cereviciae* genome (R64 in SGD) using bowtie (parameters: –best -a -m 2 -strata -5 10). After alignment to the genome, samples that had less than 150,000 reads were discarded from the analysis in order to prevent an artificial enrichment for highly expressed genes. The expression at those time points was calculated as the mean between the two closest time points in the time course. For every sequence we normalized for PCR bias using the unique molecular identifier (UMI), scoring each position on the genome by the unique number of UMI's it had out of all possible UMI's. For each gene we summed all the reads aligned to 400bp upstream its 3' end to 200bp downstream, in order to get the total expression of that gene. Reads that were aligned non-uniquely were split between the aligned loci according to the ratio of all other uniquely mapped reads in these regions. Number of reads for each sample was normalized to 10^6.

Msn2-Promoter library preparation

We used 140 synthetic promoters from Keren et al.[33] , pooled together and transformed them to replace the native *MSN2* promoter, in a strain with Msn2 tagged with YFP and deleted of msn4. We collected ~200 colonies after the transformation, and measured YFP fluorescence with flow cytometer (BD LSRII system from BD Biosciences). We picked 50 strains that spanned the expression range and was highly similar between the different repeats.

Growth experiment in harsh stress

MSN2 promoter library strains – we grew the cells to stationary phase in SC media in 96-well plate under constant shaking and 30°c. Next, we diluted the cells with fresh SC media, in deep well plates with one glass bead in each well to generate proper shaking, so they will reach the wanted OD in the next morning. Then, right before stressing the cells, we took 150µl to measure ODs (using infinite200 reader, Tecan Inc.), and 100µl to the a flow cytometer to measure Msn2-YFP fluorescence.

Single and double deletion trains – we grew cells overnight to stationary phase in SC media and constant shaking, at 30°c. Next, we serial dilute the cells with fresh SC media in a 96-well plate to reach sequential different ODs in the next morning. Then, right before stressing the cells, we took 150µl to measure ODs (using infinite200 reader, Tecan Inc.)

Stress and growth measurements – we took 30µl of growing cells into plates with 120µl of H2O2. We inserted the plates into an automated handling robot (EVOware, Tecan Inc.) where cells were grown under constant shaking and 30°c. The robot was programed to take the plates out of the incubator every 45 minutes, vortex the plates, and measure the OD (using infinite200 reader, Tecan Inc.). Experiment lasted for ~70 hours.

Growth analysis- data from the growth measurements was parsed and processed. Time to exponential growth was calculated as the time of the maximal slope of the OD measurements in time. We calculated the median time and the std of the repeats.

Competition experiment in SC media

Cells were grown O.N to stationery phase in SC. Then diluted and grown ~8 hours in exponential growth. Each strain of the Msn2-promoter library strains were then co-incubated with wt-mCherry strain at 30˚C. Wt initial frequency was ~50%. Every ~8 hours cells were diluted with fresh SC media so they will grow exponentially at all times. In addition, a sample was taken to measure OD, and to a flow cytometer to measure frequencies of each population. Flow cytometry measurements and analysis were done using BD LSRII system (BD Biosciences). Flow cytometer was conducted with excitation at 488nm and emission at 525±25nm for GFP samples. For mCherry markers, excitation was conducted at 594nm and emission at 610±10nm. The number of generations was calculated from the dilution factor. % of wt division rate was calculated as previously described in kafri et al.[61]

Msn2,4 protein expression by flow cytometry

For this experiment, we used strain with Msn2 tagged with GFP, and a strain with Msn4 tagged with GFP and Msn2 tagged with mCherry. We grew the cells overnight in 5 mL SC media and 30°C to reach stationary phase, then to reach OD_{600} ~0.4 after ~8 hours. Then, we serial diluted the cells in 96-well plate by diluting each column to the next one with 1:1 ratio with SC media, ending up with 120µl in each well, and 1:2 ratio of cells in adjacent columns. After overnight incubation in 30°C, and constant shaking, we measured fluorescence using flow cytometer.

Flow cytometry measurements and analysis were done using BD LSRII system (BD Biosciences). Flow cytometer was conducted with excitation at 488nm and emission at 525±25nm for GFP samples. For mCherry markers, excitation was conducted at 594nm and emission at 610±10nm. The average number of cells analyzed was 50,000. For the samples with high OD, 100µl of DDW was added to the sample before reading it in the FACS.

We calculated cell density using the flow cytometer parameters and output. This measure was calculated as following: $\frac{N}{V}C$ where R = flow rate (µl /sec), T = total flow time (sec), V = R*T = total volume read (µl), N = number of cells read (cells), C = dilution fix constant (values are either 1 for no dilution with DDW, or 1.8333 for samples that were diluted with 100 µl DDW).

We filtered G1 cells by the width size measure FSC-W, which has a bimodal distribution that corresponds to cells in G1 (smaller) and cells in G2/M (bigger). Next, we filtered outliers by two area measures – FSC-A and SSC-A. We used linear regression to describe FSC-A with SSC-A, and removed cells that were far from the regression line. We then applied linear regression to describe SSC-A with FSC-A and removed outliers in a similar manner.

In order to eliminate the background fluorescence we used a linear regression model that predicts background fluorescence. The independent variables were size parameters (FSC-W and SSC-W) and the cell density of the population. The dependent variable was the background fluorescence (GFP/mCherry). Model was trained on BY4741 cells with no fluorescent markers, then used to predict background in the other strains. Predicted background was subtracted from observed fluorescence for each cell.

Time laps microscopy experiment

We used a strain with both Msn2-mCherry and GFP-Msn4 and a strain with GFP-Msn4 and a deletion of Msn2. We grew the cells overnight in SC media to reach stationary phase, then diluted them to reach the desired OD600 of ~7 after ~8 hours. When reaching the desired OD, cells were transferred to a microfluidics plate (catalog number: Y04C-02-5PK) for haploid yeast cells. We used ONIX CellAsic microfluidics system that allows changing the cells' media at a fast rate, in a predefined set time, while not interfering with the imaging process. During imaging, after ~15 minutes of normal media, a media with NaCl was added to the cells (0.4/1.2/1.4 M). Every minute, for each position on the plate, three images were taken – bright-field image, GFP image and mCherry image. Two positions were taken for each strain. Imaging under stress lasted for 4-8 hours.

We used a Zeiss AxioObserverZ1 inverted microscope equipped with Hamamatsu Flash4 sCMOS cameras. In every imaging instance, three images were taken: Bright-field image, GFP image using GFP filter with 20% intensity of HSP120 V lamp and with exposure time of 200 ms, and mCherry image using mPlum filter with same intensity and exposure as GFP. we used 2x2 binning resulting in 1024x1024 pixels of image size.

Processing microscopy images and estimate nuclear localization levels

Tracking and segmentation - All images were subsequently analyzed using custom MATLAB software that segments and tracks individual cells along the movie in each bright field image frame, as previously described[60]. Briefly, cell borders were detected automatically in the last frame. Then the program goes back to the beginning of the experiment frame-by-frame, and for each cell in the image uses the centroid coordinates of the cells from the previous frame. Each centroid is expended until the borders of the cell in the current frame if found. The program also outputs a score for the segmentation that was used to filter out cells with low quality segmentation.

Image processing - Median filter: we ran 3x3 median filter on all GFP and mCherry images. Background removal was done by running mean filter of 50x50 on each image, then subtracting the filtered image from the original one. Rare events of missing frames (mainly due to focus issues) were interpolated to be the mean of the two adjacent frames.

Calculating nuclear localization measure - Our method uses image filtering with a filter shaped like a nucleus with radius of 3 pixels. We run the filter on each cell GFP/mCherry track, and then find the maximal coordinate of the filtered cell image, defining it as the center of the hypothetical nucleus. Our measure is the average over the pixels in the hypothetical nucleus divided by the pixels average in the hypothetical cytoplasm. For normalization, we divide each cell's localization in time by the minimal value for this cell. As a result, this method in fact gives signal-to-noise ratio. In order to align the GFP and the mCherry tracks together, we used z-score normalization (subtracted the mean and divided by standard deviation for each cell).

