עבודת גמר (תזה) לתואר
דוקטור לפילוסופיה

מוגשת למועצה המדעית של
מכון ויצמן למדע
רחובות, ישראל

Thesis for the degree
Doctor of Philosophy

Submitted to the Scientific Council of the
Weizmann Institute of Science
Rehovot, Israel

מאת
אופיר שלם

By
Ophir Shalem

בקרת בתר שעתוק והאינטראקציה שלה עם מנגנוני שעתוק העולה מתוך מחקר גנומי וגנטיקה סינטטית

Post transcriptional regulation and its interaction with transcription revealed by genome-wide and synthetic gene approaches

מנחה:
פרופ' יצחק פלפל

Advisor:
Prof. Yitzhak Pilpel

ט"ז בכסלו, תשע"ב

December 12, 2011

Table of Contents

Abstract

Gene expression is a complex process composed of multiple consecutive steps encoded in distinct regulatory regions. Here,we study this complexity by dissecting it to its basic determinants,both at the level of the different layers and of different regulatory regions using the yeast *S. cerevisiae*as a model system. We use direct genome-wide measurements of mRNA stability and abundance in different conditions to expose the dynamic properties of mRNA stability.Specifically, we show mRNA production and degradation are coordinately regulatedto determine the temporal kinetics of the transcriptome in response to stress. We propose a model for this coordination, based on a previously discovered molecular mechanism involving co-transcriptional imprinting of mRNA molecules by RNA polymerase II.

At the level of different regulatory regions, we study the independent effect of 3' UTR sequences on protein expression. For this we construct a synthetic yeast reporter library composed of strains differing by a unique 3'UTR sequence integrated downstream to a florescence gene driven by a constant promoter. We find that each 3' UTR sequence displays a unique and reproducible effect on protein expression, which spans a continuous range of more than 10-fold. We find that increased A/T content upstream of the polyadenylation site (3' UTR end) correlates with higher expression, highlighting the importance of this genomic region and suggesting that differences in transcription termination efficiencies might be partly responsible for the observed effect. We also use single cell measurements to show that 3' UTR sequences change protein levels with distinct dynamics compared to changes mediated by promoter activation levels putting forth a unique additional level by which gene expression might be controlled.

סיכום

תהליך הביטוי הגנטי מורכב משלבים עוקבים רבים המקודדים באזורים שונים בגנום. בעבודה זאת ניסינו לחקור את מורכבות התהליך הנ"ל על ידי פירוקו למרכיביו הבסיסים הן ברמת התהליכים עצמם והן ברמת האזורים המקודדים בגנום תוך כדי שימוש בשמר ההנצה כאורגניזם מודל. בעזרת שימוש במדידות קצבי הפרוק של מולקולות רנ"א שליח בתנאי סביבה שונים חשפנו את האופי הדינמי של תהליך זה. בפרט, מצאנו שקיים צימוד בין השינויים בקצבי הפרוק לשינויים בקצבי ייצור, תכונה המאפשרת לתאים לבקר את הפרופיל הקינטי של השינויים ברמות המולקולות רנ"א שליח כתלות בזמן. אנו מציעים מודל מולקולרי המבוסס על מנגנון שתואר בספרות עבור גנים בודדים המשלב שליטה על קצבי פרוק המבוקרת כבר במהלך תהליך השעתוק על ידי תתי יחידות של רנ"א פולימארז.

ברמת האזורים הגנומיים בצעתי מחקר שהתמקד בקצה ה3' הלא מתורגם (3'UTR) ובהשפעת אזורים אלו על רמות ביטוי חלבון. לצורך משימה זאת בניתי ספרייה של זני שמר הנבדלים ברצף 3'UTR המהונדס לתוך הגנום מיד לאחר גן דווח פלורסנטי המאפשר מעקב פשוט על רמות החלבון. מצאתי שכל אחד מהרצפים שנבדקו גרם להשפעה ייחודית על רמות החלבון כך שהספריה כולה פורשת טווח דינמי של סדר גודל ומעלה. מצאתי שרצפים עשירים בAT מיד לפני סוף ה3'UTR מראים קשר חיובי עם רמות החלבון. ממצא זה מדגיש את חשיבות האזור הגנומי הזה ומציע שחלק מההבדל בין הזנים נובע מהבדלים ביעילות סיום שעתוק. בנוסף, בעזרת מדידות של רמות החלבון בספרייה בתאים בודדים מצאתי שהשנויים המתווכים על ידי רצפי 3'UTR מראים דינאמיקה שונה משינויים המתווכים על ידי הבדלים ברמות בקרי שעתוק. ממצאים אלו מציעים את רצפי 3'UTR כשכבת בקרה ייחודית של רמות חלבון.

