Supplementary data for manuscript:

Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation

Ophir Shalem, Orna Dahan, Michal Levo, Maria Rodriguez Martinez, Itay Furman, Eran Segal, and Yitzhak Pilpel

1. Preprocessing decay data

The first step in any microarray experiment, after hybridization and scanning, is data preprocessing in which gene specific expression values are calculated from probe level data. One of the important aspects of this step is to separate between biological to non-biological variation in the measured intensities that occur both within and between samples. When comparing multiple samples most preprocessing algorithms use a normalization step in order to bring all samples to have the same global distribution of intensity values. This is done under the assumption that global deviations, e.g., in the mean intensity between samples, represent mostly technical artifacts that result from a difference in the processes that the samples undergo after RNA extraction till hybridization and scanning. Thus, a net change in the measured intensity distribution would be ignored and biological insight would be inferred only based on changes in the relative ranking of individual genes within the intensity distribution of each sample. A unique aspect of mRNA decay measurement is that the above assumption is by definition not valid: Due to transcription inhibition a global decrease in the total amount of mRNA is expected as the time course advances, which makes standard preprocessing procedure non applicable to this data.

In order to normalize microarray samples with respect to each other we followed a procedure similar to what was done previously[1]. An internal standard, containing a pool of 4 "spikes" - *in vitro* transcribed polyadenylated *B.subtilis* genes (poly(A) control kit supplied by Affymetrix) in different dilutions (1:100,000; 1:50:000; 1:25,000 and 1:7,500), was mixed with each RNA sample following RNA extraction. Each *B.subtilis* gene is represented on the microarrays by several probe sets. Figure S4 shows that indeed a linear relationship is observed between the known concentrations to the measures probe set intensities of the spiked-in genes; the difference between two microarray samples is also observed in this figure. The amount of internal standard added to each sample was determined such that the intensity values of the four *B.subtilis* genes would span most of the signal probes intensity distribution.

mRNA accounts for less than 5% of the total amount of RNA in each sample; therefore, although the same amounts of extracted RNA were taken for processing and hybridization, we expect that mRNA degradation will result in a decrease in the relative amount of mRNA in each sample as the time course proceeds. For cDNA preparation poly(T) primers were used in order to only amplify polyadenylated RNA together with

the polyadenylated "spiked-in" RNA. Due to a decrease in the amount of mRNA in each sample, after cDNA preparation, since we take equal amounts of cDNA for further processing we expect an increase in the concentration of the "spiked-in" RNA relative to the mRNA which accounts for most of the cDNA in the sample. The increase in the mean spike intensities for the non treated time course (Figure S5) indicates that indeed global mRNA degradation is captured in the ratio between the mean signal intensity to the intensity of the spiked-in RNA.

We describe below the procedure which was eventually chosen in order to compute expression values and produce the results presented in this article. Other procedures were also considered: ranging from different algorithms (RMA, MAS5 and more simple summation methods) used to calculate the initial expression values, to different ways to scale the samples according to the intensity values of the internal standard. The results presented here did not change qualitatively between different procedures (data not shown) and the chosen procedure was determined according to technical parameters alone.

Preprocessing was composed of two main stages: an initial calculation of expression values irrespective of the effects of degradation and rescaling of each sample in order to recapitulate its effect. The first step was preformed using the standard RMA algorithm[2] which includes background correction for each sample separately, median polish summation and quantile normalization between samples.

In the second step, in order to capture the effect of degradation, each sample was scaled as follows: a least square error linear fit was preformed on the spike intensities against their known concentrations in double logarithmic scale. The intensity values of each microarray were divided by the slope of the fitted line to get a slope of unity for all microarrays and the intercept differences between every microarray to the microarray with the minimal intercept were subtracted. As expected from the fact that this procedure was done in log scale, most of the scaling between samples resulted from the intercept subtraction and the slope correction had a very slight effect on the scaled intensities. Figure S6 shows the mean signal intensity for all samples of three time courses against the time at which the samples were taken relative to transcription inhibition. It is apparent that global mRNA amounts decay monotonically due to transcription inhibition and that all three conditions display relatively similar rates. To show that indeed most of the genome decays, and in a constant rate throughout the experiment, each gene is fitted to an

exponential decay model and an r-square value (goodness of fit) is calculated for each gene (figure S7).

A complete biological replicate of the reference decay profile was preformed in order to asses the reproducibility of half life measurements. Due to technical reasons of outlier samples the half lives for the second biological replicate of the reference decay profile were calculated without the last 3 time points. We assessed the agreement between the two replicates in terms of the Pearson correlation coefficient between the estimated half lives of genes in the two replicas. We collected all the genes that passed various R-square cut-offs in the two replica and measured correlation for each such set of half life pairs. For R-square cut-offs of 0.6, 0.7, 0.8 and 0.9 4014, 3171, 2160 and 960 genes respectively passed the thresholds and they showed correlation coefficients of 0.73, 0.78, 0.85 and 0.92 respectively. The conclusion is that even for a complete biological replicate of the entire experiment half life value estimations are quite robust, especially for genes that show a good fit to the exponential decay model.

A different kind of replicate was also preformed in order to increase confidence in key reference measurements. For each of the four decay time courses, two samples were taken for the first time point (at the point of transcription inhibition) and both were hybridized to the arrays. Figure S8 shows the correlation between two replicate arrays. In all further calculations the mean intensity of the two replicates at time zero was used as a reference for the expression level of each gene at the moment of transcriptional arrest. Figure S8 also shows for comparison the correlation between two consecutive arrays with 20 minutes difference between them. The replicate arrays display high reproducibility while degradation is clearly observed between the two consecutive arrays.

Although the three conditions display similar global mRNA decay rates, there are still differences between the profiles which probably result from inaccurate measurements of the relative spike intensities between conditions which were hybridized in different batches. These differences might result in biases when comparing gene specific decay values between conditions. For this reason, before comparing conditions, all samples were scaled such that the three conditions would have the same decay of the mean sample intensity. This step assumes that the mean mRNA amount decays in a similar rate between conditions. We find this a reasonable assumption based on the results in figure S6. This step assures that genes determined to have different decay rates between two conditions are only those that change their rankings compared to the whole genome; in other words such genes do not represent global decay changes between the

conditions or scaling inaccuracies. The assumption, that the mean mRNA amount at a given time following transcription inhibition is equal between conditions, is equivalent to the assumption that there is no net change in mRNA distribution between samples used by any standard micro-array preprocessing algorithm. Still, the same results, as presented in this manuscript, were achieved even without this last step (data not shown).

2. Comparison to previous mRNA abundance measurements

2.1 Oxidative stress

We compared our mRNA abundance results to three more studies that measured changes in mRNA abundance following oxidative stress: Gasch et al.[3], Mendes et al.[4], and Molina-Navarro et al.[5]. Table S1 sums up the treatment that each group used and the platform that was used for hybridization.

	Treatment	Platform
Gasch et al.	0.3 mM Hydrogen Peroxide	cDNA arrays
Mendes et al.	0.19 mM CHP	Affymetrix Yeast Genome S98 arrays
Molina-Navarro et al.	0.1 mM t-BOOH	cDNA arrays
Shalem et al.	0.3 mM Hydrogen Peroxide	Affymetrix Yeast Genome 2.0 arrays

Table S1 – List of treatment and platforms used by each study.