Filtering bad cells- we filtered out "bad" cells in two rounds, once after running segmentation, and once after calculating the dynamical attributes. In the first round, we filtered cells that answered one or more of the following conditions: (1) Cells with area outside the range defined $Median \pm 3 \cdot MAD$ (MAD is the mean absolute deviation) over all cells, at least 10% of the time. (2) Cells with

segmentation score <8 at least 10% of the time. (3) Cells that did not appear from the beginning of the experiment. In the second round we removed cells with response amplitude below 1.1 or above 6.

MSN2,4 homologues expression in growth

Data was taken from Thompson et al.[51] For each yeast species in this experiment, 5 time-points were measured as a ratio to mid-log sample: lag, late-log, diauxic-shift, post diauxic shift, plateau. For each time point 3 repeats were made. We show the average of the repeats. We exclude the first time-point (lag) from the figure due to a large number of missing data and repeats.

5'mRNA sequencing

Cells were grown overnight to stationary phase in SC media in 30°C, and then diluted and exponentially grew for ~12 hours in constant shaking. Samples were fixed by mixing with cold (-80°C) methanol. RNA was Poly(A)-selected, reverse transcribed to cDNA and barcoded at the 5' end using Dynabeads® Oligo$(dT)_{25}$ magnetic beads. cDNA was pooled and fragmented by a hyperactive variant of the Tn5 transposase. Tn5 was stripped off the DNA by SDS 0.2% treatment followed by SPRI cleanup and the cDNA was amplified and sequenced with Illumina NextSeq 500. Number of reads for each sample was normalized, and genomic tracks were created from the sequence reads, representing the enrichment on each position of the genome.

CHeC-seq strains

We fused Msn2 or Msn4 to MNase (Amino Acids 83-231) using pGZ108 (pFA6a-3FLAG-MNase-kanMX6). This plasmid was a gift from Steven Henikoff (Addgene plasmid # 70231).

CHeC-seq experiment

Cells were grown overnight to stationary phase in SC media in 30°C, and then diluted and grew for ~15 hours in 30°C and constant shaking until they reached OD600 of ~4. Then. CHeC-seq was performed as described in Zentner et al.[40] with 30 seconds of activated Mnase, and changes in the ethanolic precipitation (1h in -80°C), and SPRI beads size selection (0.8x). Library preparation was performed as describe in Orsi et al.[62] , with converting the S-300 column cleanup following the phenol-chloroform step to ethanolic precipitation. Libaries was sequenced with Illumina NextSeq 500.

CHeC-seq analysis

Reads were mapped to the S.cereviciae genome (R64 in SGD) using bowtie2. The first nucleotide of every read was counted as binding signal. All samples had $>10^6$ reads. Each sample was normalized to 10^7 reads. Promoter length was defined as 700 base pairs upstream to the transcription start site or the distance to the upstream transcript (the shorter between these two). Transcription start and end sites were taken from Pelechano et al.[63] For the motif analysis, the average of the sum of signal of each 7-mer appearance (±10bp) in all of the promoter regions was calculated.

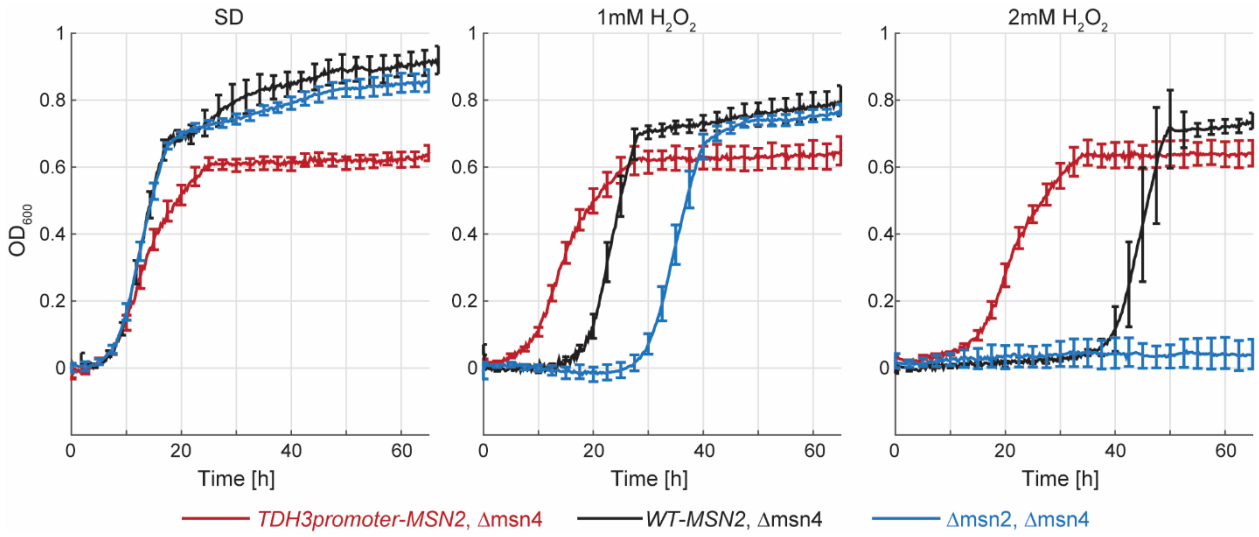

Fig. S1. *Growth curves in SD or H_2O_2 in Msn2 over-expression or deletion strains.* Cells were grown in the indicated condition under constant shaking and 30°c in 96-well plates in an automated handling robot (EVOware, Tecan Inc.). OD was measured automatically every ~30min for 65 hours using infinite200 reader.