Introduction

Research outline and list of related publications

My research aims were to deepen the understanding of the basic determinants of gene expression by studying the independent effects bothof different regulatory layers, e.g. transcription and mRNA degradation, and of different regulatory regionssuch as 3'UTRs and promoters. Most of my work lied within two main lines of research: The first examined the interplay between two layers in the process of expression, transcription and mRNA degradation, using direct measurements of mRNA decay and abundance in different conditions. The second aimed to study the independent contribution of 3'UTR sequences on protein expression byconstructing yeast synthetic reporter strains differing by native 3'UTR sequences integrated downstream to a florescence reporter gene. Most of the work is described in detail in the following publications appearing in the results section of this thesis:

- The effect of 3'UTR sequences on protein expression in Yeast. **Shalem O**, Pilpel Y and Segal E. *Submitted*
- Transcriptome kinetics is governed by a genome-wide coupling of mRNA production and degradation: a role for RNA polII. **Shalem O**, Groisman B, Choder M, Dahan O and Pilpel Y. 2011 *PLoS Genetics*
- Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation. **Shalem O**, Dahan O, Levo M, Martinez MR, Furman I, Segal E, Pilpel Y. *MolSyst Biol.* 2008;4:223. Epub 2008 Oct 14.

Additional publications I contributed to:

- Axonal Transcription Factors Signal Retrogradely In Lesioned Peripheral Nerve. Ben-Yaakov K, Dagan S, Segal-Ruder Y, **Shalem O**, VuppalanchiD,Willis D, Michaelevski I, Rishal I, Blesch A, Pilpel Y, Twiss, Fainzilber M. *under revision*
- Signaling to transcription networks in the neuronal retrograde injury response. Michaelevski I, Segal-Ruder Y, Rozenbaum M, Medzihradszky KF, **Shalem O**, Coppola G, Horn-Saban S, Ben-Yaakov K, Dagan SY, Rishal I, Geschwind DH, Pilpel Y, Burlingame AL, Fainzilber M. *Sci Signal.* 2010 Jul 13;3(130):ra53.

Primer on mRNA decay in yeast

Mature eukaryotic mRNAs are created and exported into the cytoplasm with two integral stability determinants: the 5' cap structure and the poly(A) tail. Several pathways exist by which these structures are removed and the mRNA is degraded: deadenylation-dependent mRNA decay, deadenylation-independent decapping, endonuclease-mediated mRNA decay and mRNA surveillance pathways[1]. miRNA-mediated mRNA decay also exists, although it lacks from the yeast *S. cerevisiae*, the model organism used in this work[2]. In eukaryotes and especially in yeast, most mRNAs decay by the deadenylation-dependent mRNA decay pathway, and only a few mRNAs were shown to bypass this route and decay by other pathways[1]. The rates of deadenylation are highly correlated with the decay rates of total mRNA and this step is thought to be the rate limiting step of the decay process[3]. The first step in deadenylation-dependent decay is the shortening of the poly(A) tail which can be catalyzed by several enzymes (CCR4-NOT is the main deadenylasecomplex in S.cerevisiae). Once the poly(A) tail is shortened below a certain length, the mRNA molecule is exonucleolitically degraded either from the 5' or 3' end.mRNA decay is carried out in distinct cytoplasmic compartments (P-bodies)[4]: These sites have been characterized in both human and yeast and the enzymes known to be associated with the process of mRNA decay were shown to localize to these compartments together with mRNA decay intermediate products. Once the mRNA is exported to the cytoplasm the poly(A) tail and 5' cap structure are bound by the cytoplasmic proteins poly(A)-binding protein (PABP) and eIF4E respectively. During translation the mRNA is thought to be circularized by the interaction of these proteins and other translation and stability protein factors. This interaction enhances translation and prevents degradation by protecting the poly(A) tail from deadenylation[5].