We compared two parameters of the response: The first is the general temporal dynamics and time scales of the response and the second is the number of genes which respond either by induction or repression and the overlap of these groups between different studies. Figure S9 shows the mean mRNA abundance response for the four data sets. The mean of induced and repressed genes, defined by the fold change compared to the first time point (of genes that responded by at least two fold), is plotted as a function of time for each study. In the current study, and in the studies by Gasch et al.[3] and Molina-Navaro et al.[5] induced genes show transient kinetics. In addition the general temporal behavior of genes is similar across the studies in terms of time to peak and activation/relaxation speed (not shown). In the data by Mendes et al.[4] the averaged response is not transient but sustained (Figure S9). Yet more detailed clustering analysis

of these data shows two main groups of genes, one that shows a fast transient response and another that shows a more sustained response (Figure S10).

Second, we checked the overlap in responsive genes, induced and repressed, between the four studies. The results are shown in Table S2.

Induced genes					Repressed genes				
	Mendes et al. (992)	Gasch et al. (1688)	Molina-Navarro et al. (565)	Shalem et al. (679)		Mendes et al. (787)	Gasch et al. (963)	Molina-Navarro et al.(2065)	Shalem et al. (535)
Mendes et al. (992)					Mendes et al. (787)				
Gasch et al. (1688)	463				Gasch et al. (963)	325			
Molina-Navarro et al. (565)	226	278			Molina-Navarro et al. (2065)	449	523		
Shalem et al. (679)	378	433	197		Shalem et al. (535)	297	326	360	

Table S2 – The number of induced and repressed genes in each study with the sizes of intersections: For each of the four studies examined, the number of induced and repressed gene was determined by taking all the genes with a fold change of above (or below) two fold, for at least one time point. This number is given in brackets alongside with the name of the study's first author. The numbers is the middle of the table represents the intersection between the two sets: the number of genes which are induced or repressed in both studies.

Generally this simple analysis shows that our mRNA abundance data is in relatively good agreement with previous studies, compared to the differences between different previous works. The difference in the temporal dynamics and the identity of responsive genes, between the different studies may be due to the different severity of the treatments and different genetic background of the yeast strains used for each study.

2.2 DNA damage stress

We have compared our mRNA abundance results to a work by Gasch et al.[6], which also measured changes in mRNA abundance following DNA damage stress. They have used a 0.02% of MMS in order to introduce DNA damage and used cDNA arrays for hybridization, while we used 0.1% MMS and used Affymetrix arrays. Here too, the same two parameters of the response were compared, namely the general temporal dynamics and the identity and amount of responsive genes. Figure S11 shows the mean

expression profile of all responsive genes (induced/repressed above/below two fold). Interestingly in both data sets the response to DNA damage is sustained compared to the transient response to oxidative stress although Gasch et al. used a significantly lower concentration of MMS. This is reflected by the number of genes which respond, i.e. changing by more than two fold (Figure S11). 67 and 73 percent of the genes which are induced and repressed above two fold in the study by Gasch et al. were also found to respond in our study. The differences might be ascribed to the different severity of treatments, different array platform, and the different genetic background between yeast strains.

3. Comparison to transcription rate measuremnts

In the study of Molina-Navarro et al.[5] the authors measured transcription rates and balanced changes in mRNA abundance in several time points during 70 minutes following the application of oxidative stress. They calculated decay rates (Kd) based on the discrepancies between these two measurements and use this in order to report changes in mRNA stability following oxidative stress.

The most notable discrepancies, between changes in transcription rates to changes in mRNA abundance, which result in predicted changes in mRNA stability are in the first two time points (7 and 16 minutes), but also in later time points, sometimes showing opposite trends between early and late time points for the same genes (figure 2 and 3 in Molina-Navarro et al.). For example cluster 23 shows a general induction with early stabilization (decreased Kd) and late destabilization (increased Kd).

To generally test the agreement between our results to the results published by Molina-Navarro et al. we ran an analysis, similar to the one in our manuscript, on the data published by these authors. We first checked whether induced genes have a general tendency towards stabilization or destabilization by plotting the Kd difference, at each time point, against the maximal fold change. Kd difference was calculated by subtracting the Kd value at time point 1min from the Kd value at each time point. We used subtraction instead of the logarithm of the ratio due to the presence of negative Kd values. Because these authors calculated a Kd value for each measured time point, the relationship between fold change to the change in decay constant will depend on the time

point at which the Kd value is calculated. We find in the first time points (7 and 16 minutes) that induced genes show a general tendency towards stabilization while later, in time points 26 and 41 minutes, this trend is reversed and induced genes show clear destabilization (Figures S12 A and B). In our experiment we stopped transcription 25 minutes following treatment, thus our destabilization is in agreement with their destabilization observed at time points 26 and 41 minutes. In addition, the early stabilization observed in the first time points is actually based mostly on genes for which the calculated Kd is negative. These values are probably due to an under estimation of the transcription rate values, as indicated by the authors and do not represent biological relevant Kd values (Red points in Figure S12 A). Thus drawing conclusions based on these results is highly problematic. The destabilization of induced genes, observed from time point 26 and on, is enhanced if only the most transient genes are taken for analysis in agreement with our results (Figure S12 C). The early stabilization and late destabilization of transient induced genes was actually validated by the authors for two genes (figure 4 A and B Molina-Navarro et al.).

Repressed genes show a general destabilization and also a sustained repression, this also is in agreement with our results (Figure S12 D). Specifically Ribosomal and rRNA processing genes which are indicated by Molina-Navarro et al. as repressed and destabilized following oxidative stress show the same behavior in our data (table 1 in manuscript).

Thus the two works are in general agreement although using very different methodologies – in both works balanced mRNA response is measured; in our case transcription is arrested, in their case transcription rate is measured directly. In both works a transient induction is accompanied with destabilization which is observed approximately from the point were the genes reach their peak in mRNA abundance. The connection between transient induction to destabilization is strengthened by the fact that taking the most transient genes enhances the signal as observed in both studies.

Although the general repression profile, in response to oxidative stress, is more sustained in the results of Molina-Navarro et al. the general result, by which sustained repression is accompanied by destabilization is in agreement between the two studies. The early stabilization of induced genes, which is observed by Molina-Navarro et al., is based mostly on negative Kd values.

We have also used their direct measured transcription rates in order to check if destabilized genes tend to have higher transcription rates and the opposite for stabilized

as expected by our hypothesis. We strikingly find that this is indeed the case as can be observed in Figure S13.

4. Validation using Real Time PCR

Microarray results were validated for several genes: Following oxidative stress induction accompanied with destabilization is observed, and following exposure to MMS sustained induction along with stabilization is observed, in agreement with the genome-wide arrays observations (Figures S14 A to D).

5. List of supplementary data

S-data1.xls - Contains the full data sets, mRNA abundance measurements in both conditions and all decay profiles together with fitted half life values and R-squares.

S-data2.xls - Full results of data mining, enriched categories in both conditions.