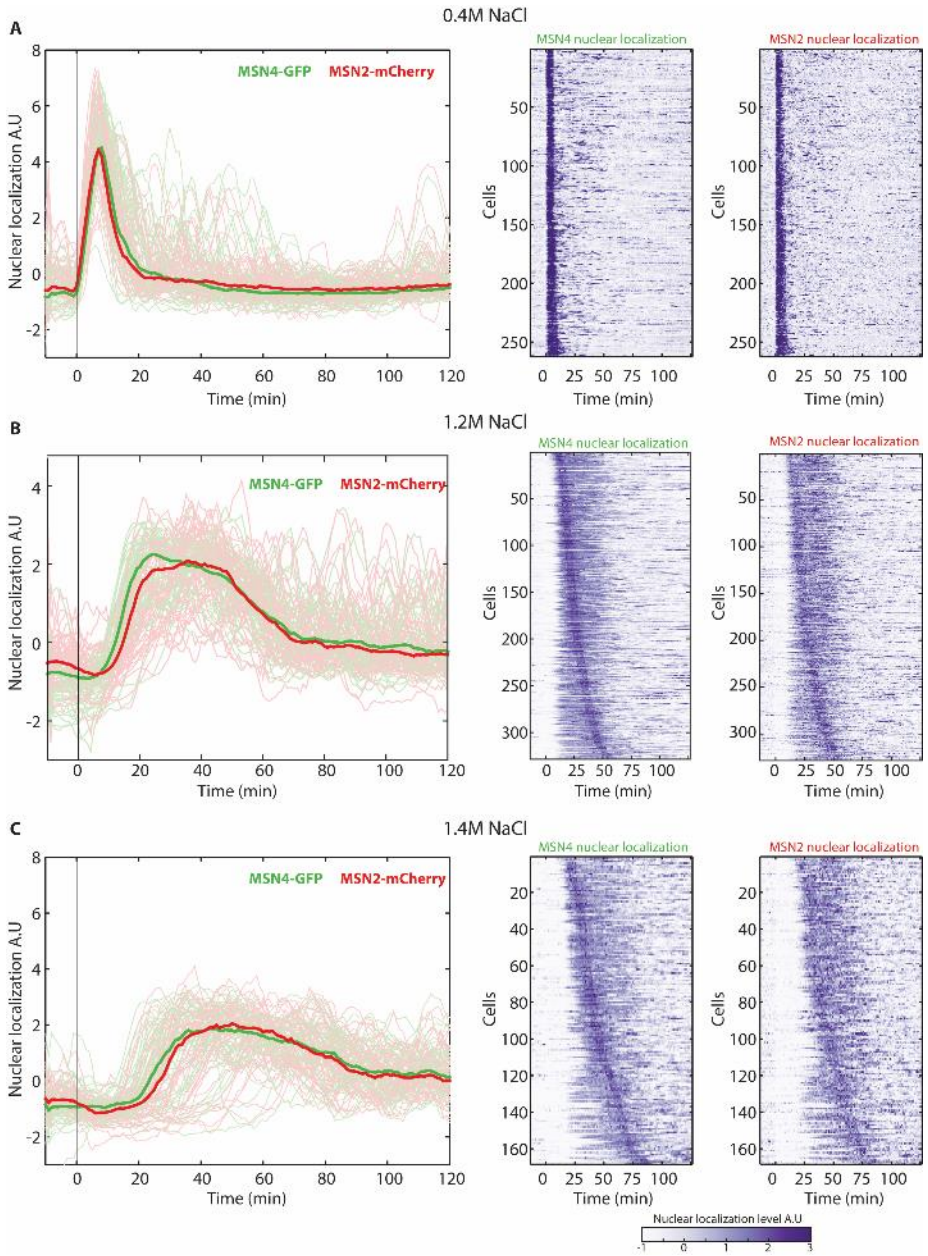

Fig. S2. Nuclear translocation of Msn2,4: Single cells expressing Msn4-GFP and Msn2-mChrerry fusion proteins were tracked using microfluidics-coupled live microscopy in 0.4/1.2/1.4 M NaCl. (left) Localization dynamics following exposure to stress is shown as the medians, and the single cell traces as shaded lines. (right) Individual nuclear localization traces of both Msn2 and Msn4 are shown, with cells in both columns presented in the same order.

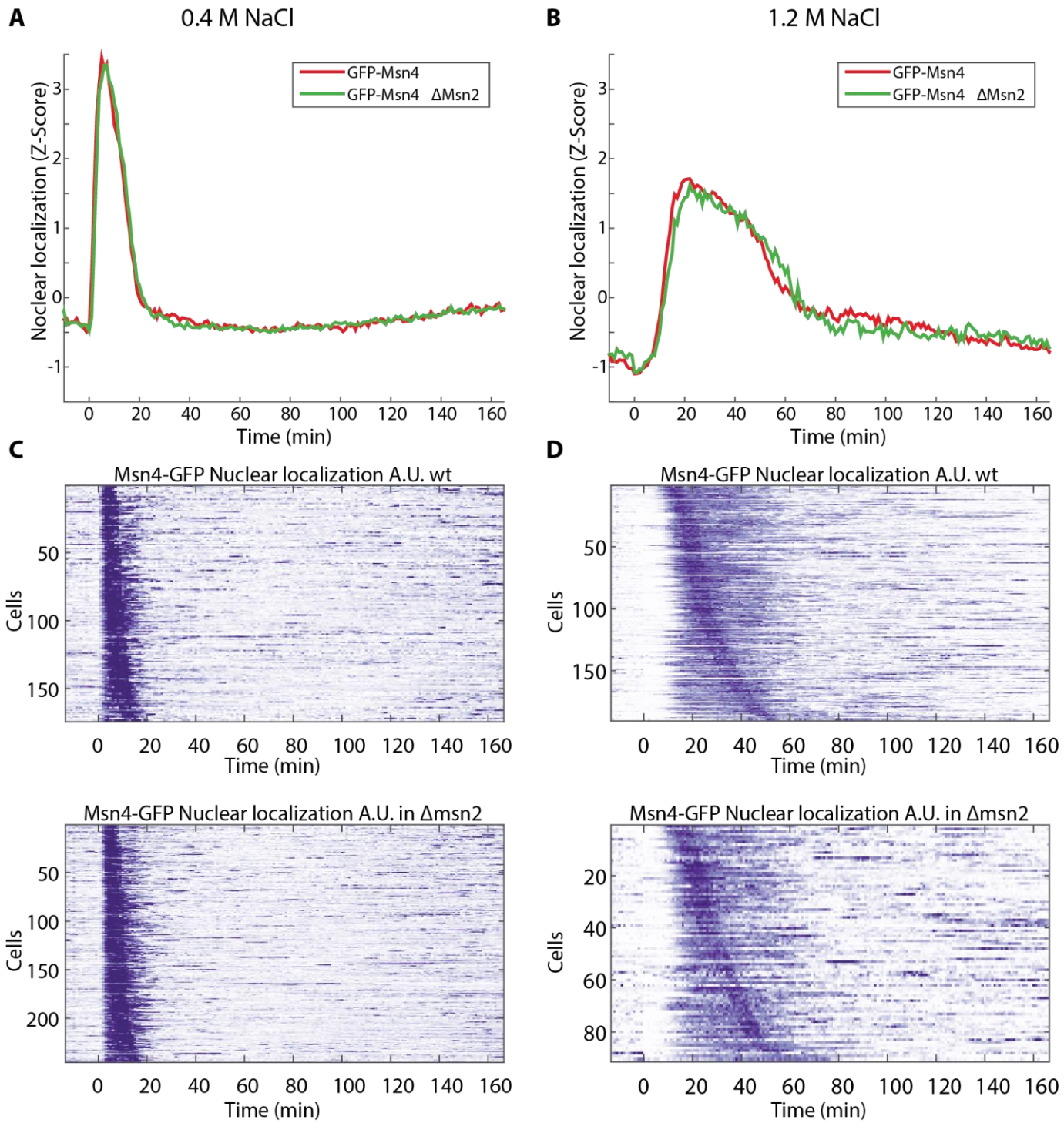

Fig. S3. Nuclear translocation of Msn4 in wild-type cells and cells that are deleted of msn2. (A,B) Localization dynamics following exposure to 0.4/1.2M NaCl is shown as the median. (C,D) Individual traces of Msn4 in wt cells of cells deleted of msn2.

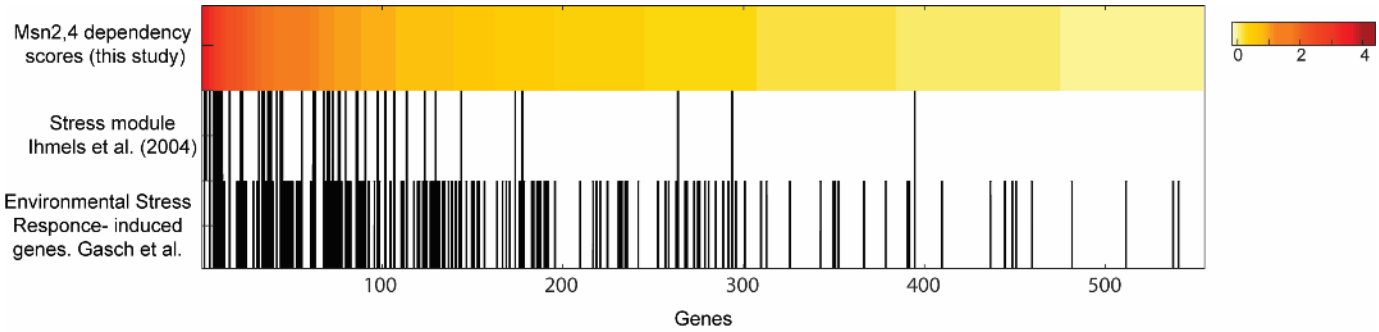

Fig. S4. Msn2,4 dependent genes. We calculated Msn2,4 dependency score for each gene as the average over all conditions, of ratio between wild-type induction and the double msn2,msn4 deletion strain induction. 500 top Msn2,4 dependent genes are ordered by this score. Shown are the scores and an indication if the genes are part of the written published datasets[12,64] (black: gene is part of the group, white: gene is not part of the group.).