Genome wide measurements of mRNA stability in permissive conditions were previously performed in several organisms[3,6–8]. Such experiments revealed that genes encoding for proteins that act together in stoichiometric complexes, or

share a physiological function, tend to decay in similar rates proposing that mRNA decay is tightly regulated and plays a functional role.

Coordinated regulation of mRNA production and decay

Traditionally, the process of gene expression was viewed as a series of independent steps, each controlled by separate regulatory mechanisms. Nowadays it is becoming clear that the different regulatory layers are coordinately regulated[9]. Coordinated changesin distinct expression layers, e.g. transcriptionand translation, can be achieved by one of two general means: either a shared regulatory cascade controls the two process in parallel, or a direct mechanistic coupling by which consecutive layers share regulatory factors such that an effect in one is directly propagated to the next level. Specifically, for the case of transcription and mRNA decay, two separateprotein complexes were shown to affect both of these two opposing processes. The first is the CCR4-NOT complex, a major yeast deadenylase that controls the initial steps of mRNA degradation[10]. In addition to its major role in mRNA decay it was recently shown to interact with proteinsinvolved in transcriptioninitiation[11–14].Additional evidence for direct mechanistic coupling between mRNA production and degradationcomes from work on the RNA Polymerase II itself. Detailed biochemical work, on specific genes in the yeast *S. cerevisiae*suggest that Rpb4 and Rpb7, two subunit of RNA Pol II, play a role not only in transcription but also in the later stages of gene expression as mRNA degradation and translation[15–19]. This effect is achieved when an mRNA molecule is co-transcriptionally imprinted, by the binding of these two subunits. This mRNP complex is then exported into the cytoplasm and the binding of the two subunits affects both the rate of degradation and translation of the bound mRNA[17,18,20]. The idea introduces a new concept in gene expression regulation by which the post-transcriptional fate of an mRNA molecule is allready determined at the initial steps of transcription[21]. Yet, this mechanism was figured out by investigation at the single gene level, and how relevant it is to the entire genome level was left to be explored. Further, it was not known what might be the physiological effect of such coupling, or at least how might it affect the kinetics of gene expression.

I have initially set to study the dynamical properties of mRNA decay by measuring genome-wide mRNA stability at different growth conditions. By genome-wide investigation carried out in the yeast *S. cerevisiae*I found, that in addition to extensive, condition specific modulation of mRNA stability, changes in transcription and changes in mRNA degradation are frequently coordinated, an interaction which shapes the temporal kinetics of mRNA abundance in response to external stimuli[22]. These findings included a mode of coupling in which fast responding genes, which exhibit a transient temporal behavior, feature an increase in both the rates of production and degradation. Following our publication several additional works reported in the literature a similar coupling mode that was often assumed to be responsible for spiked kinetic profiles in response to environmental changes[23–27].

In order to uncover a potential mechanistic basis for such coupling I extended this work to propose a molecular model that governs this coordination based on the previously discovered RNA Pol II mRNA imprinting mechanism described above. We used a yeast mutant strain, which enabled the de-coupling of Rpb4/7 mediated effect on production and degradation,and show that this mechanism [28,29] can account for the counter-action mode of coupling, and its effect on the kinetics of mRNA abundance,which we previously observed. To explain these results we propose a simple model in which changes in the imprinting probability in response to stress result in a genome wide redistribution of the general decay machinery in the cytoplasm enabling the cell to control transcriptome response kinetics.