6. Supplementary figures

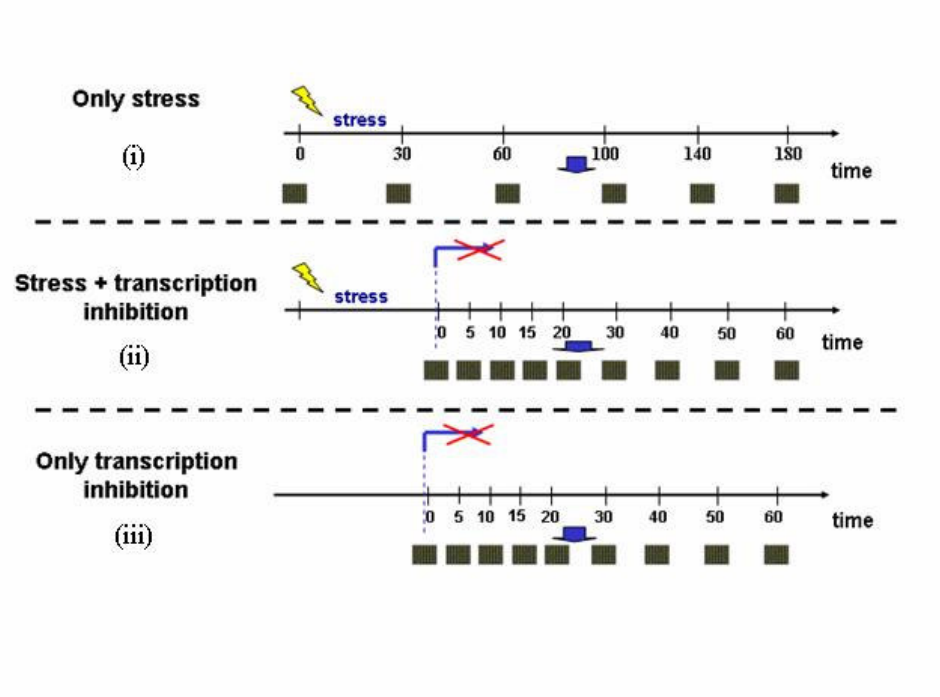

Figure S1: **A schematic illustration of the experimental procedure:** Three types of experiments were conducted: (i) A conventional microarray experiment where mRNA abundance was measured following the perturbation, (ii) a stress followed by transcription inhibition to measure condition specific decay kinetics, and (iii) a reference decay experiment where decay kinetics was measured after transcription inhibition without applying additional stress. The blue arrows with the red mark in the two later experiments indicate the point where transcription is inhibited.

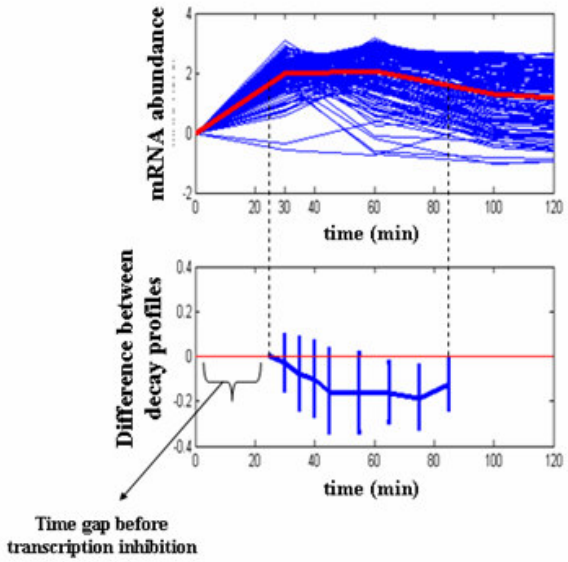

Figure S2: **Difference between decay profiles as a function of time:** For each gene, at each time point, the reference decay profile is subtracted from the oxidative stress decay profile, then at each time point the mean and standard deviation of the difference is calculated and plotted (lower panel) alongside with the balanced mRNA abundance response profile (upper panel). In the induced genes it can be seen that the destabilization starts before most genes start their relaxation stage. This is not the case for the repressed genes where stabilization is more coordinated with the point where relaxation begins (not shown).

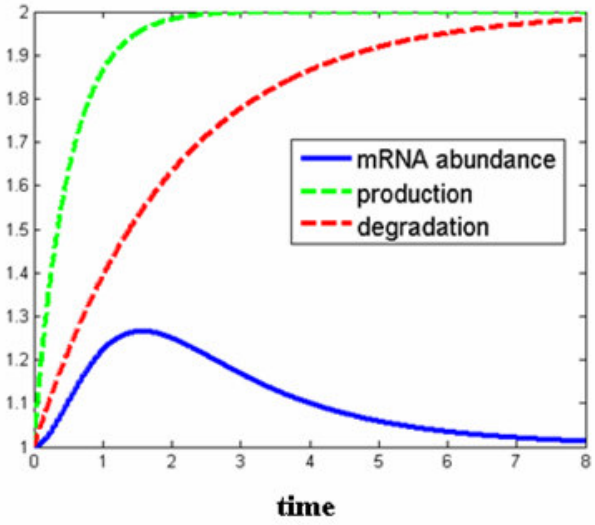

Figure S3: **A proposed model that accounts for the transient changes in mRNA abundance:** Increase in both production and degradation creates a transient increase in mRNA abundance if we assume that the increase in degradation is slower than the increase in production. This time delay will produce a period of time where production exceeds degradation to produce an overshooting of mRNA abundance with respect to the final steady state defined by the ratio between the final production and degradation rates. A similar model can also account for the transient repression where slower decrease in decay rate follows decrease in production rate.

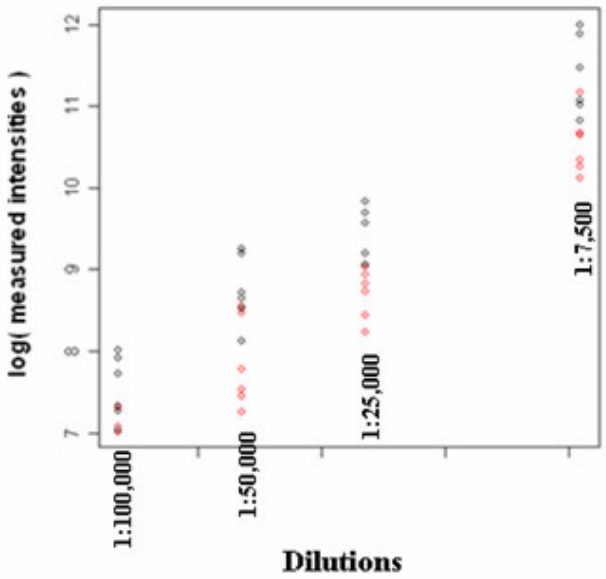

Figure S4: **linear relationship between the spikes known dilutions to measured intensities:** Plot of the known dilutions vs. the measured intensities of all probe sets representing the "spiked in" RNA genes. Each of the four genes is represented by several probe sets. The two colors represent two different microarray samples showing that most of the differences between arrays reside in additive errors after log transformation.

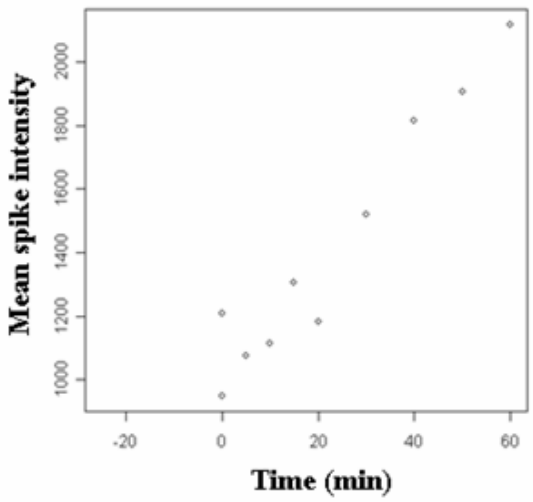

Figure S5: **Increase in the relative concentration of "spiked in" RNA:** The mean spike intensity for each sample is plotted against the time at which it was taken after transcription inhibition. A monotonic increase is apparent as expected.