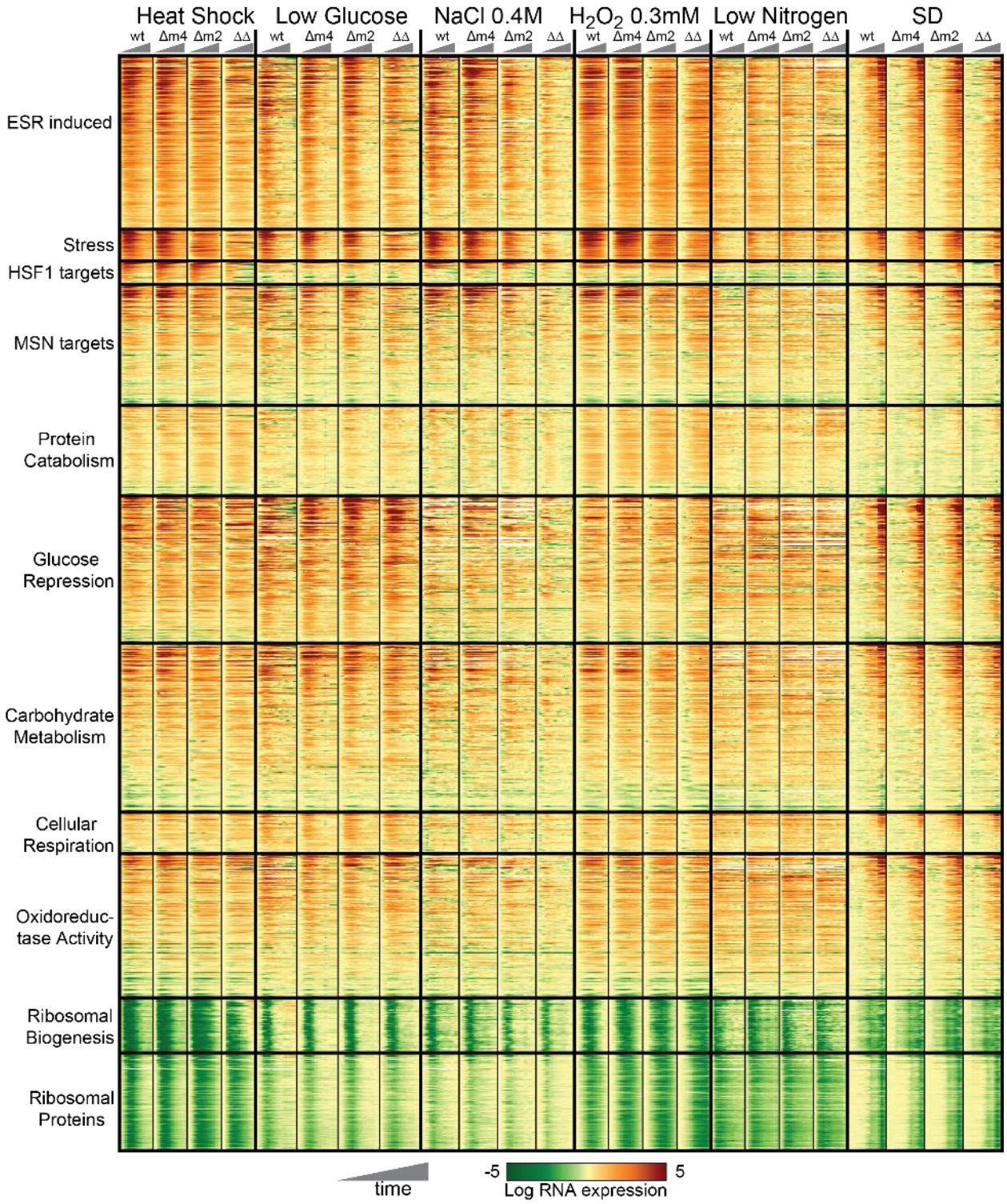

Fig. S5. *RNA expression in all Conditions.*

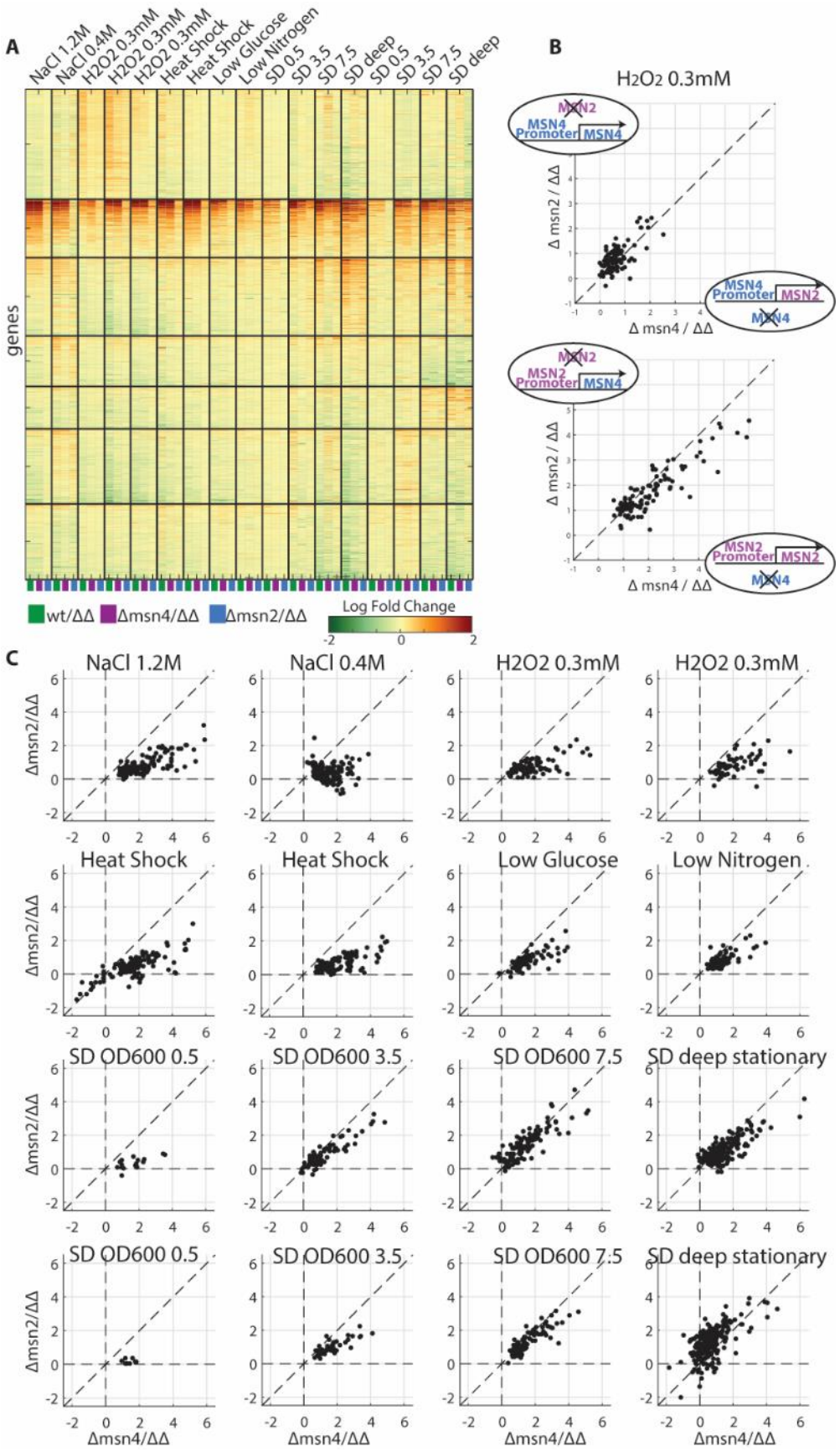
A
NaCl 1.2M
NaCl 0.4M
H2O2 0.3mM
H2O2 0.3mM
H2O2 0.3mM
Heat Shock
Heat Shock
Low Glucose
Low Nitrogen
SD 0.5
SD 3.5
SD 7.5
SD deep
SD 0.5
SD 3.5
SD 7.5
SD deep
genes
wt/ΔΔ
Δmsn4/ΔΔ
Δmsn2/ΔΔ
Log Fold Change
B
H2O2 0.3mM
MSN2
MSN4
Promoter
Δ msn2 / ΔΔ
Δ msn4 / ΔΔ
C
NaCl 1.2M
NaCl 0.4M
H2O2 0.3mM
H2O2 0.3mM
Heat Shock
Heat Shock
Low Glucose
Low Nitrogen
SD OD600 0.5
SD OD600 3.5
SD OD600 7.5
SD deep stationary
Δmsn2/ΔΔ
Δmsn4/ΔΔ