Effect of 3'UTR sequences on protein expression in Yeast

In a second line of work I aimed to decompose the effect of particular DNA regulatory region – the 3' un-translated region (UTR) on gene expression by studying its independent effect on protein levels. For that I constructed a synthetic yeast reporter library, in which each strain has a unique native 3'UTR from the yeast genome integrated downstream to a yellow florescence reporter gene. Using the library I showed that each native 3'UTR tested has a unique and highly reproducible effect on protein expression spanning a continuous range of

more than 10 fold in expression values. Using flow cytometry measurements of single cells I showed that the observed effect is due to differences in expression burst size and not frequency, as would be expected from regulation at post transcriptional-initiation level. Because the differences in expression are due to the sole effect of the different 3'UTRs, using simple sequence analysis we found a sequence signature which is based on the GC content at a small 15-20bp window located at a specific location upstream to the poly adenylation site which correlates with the library YFP expression. Following the detection of this sequence signature in the synthetic context of the library I wanted to examine the effect of this sequence feature in natural genes. For that I used published RNA-seq measurements of mRNA abundance[30] to show that a similar sequence signature also shows a highly significant, although lower, correlation on a genome wide level with expression level. I thus concluded that this 3' UTR nucleotide composition signal plays a regulatory role in the native genomic context, and on a genome wide level. The proximity of this region to the poly-adenylation site, together with other accumulating molecular evidence, suggest that native 3'UTR sequences encode different transcription termination efficiencies resulting in a wide effect on protein expression. These findings thus put forth an additional layer at which gene expression might be controlled.

Methods

Complete description of the methods used is available in the three accompanying papers. Here I briefly introduce the main experimental procedures with a short description of each.

Microarray measurements of mRNA abundance and decay

DNA Microarrays is a well-established technology used to quantify genome-wide mRNA levels[31,32]. Here I use this method to quantify the levels of two cellular parameters in different conditions and strains. The first is standard measurements of mRNA abundance and the second is mRNA half-lives.

mRNAabundance measurements are preformed using standard protocols:A yeast culture is grown to mid log phase and then subjected to the desired treatment. Following that, samples are taken over a short time course and RNA is extracted using commercially available kits. RNA is then converted to cDNA using polyT primers to capture only polyadenylated mRNA which is then processed using standard Affymetrix protocols and hybridized to a DNA CHIP containing all yeast open reading frames (ORFs). Raw intensity measurements are processed using the RMA algorithm in order to achieve an expression value per gene.

Briefly, to measure mRNA half-lives, a yeast culture is grown to mid log phase when transcription is inhibited. Few culture samples are then taken during the course of an hour and mRNA abundance is quantified in each sample as described above. We have used two ways to inhibit transcription: in the first paper[22] we used a yeast mutant containing a temperature sensitive Pol II mutant which is inactivated when yeast cells are transferred to a non permissive temperature[33]. The advantage of this strain over chemical inhibition is the fact that it inhibits only Pol II transcription. In order to achieve a decay profile per gene I developed a pre-processing pipeline, which uses spiked-in RNA as a reference to recover genome-wide decay kinetics. A special pre-processingprocedure was needed here because standard techniques to process microarray experiments assume that the intensity distributionbetween samples remains constant while in our case we expect a global decrease in the genome mRNA levels due to the inhibition of transcription. Following normalization each gene's decay profile is fitted to a first order exponential decay curve to calculate a half life per gene and a goodness of fit measure. In the second paper[34] we used chemical inhibition of transcription using the drug 1, 10-phenantroline. The different methods to inhibit transcription have been compared and shown to have a similar effect on genome wide mRNA abundance[6]. Here, because not only Pol II transcription is inhibited, a different normalization scheme was applied[34].

Construction of a 3'UTR reporter strain library

Construction of a yeast 3' UTR reporter strain library followed two main steps: the first included the construction of a master strain, which was used as a template for the transformation of the individual 3' UTR strains. The second step included the creation of the library using standard transformation protocols converted to 96 well,semi-automated, format using a robotic liquid handling system. For the creation of the master strain I cloned onto a plasmid an mCherry florescence gene driven by the TEF2 promoter used as a control, YFP florescence gene driven by the Gal1/10 promoter, a truncated URA3 gene located downstream to the YFP gene and NAT yeast antibiotic resistance gene. The complete construct was amplified using PCR from the plasmid and genomically integrated into the Y8205 strain using homologues recombination to the HIS deletion locus.

Following master strain construction a library of downstream intergenic regions were amplified from the genomeby PCR. Each PCR product was extended to contain the URA3 promoter and start codon, downstream to the tested genomic region, enabling the integration of each constructdownstream to the YFP gene in the master strain. We converted standard yeast transformation protocols to 96 well plates to create a library of 87 yeast strains each differing by a 3' UTR sequence downstream to the YFP reporter gene. Strains in the library were then validated by several means: colony PCR, clone comparison of YFP levels, selective sequencing and comparison of mCherry and OD to detect global deficiencies in growth, transcription or translation. Library strains were arranged in a 96 well plate together with a few control strains.