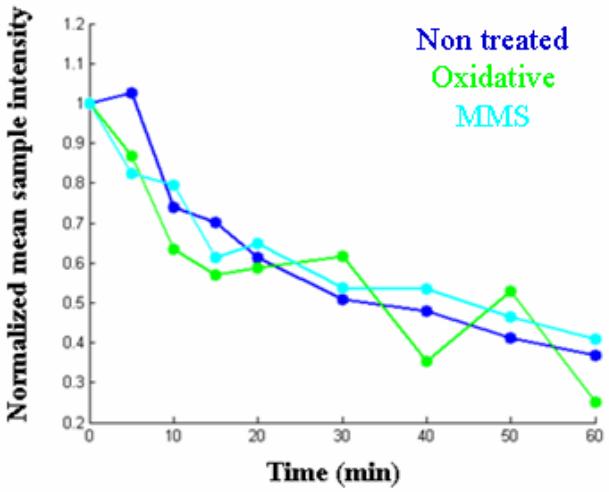

Figure S6: **Mean mRNA amount decreases with time:** The mean mRNA amount for each sample is plotted as a function of time following transcription inhibition for all three conditions. The values for each time course are normalized to the first time point in order to compare the decay kinetics of each time course. Similar profiles are observed for all three conditions, still some difference are observed which probably result from inaccuracies in the evaluation of spike intensities.

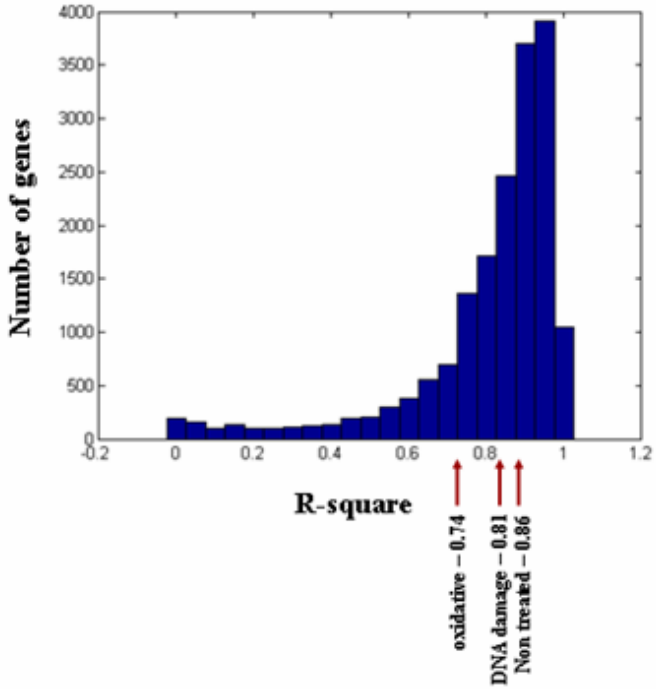

Figure S7: **Most genes decay in a constant rate during the measured time course:** Histogram of R-square ($1-\frac{\sum_i (y_i - f(x_i))^2}{\sum_i (y_i - \hat{y})^2}$) values for the fit to a constant rate decay model for all genes in all three conditions together. The values represent the goodness of fit showing that most genes display constant rate decay kinetics. The mean R-square for each condition is marked below the histogram. The differences in mean values result from the differences in the accuracy of evaluation of spike intensities as apparent from figure S6.

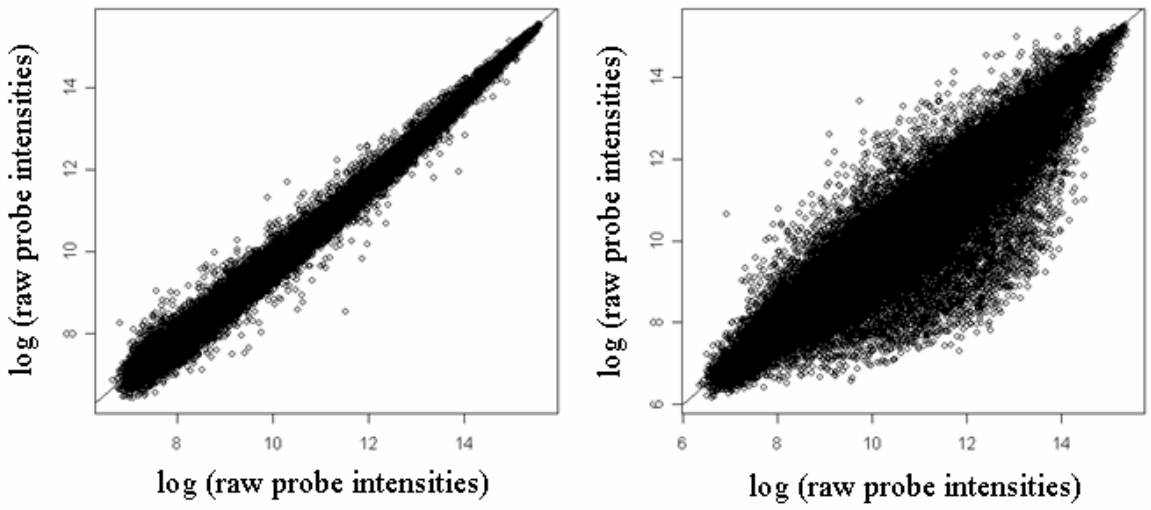

Figure S8: **Dot plot of replicate arrays and consecutive arrays:** Dot plot of raw probe intensities, left: correlation between replicate arrays, right: correlation between two different time points (time point zero and time point 20 minutes).

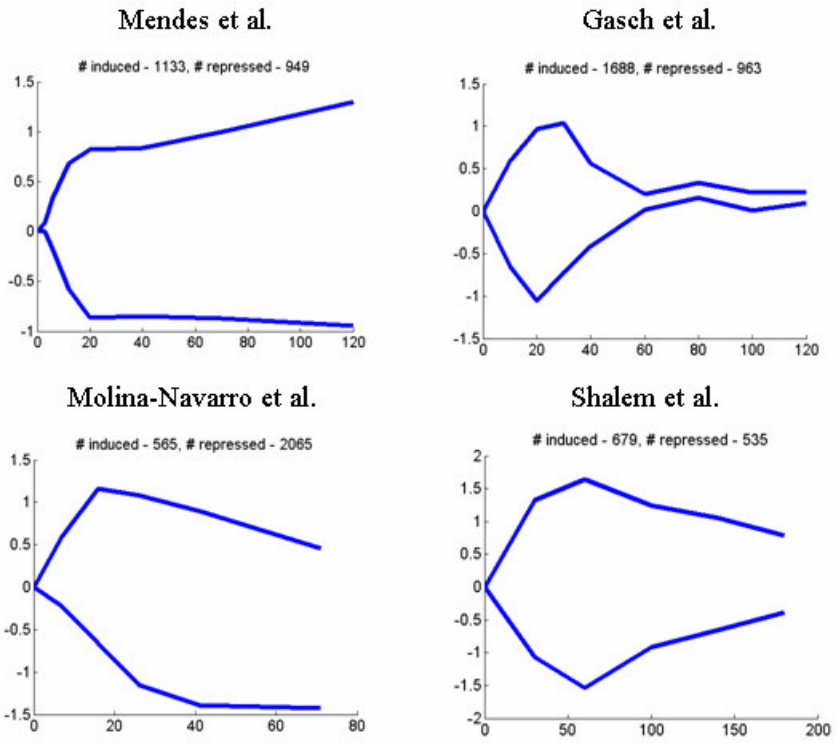

Figure S9: **Mean change in mRNA abundance following oxidative stress:** For each on the four works that measured changes in mRNA abundance following oxidative stress the mean profile of all responsive genes (that respond at least two fold relative to the reference time point) is plotted as a function of time, separately for induced and repressed genes.