Fig. S6. *Msn2 and Msn4 induce the same target genes.* (A) Clustering of all genes in all the conditions and repeats that we checked. For each experiment of the stress perturbations we calculated for each strain the A.U.C, and for cells growing into stationary phase we used expression in different ODs. We then calculated the fold change of wt or single deletions to the double deletion strain, and used these values to cluster genes. (B) Swapping Msn2,4 promoter. Shown is the fold change of single deletions to double deletions, in the indicated strains. (C) Each dot represent a gene that was >2 fold higher in the wt then the double deletion. Shown are fold changes of single deletions to double deletion.

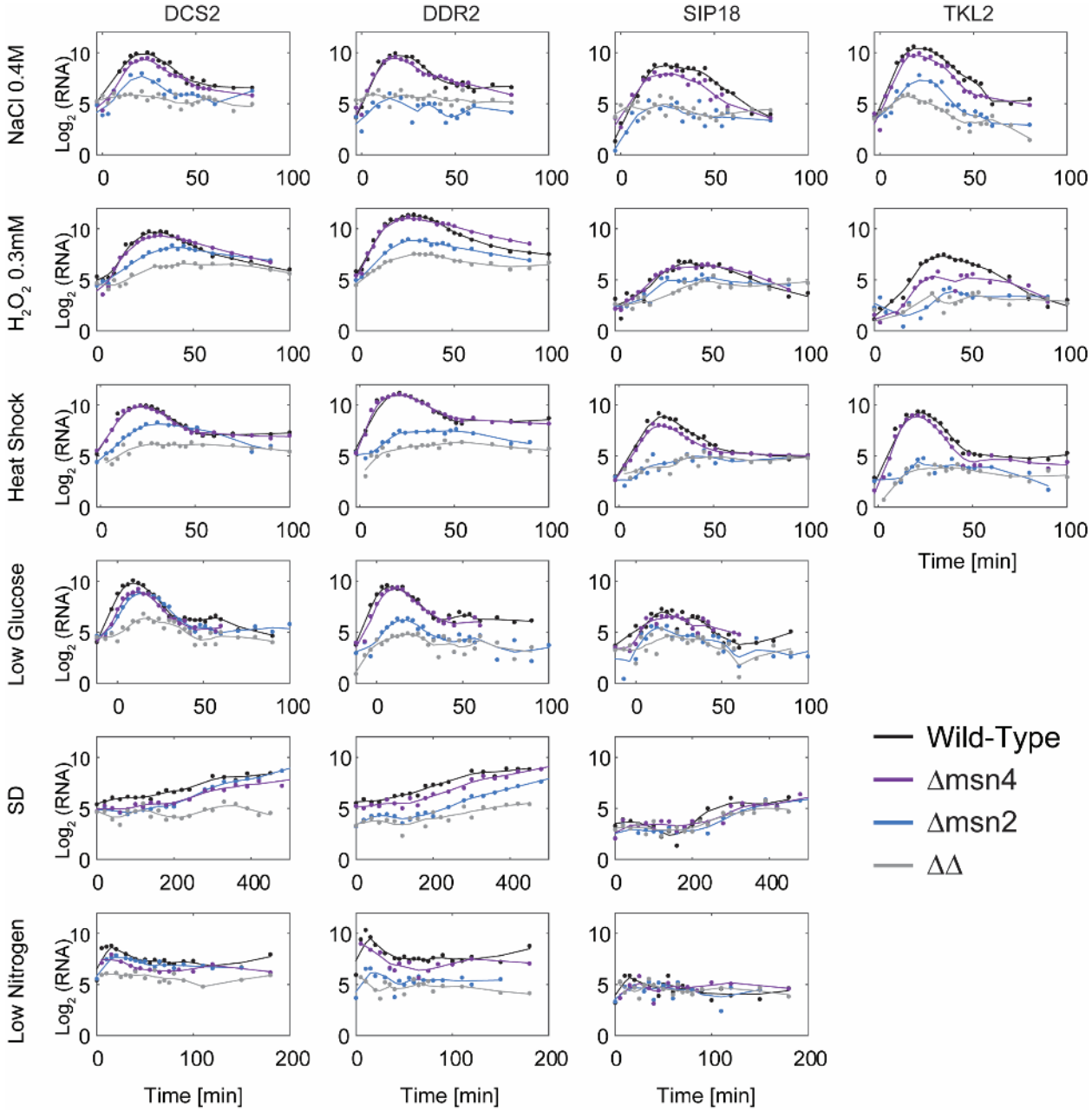

Fig. S7. *Response to stress of reported genes from AkhavanAghdam et al*[39]. Plotted are mRNA measurments (from this study) of the response of 4 genes reported in AkhavanAghdam et al. Shown are mRNA measurments for the wild-type, single and double deletion msn2,4 strains. Dots represents the data measurments and lines is the smoothened signal. In our high temporal resolution data there is no fundamental difference in Msn2,4 contribution to the reponse between the first two genes(DSC2,DDR2) and last two genes (SIP18,TKL2) as was reported. In all of these gense Msn4 deleted strains show similar epression and dynamics to the wt strain, where Msn2 deleted stains reduces the induction significantly.

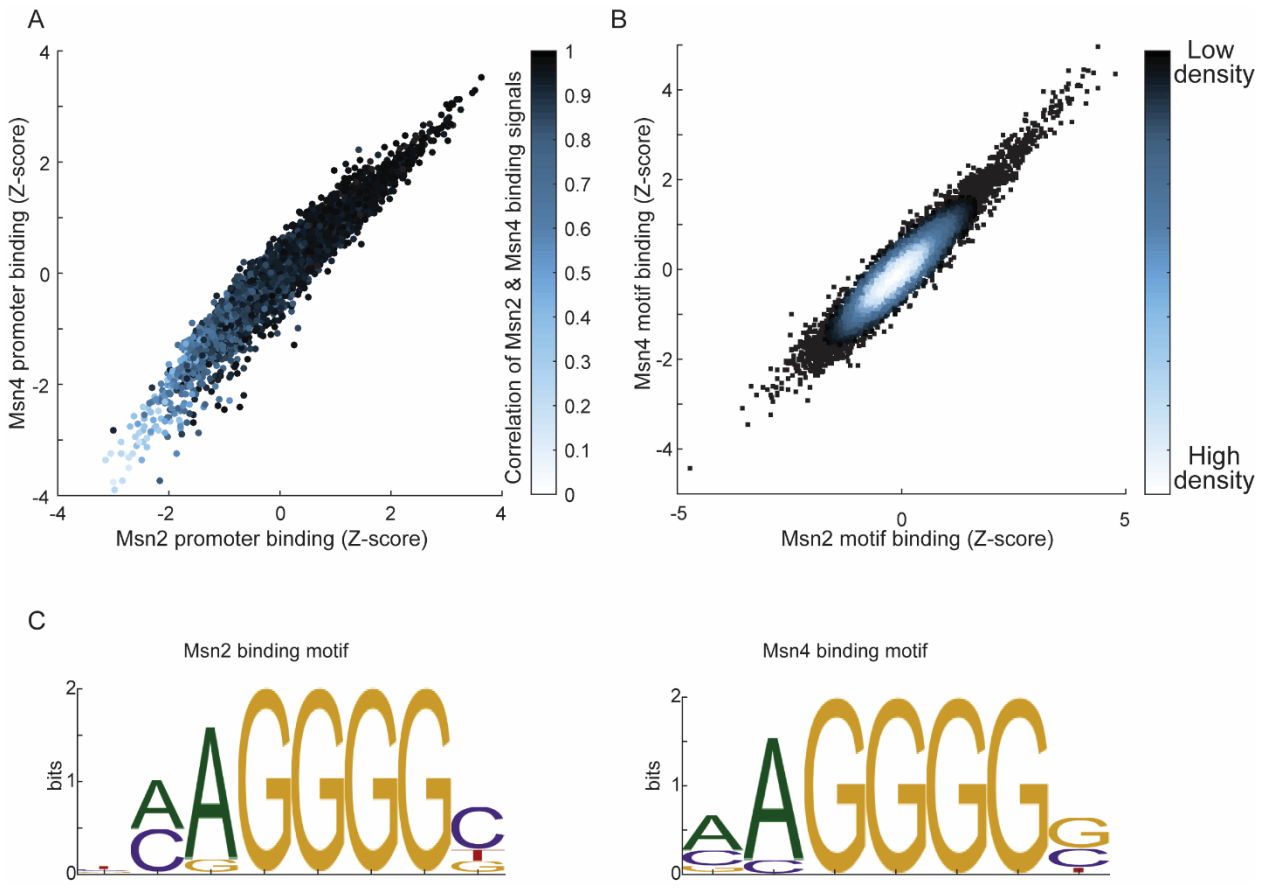

Fig. S8. *Msn2,4 prefer the same binding sequence and the same promotors in-vivo.* (A) Msn2 and Msn4 binding to all the promotors. Sum of the normalized CHeC-seq signal of each factor was calculated for all the promotes, in >4 repeats. Shown is a the z-score of the median of all repeats. Color represents the correlation of Msn2 and Msn4 binding signal on the promoters. (B) Density plot comparing Msn2 and Msn4 *in-vitro* binding to all possible (8192) 7-DNA base-pair sequences. For each 7-mer, the mean signal of all its appearances in all promoters was calculated for Msn2 and Msn4. Shown is the density plot of the z-scores of all possible 7-mers. (C) Motifs found in our data for Msn2 and MSN4.

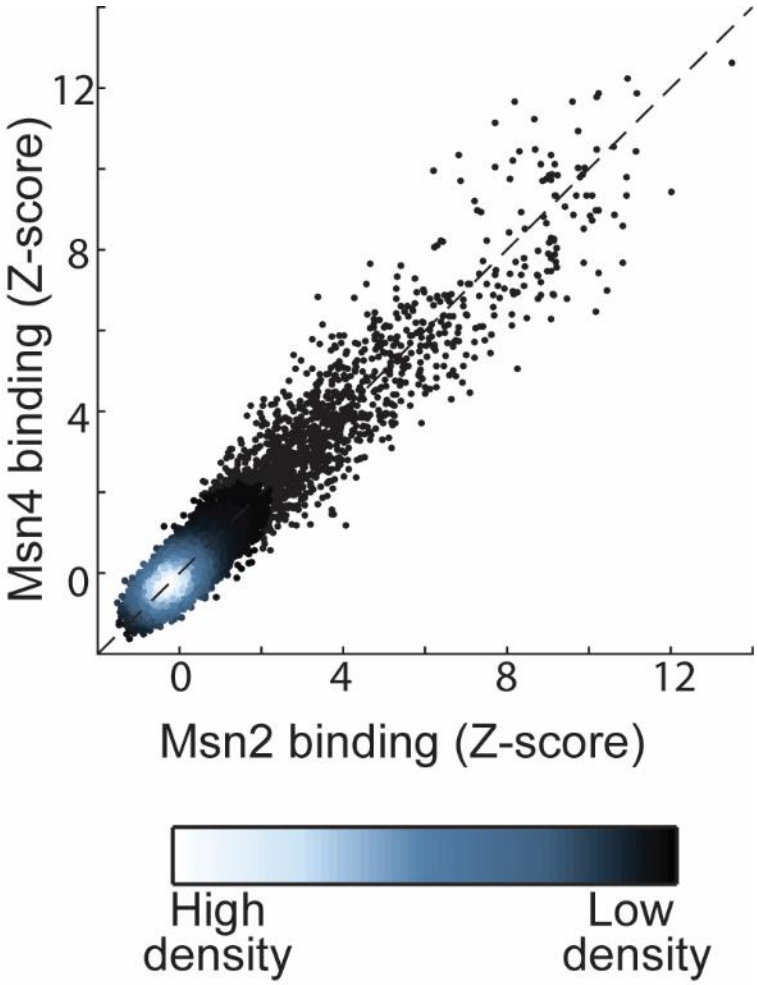

Fig. S9. *Msn2,4 prefer the same binding sequence site in-vitro*. Density plot comparing Msn2,4 *in-vitro* binding to all possible (32,896) 8-DNA base-pair sequences. Data from Siggers et al.[41]

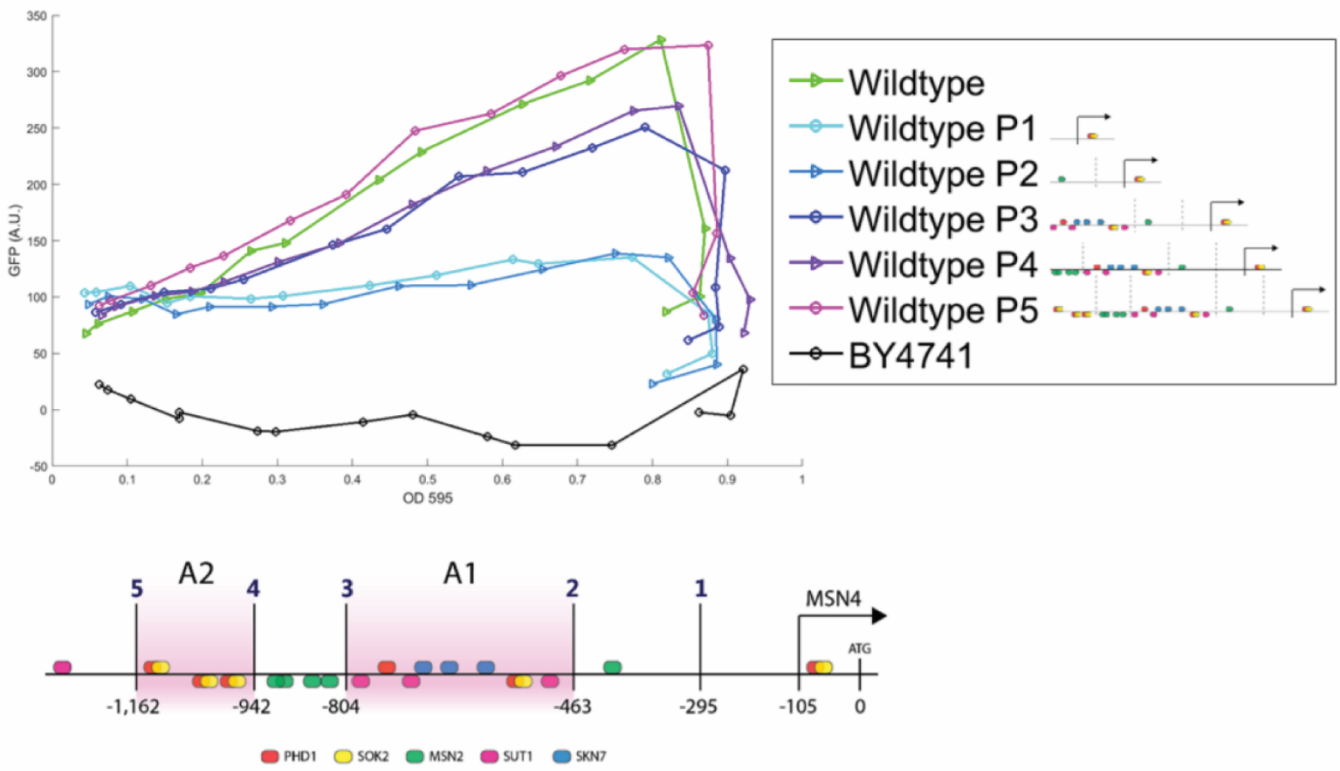

Fig. S10. *Msn4 promoter regions*. Median expression of Msn4-GFP along the growth curve in strains with full and partial *MSN4* promoter. The scheme shows the promoter regions that induce Msn4 at high ODs.