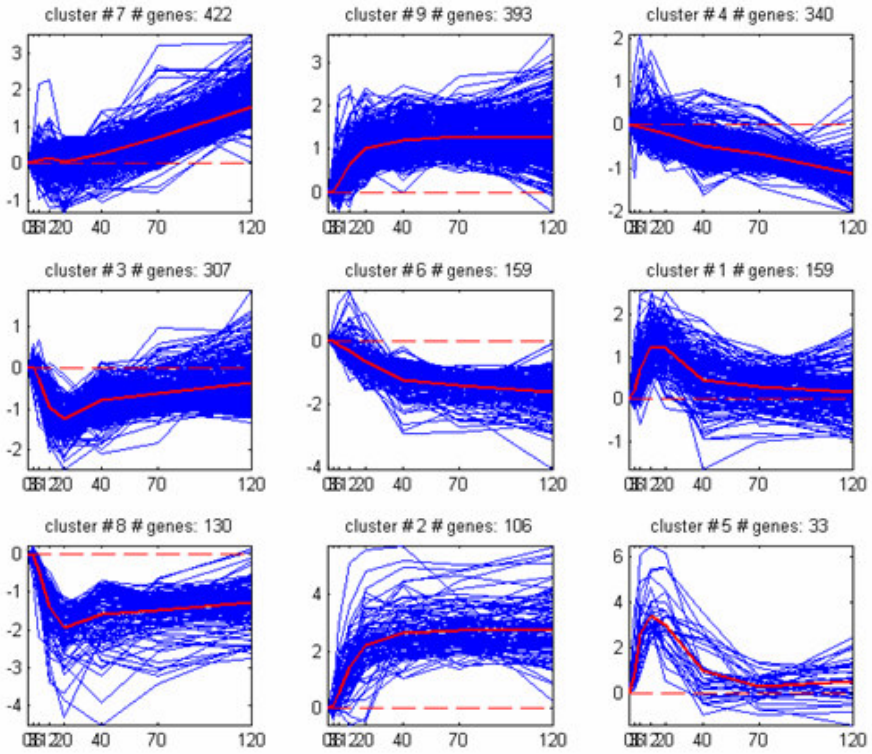

Figure S10: **Clustering of mRNA abundance changes following oxidative stress (data of Mendes et al.):** Responsive gene profiles of change in mRNA abundance following oxidative stress in the data of Mendes et al. are clustered to show two types of behaviors. Clustering was performed using k-means with k = 9. Some clusters display a highly transient response (clusters 6 and 9) while other cluster show a sustained behavior.

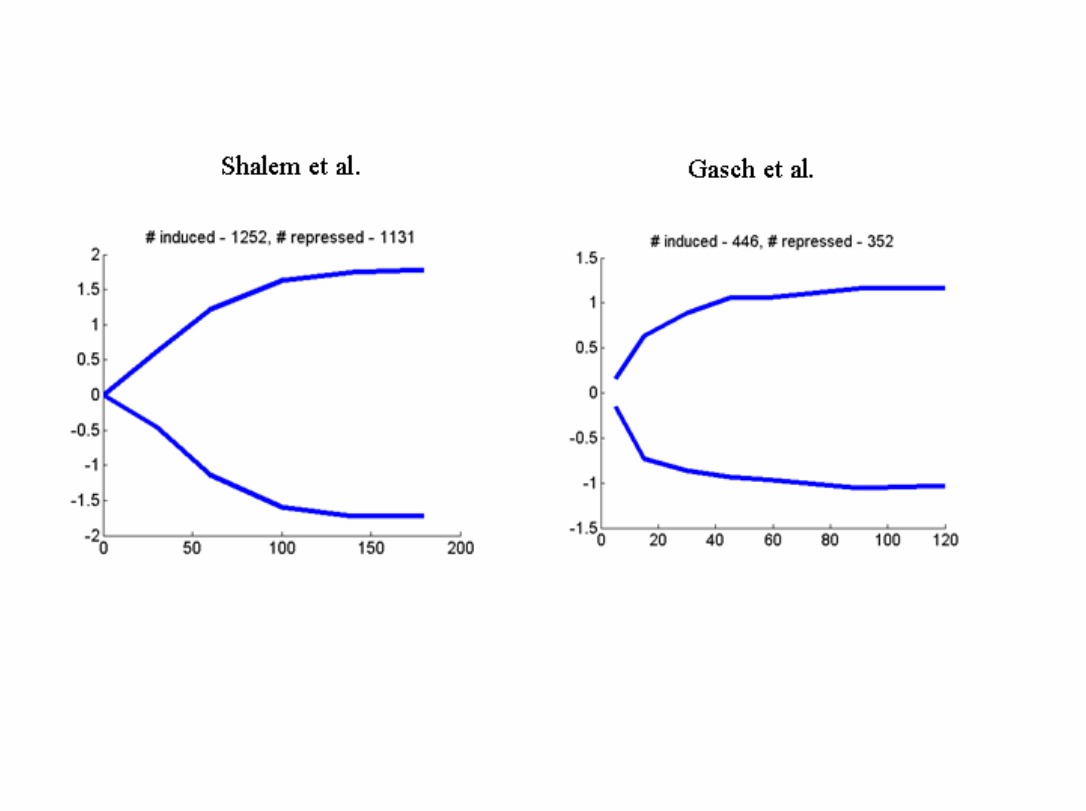

Figure S11: **Mean change in mRNA abundance following DNA damage:** The mean profile of all responsive genes (that respond at least two fold relative to the reference time point) is plotted as a function of time, separately for induced and repressed genes. Our data is compared to the data of Gasch et al. showing that in both cases a sustained response in observed.