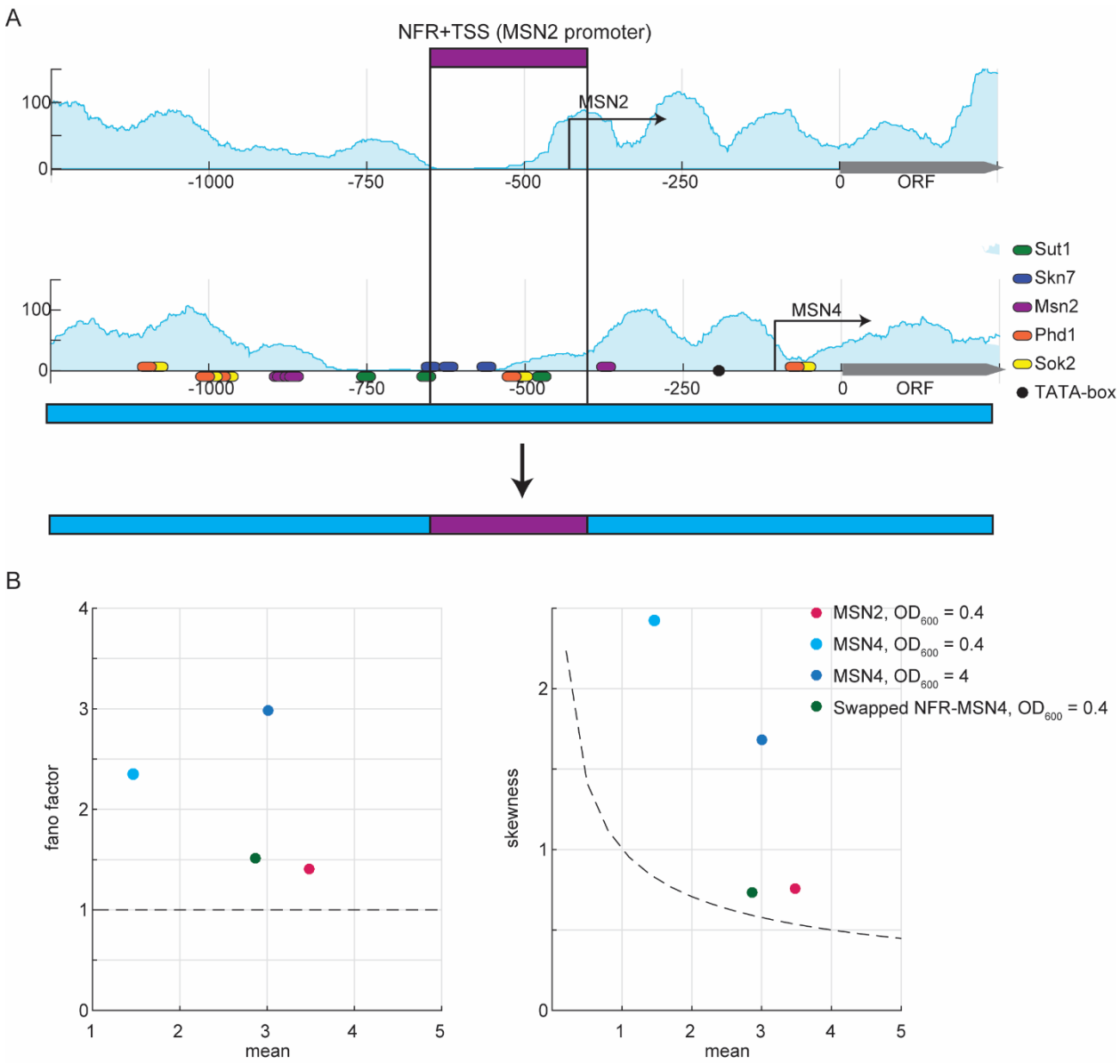

Fig. S11. *Msn2 NFR and TSS promoter region determine the expression level and noise.* (A) A scheme of the strain we used – *MSN4* promoter with a swap with *MSN2 NFR+TSS* in the same position. (B) smFISH results of the swapped strain and the wt Msn2,4 in the indicated ODs.