Figure S12 **(A)**:

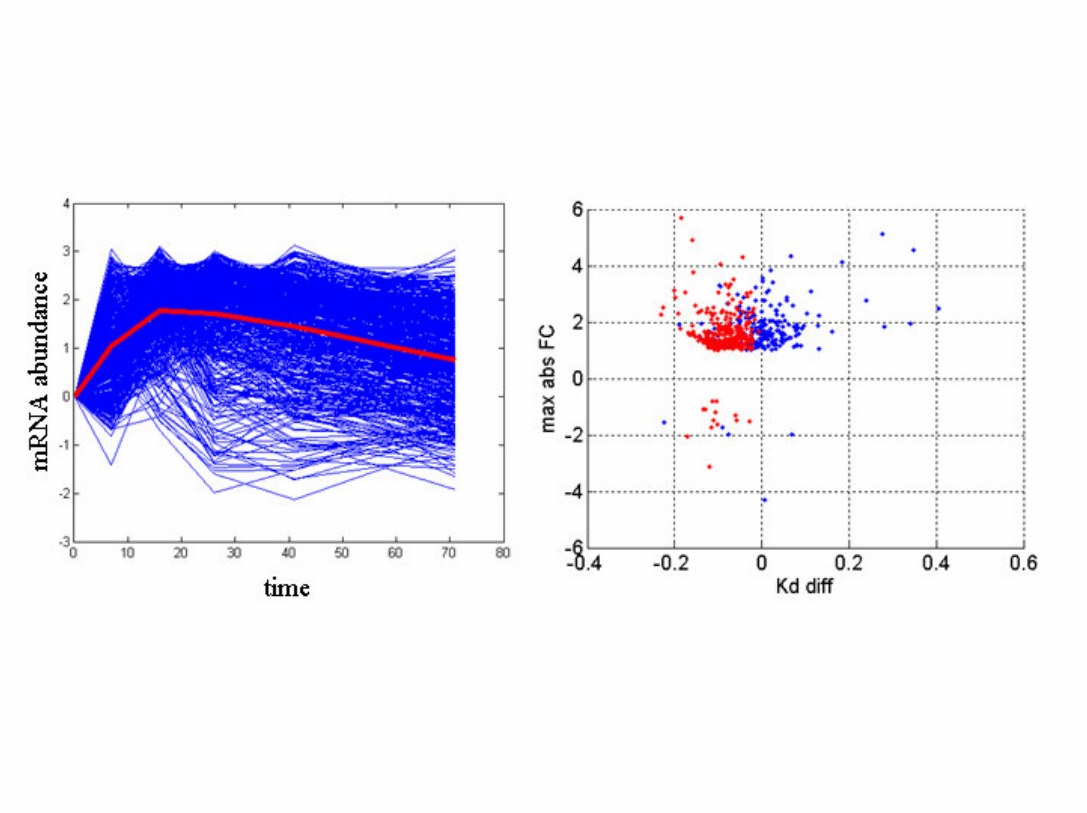

Figure S12 **(B)**:

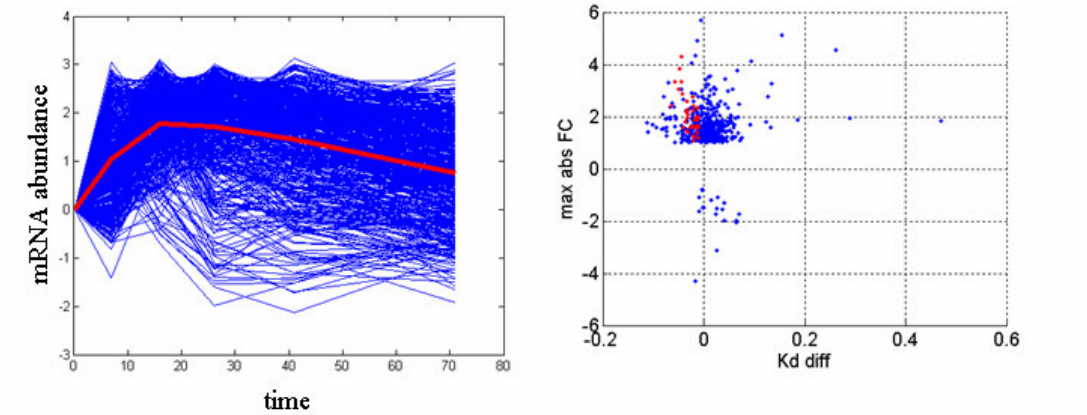

Figure S12 **(C)**:

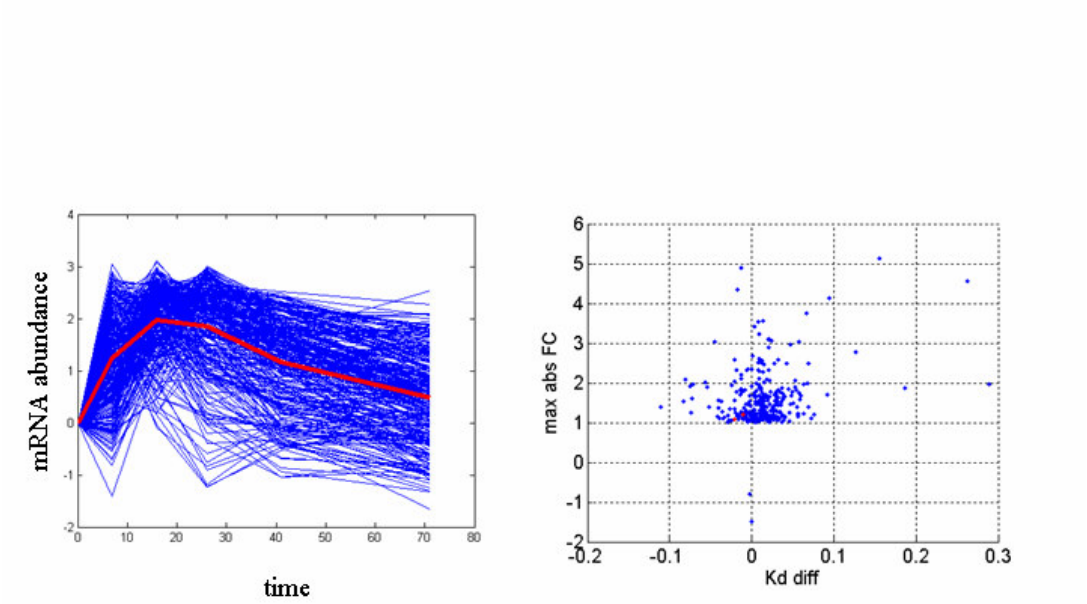

Figure S12 **(D)**:

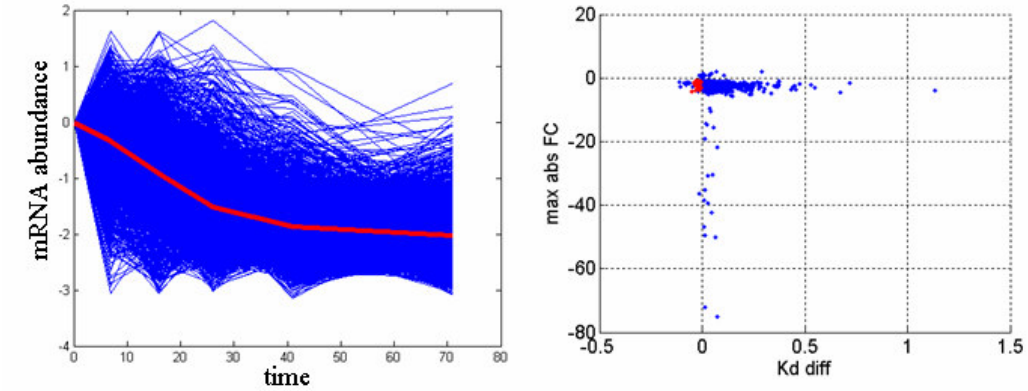

Figure S12: **Change in mRNA abundance Vs. change in Kd values (Molinha-Navarro et al. data): (A)** mRNA abundance profiles (standard deviation and mean normalized) of all induced genes are plotted alongside with the relationship between the maximal fold change to the difference in Kd in time point 7 compared to time point 1. The Kd difference is calculated by subtracting the Kd value at time point 1 from the Kd value at time point 7. Red points represent Kd differences for which the Kd values at time point 7 were negative. Although at this point most induced genes show a tendency towards stabilization (decrease in Kd) it is based mostly on non-physical negative Kd values. **(B)** Same plot as A only for time point 41, induced genes show a general tendency toward destabilization (increase in Kd). **(C)** Same plot as B, yet here only the most transient genes are taken (reach their peak before 30 minutes), the destabilization is enhanced as expected by our results. **(D)** Same plot only for repressed genes, repressed genes show a general destabilization alongside with sustained repression.

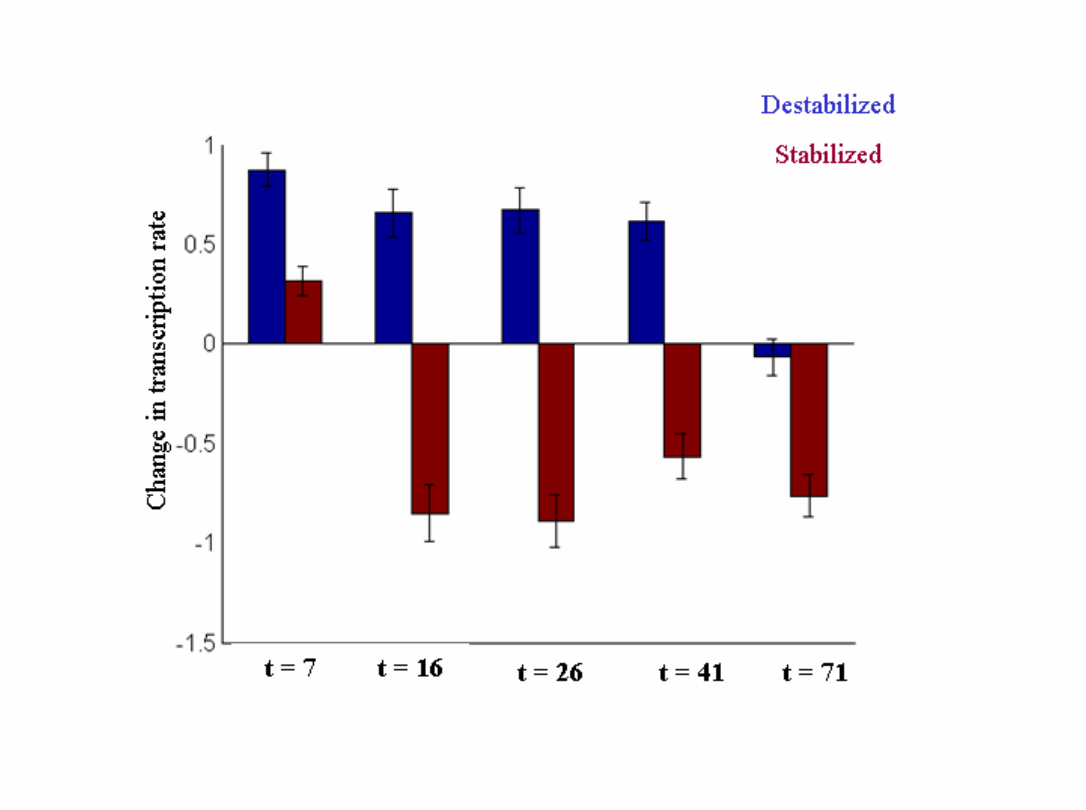

Figure S13: **Changes in transcription rates for stabilized and destabilized genes:** Bar plot of changes in transcription rates at each time point for stabilized and destabilized genes (log2 of half life ratio of above 0.8 or below -0.8), mRNA abundance and stability values are taken from our data while changes in transcription rates are taken from Molinha-navarro et al. Only responsive genes (above/below two fold change in mRNA abundance) are taken for the analysis. Destabilized genes have increased transcription rates while stabilized genes have decreased transcription rates as expected from our data. Bars represent standard error. The units of the change in transcription rates are log2 of transcription rate (TR) at each time point divided by TR measurement at time point 1min.