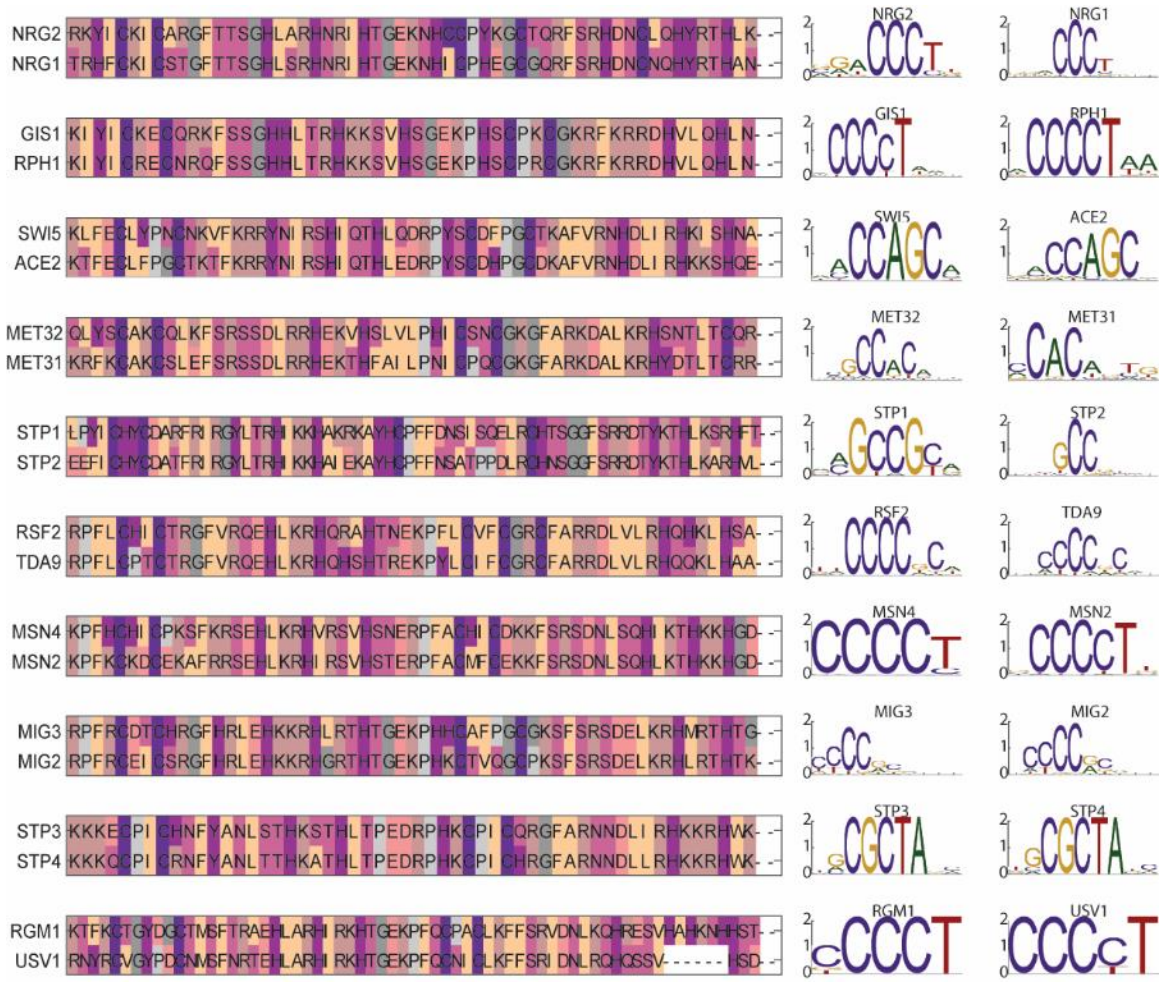

Fig. S12. *All S. cerevisiae Zinc finger TFs duplicates from WGD.* (left) Alignment of binding domains of all duplicated pairs. (right) DNA binding motifs of the pairs from YeTFaSCo[65].

Table S1. Yeast strains used in this study.

	Strain	Genotype	Source
1	BY4741	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	-
2	BY4742	MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0	-
3	wild-type	BY4741	-
4	Δmsn2	BY4741 msn2Δ:: hphNT1	This study
5	Δmsn4	BY4741 msn4Δ:: natNT2	This study
6	Δmsn2Δmsn4	BY4741 msn2Δ:: hphNT1 msn4Δ:: natNT2	This study
7	Δmsn2, MSN4-promoter swap	BY4741 msn2Δ:: hphNT1 MSN4promoterΔ:: natNT2-MSN2promoter	This study
8	Δmsn4, MSN2-promoter swap	BY4741 msn4Δ:: natNT2 MSN2promoterΔ:: hphNT1-MSN4promoter	This study
9	Δmsn4, mCherry	BY4741 msn4Δ:: natNT2 HO::Kan+TDH3prom-mCherry	This study
10	Δmsn4 & MSN2-YFP	BY4741 msn4Δ:: natNT2 MSN2-YFP::kanMX4	This study
16	Δmsn4 & MSN2 promoter library - 50 strains	BY4741 msn4Δ:: natNT2 MSN2-YFP::kanMX4 MSN2promoterΔ::URA-promoterX	This study
17	GFP-Msn4	BY4742 GFP-MSN4 (N terminus tag), different colonies	This study, based on a strain from Yofe et al.[66]
18	GFP-Msn4 & Δmsn2	BY4742 GFP-MSN4 msn2Δ:: hphNT1	This study, based on strain 16
19	GFP-Msn4 & Msn2-mCherry	BY4742 GFP-MSN4 MSN2-mCherry::His3	This study, based on strain 16
20	Msn2-GFP	BY4741 MSN2-GFP:: kanMX4	This study
21	Msn2-Mnase	BY4741 MSN2-3FLAG-MNase:kanMX6	This study
22	Msn4-Mnase	BY4741 MSN4-3FLAG-MNase:kanMX6	This study
23	K.lactis promoter-Msn2	BY4741 MSN2promoteΔ::K.lactisMSNpromoter	This study
24	MSN4 promoter with MSN2 promoter NFR+TSS	BY4741 MSN2promoter(400-650)Δ:: MSN4promoter(400-650)	This study

Table S2. MSN2 smFISH probes, CAL Fluor Red 590

PROBE #	PROBE (5'-> 3'	PROBE #	PROBE (5'-> 3'	PROBE #	PROBE (5'-> 3'
1	tgaaatcatggtcgaccgtc	17	agcatggagtctatgttcag	33	ccactttcgcaataacggac
2	ctactcatgctttctatggg	18	gctaaatcttcggcgtgata	34	tcattactcaagcctgtagt
3	ttgggttattattctccacg	19	tggttccaaaggtccaaaga	35	ttcaccttcctctgtcaaaa
4	ccgcactatctaagacagtg	20	cttgcattggaacttgaggt	36	ggagccatattcatttgagt
5	ccccaaattcagtgaagttt	21	ttgcgttcatagtagtggta	37	ttgcaagagacgtggaggat
6	gtagtggatggtatcgtttc	22	gttgccagcaatatttgagt	38	tccgaagggaagaaccgatt
7	tttgttgctatcagctttca	23	ctatggtagcgtcattgttt	39	tcaaaggcacagcagacttc
8	tcaatagttcttgcatcgcg	24	agtggatgtgcccaaattat	40	tgttgatttgtaagcggcac
9	gagccactagcattattagt	25	ggcaagcagatttgttcttg	41	gagttgacactactactgct
10	gagttgtgtattgatttgcc	26	atctttgtcatagcattggc	42	tgtcattgattttctcctgt
11	gtgacggtgattgtaacgga	27	tgctgtaattgctgacttgg	43	cgacacttgatcttctggac
12	ggattttgaggtaccgatga	28	tgtgttgaactcggtttagc	44	tcgagttcctttgttgattc
13	cttgctgtatttatgggagg	29	ttatgtgacgaggtgagctg	45	tgtgacagtggaacggtttc
14	ccatttgaagtttgaggcga	30	aaagaagctctttgccttct	46	atgccttttcaaatgttcgc
15	gcaccattaggtaaggacag	31	tgcttattaaccaggtcgta	47	ggtcgttcgttagagtgaac
16	acccgtagaattggagtttg	32	aattcggcagcatatcgttc	48	gtcttgatgtgttgcgacaa

Table S3. MSN4 smFISH probes, CAL Fluor Red 590

PROBE #	PROBE (5'-> 3'
1	ttaggtccgaagactagcat
2	ttgtttcttgtttgcgtgac
3	ggctcgttcattatagacga
4	cagaaacgttggttgtgctc
5	gctattgttcgcagaattgt
6	aaggtatattccggcgaaga
7	ctatgcggagaatccattga
8	attggagtattagtggcgtc
9	cgtattagttgtcgctgtta
10	ggcggtatcctttaaattca
11	gaatttttccacttccagtt
12	cattcacgaactgtgcgtta
13	agcattactgctttgtttct
14	cttatcaccaccatttttct
15	gggtggctactgaactatat
16	ctagttgctccaagtttttc

PROBE #	PROBE (5'-> 3'
17	catgttggtgtggcgataat
18	aggtttgtcttgggagaagt
19	atattgaaggttctgcctca
20	gtttggtaacctagcatcta
21	cgttcaatcctgttgtactg
22	cggcatgaacttcaatgctt
23	aaggaagaggtaaccacctc
24	gttgagtgggtttactgttg
25	gcagtaccttcagggataat
26	ggtgaaatggttgatccacg
27	gctcagagcttgagaaacgt
28	tttgttataccagagtcctt
29	gctatgggagatgcgattta
30	ttaagggcaaagaggcacgc
31	gcggaatcagacagaggatt
32	caccaaaagcatcgtcttca

PROBE #	PROBE (5'-> 3'
33	ggttgcatctaagctattgt
34	gggtctacgttattgtcgaa
35	taaactgttgatcctgagcc
36	cagaggcattatccgacaac
37	accagaagtcgccaattta g
38	tttggaaagctggtgtaggt
39	tggtgagcttgactagtagg
40	gttgttgatgatgttgagct
41	gacgactttcttcttcttgt
42	aaccttaccattgttgttgt
43	atagatttccttttccgagg
44	tttatcgtagttgttggggt
45	cttctcacagtctttacact
46	cggatcttatatgccttttc
47	tacaagcaaaagggcgttcc
48	cgtgctttttgtgagttttt