Figure S14 **(A)**

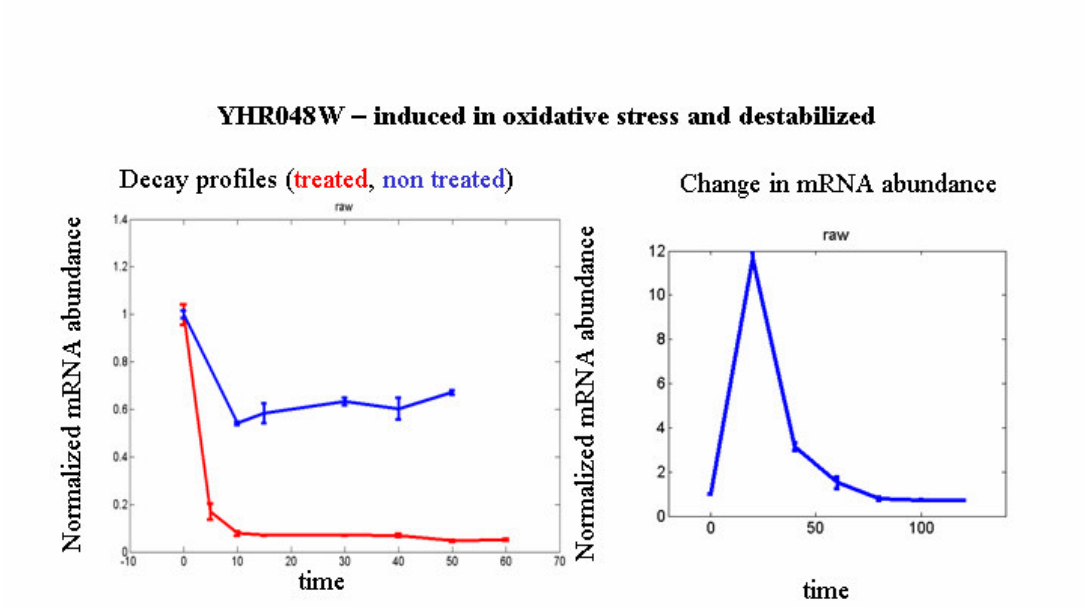

Figure S14 **(B)**

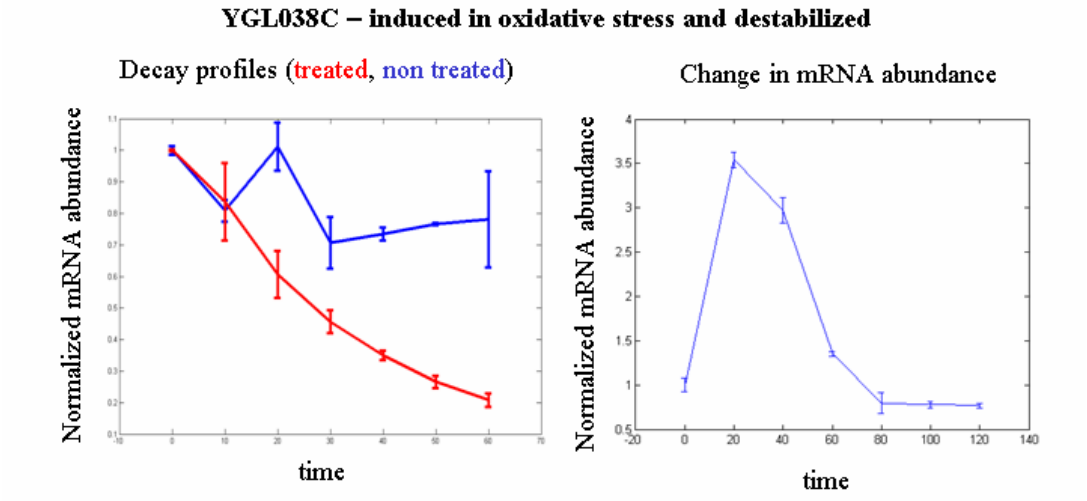

Figure S14 **(C)**

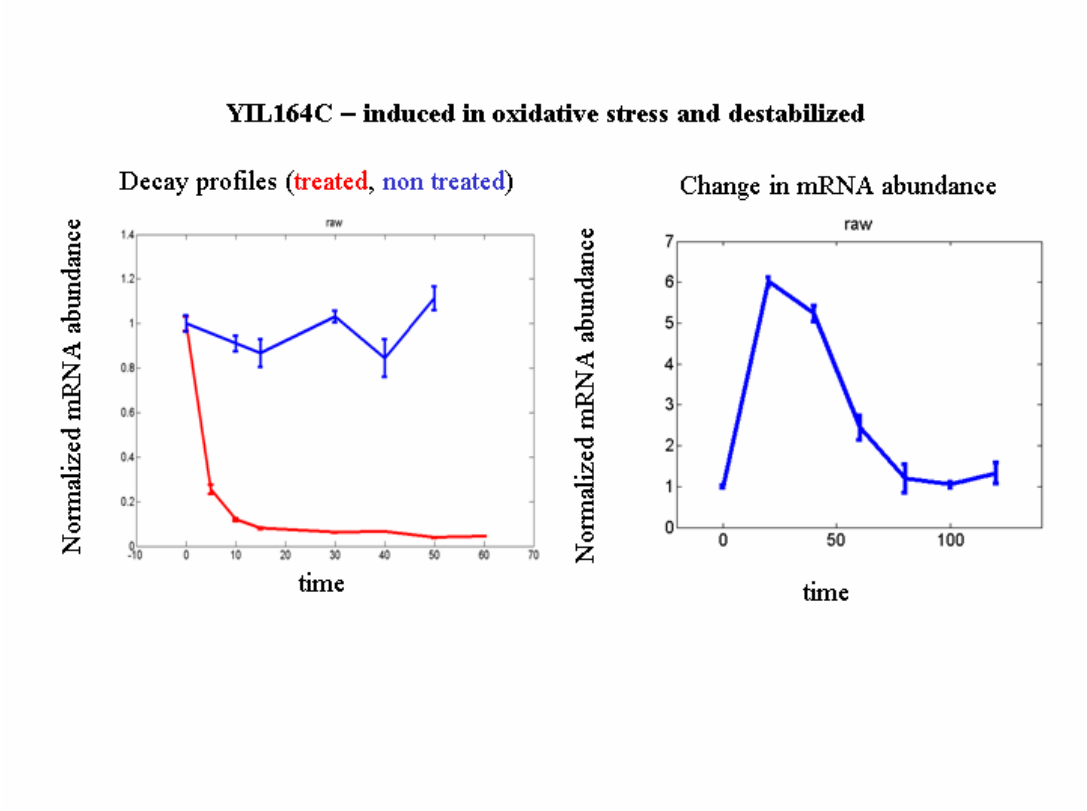

Figure S14 **(D)**

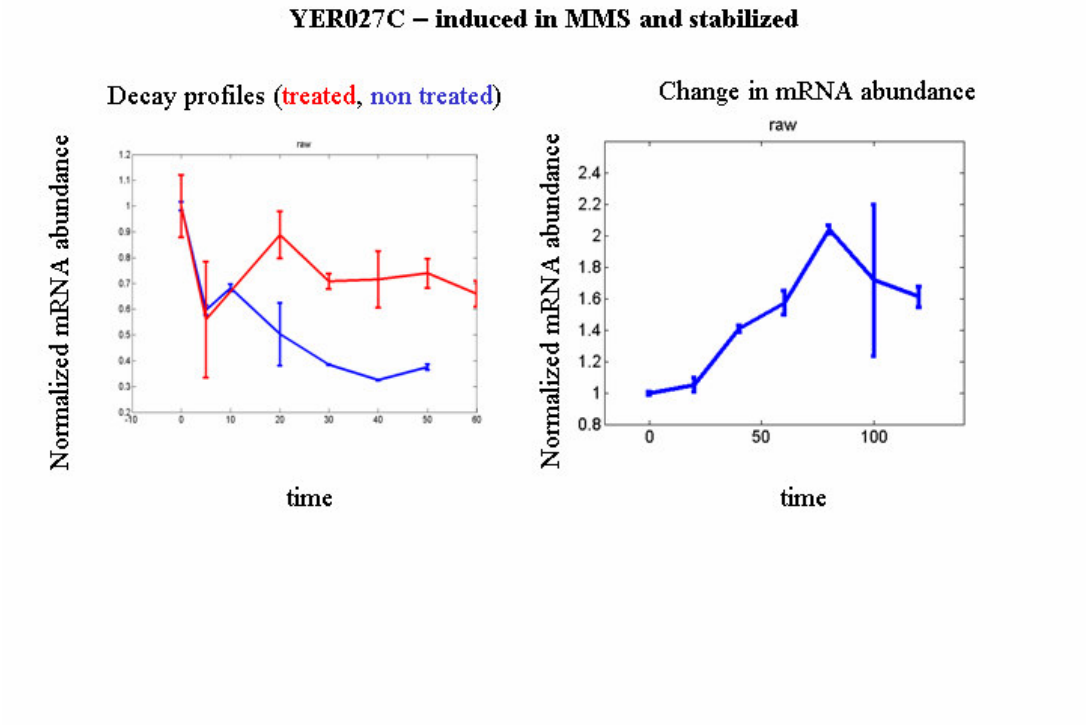

Figure S14: **Validation of microarray results using Real Time PCR for a few specific genes:** Real time PCR was used to validate microarray results, both mRNA abundance and decay profiles, for a few candidate genes (see Materials and Methods).

7. References

1. Wang, Y., et al., *Precision and functional specificity in mRNA decay.* Proc Natl Acad Sci U S A, 2002. **99**(9): p. 5860-5.
2. Bolstad, B., *Low Level Analysis of High-density Oligonucleotide Array Data: Background, Normalization and Summarization.* Dissertation. University of California, Berkeley., 2004.
3. Gasch, A.P., et al., *Genomic expression programs in the response of yeast cells to environmental changes.* Mol Biol Cell, 2000. **11**(12): p. 4241-57.
4. Sha W, M.A., Laubenbacher R, Mendes P, Shulaev V, *Expression data for Saccharomyces cerevisiae oxidative stress response.* GEO, 2007.
5. Molina-Navarro, M.M., et al., *Comprehensive transcriptional analysis of the oxidative response in yeast.* J Biol Chem, 2008. **283**(26): p. 17908-18.
6. Gasch, A.P., et al., *Genomic expression responses to DNA-damaging agents and the regulatory role of the yeast ATR homolog Mec1p.* Mol Biol Cell, 2001. **12**(10): p. 2987-3003.