Florescence measurements and analysis

Once the library was constructed we measured YFP, mCherry and optical density (OD) both in batch and in single cells.All measurements were done in 96 well plates. For bulk measurements a small volume of stationary culture was inoculated into fresh media containing Raffinose and galactose to induce the promoter. Measurements were carried out every ~20 minutes using a robotic system (Tecan Freedom EVO) with a plate reader (Tecan Infinite F500). Each measurement included OD, YFP and mCherry fluorescence. Raw data was then

processed by an automatic pipeline, which integrated YFP and OD measurements into one measure per strain of YFP production per cell per time unit.

To calculate the mean and standard deviation of YFP expression from an isogenic cell population for each 3' UTR strain we used an LSRII flow cytometry machine supplements with an High Throughput Sampler (HTS) to measure 96 well plates. Raw data was then subjected to an automatic gating procedure which controls for extrinsic variation by selecting a population of cells with a narrow range of physiological parameters (as captured by the forward and side scatter in a FACS machine). Mean and standard deviation of the gated population is then used in order to calculate burst size and frequency parameters according to an analytical framework previously proposed[35,36]. Briefly, the distribution of protein expression values *x* is modeled using a gamma distribution:

$$p(x) = \frac{1}{b^a \Gamma(a)} x^{a-1} e^{-x/b}$$

Then the burst frequency and size (a and b respectively) relate to the mean and standard deviation as follows:

$$\frac{\sigma^2}{< x >} = \frac{ab^2}{ab} = b \frac{\sigma^2}{< x >^2} = \frac{ab^2}{a^2 b^2} = \frac{1}{a}$$

Results

Molecular Systems Biology 4; Article number 223; doi:10.1038/msb.2008.59
Citation: ***Molecular Systems Biology*** **4:223**

www.molecularsystemsbiology.com

molecular systems biology

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

Ophir Shalem[1,2], Orna Dahan[1], Michal Levo[1], Maria Rodriguez Martinez[1,3], Itay Furman[1], Eran Segal[2,*] and Yitzhak Pilpel[1,*]

[1] Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel, [2] Department of Applied Mathematics and Computer Science, Weizmann Institute of Science, Rehovot, Israel and [3] Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot, Israel
* Corresponding authors. Eran Segal, Department of Applied Mathematics and Computer Science, Weizmann Institute of Science, Herzel, Rehovot 76100, Israel Tel.: +972 8 934 4282; Fax: +972 8 934 4122; E-mail: Eran.Segal@weizmann.ac.il or Yitzhak Pilpe, Department of Molecular Genetics, Weizmann Institute of Science, Herzel, Rehovot 76100, Israel. Tel.: +972 8 934 6058; Fax: +972 8 934 4108; E-mail: pilpel@weizmann.ac.il

Received 14.5.08; accepted 14.9.08

The state of the transcriptome reflects a balance between mRNA production and degradation. Yet how these two regulatory arms interact in shaping the kinetics of the transcriptome in response to environmental changes is not known. We subjected yeast to two stresses, one that induces a fast and transient response, and another that triggers a slow enduring response. We then used microarrays following transcriptional arrest to measure genome-wide decay profiles under each condition. We found condition-specific changes in mRNA decay rates and coordination between mRNA production and degradation. In the transient response, most induced genes were surprisingly destabilized, whereas repressed genes were somewhat stabilized, exhibiting counteraction between production and degradation. This strategy can reconcile high steady-state level with short response time among induced genes. In contrast, the stress that induces the slow response displays the more expected behavior, whereby most induced genes are stabilized, and repressed genes are destabilized. Our results show genome-wide interplay between mRNA production and degradation, and that alternative modes of such interplay determine the kinetics of the transcriptome in response to stress.
Molecular Systems Biology 14 October 2008; doi:10.1038/msb.2008.59
Subject Categories: chromatin & transcription; RNA
Keywords: degradation; microarray; stress; transcription; yeast

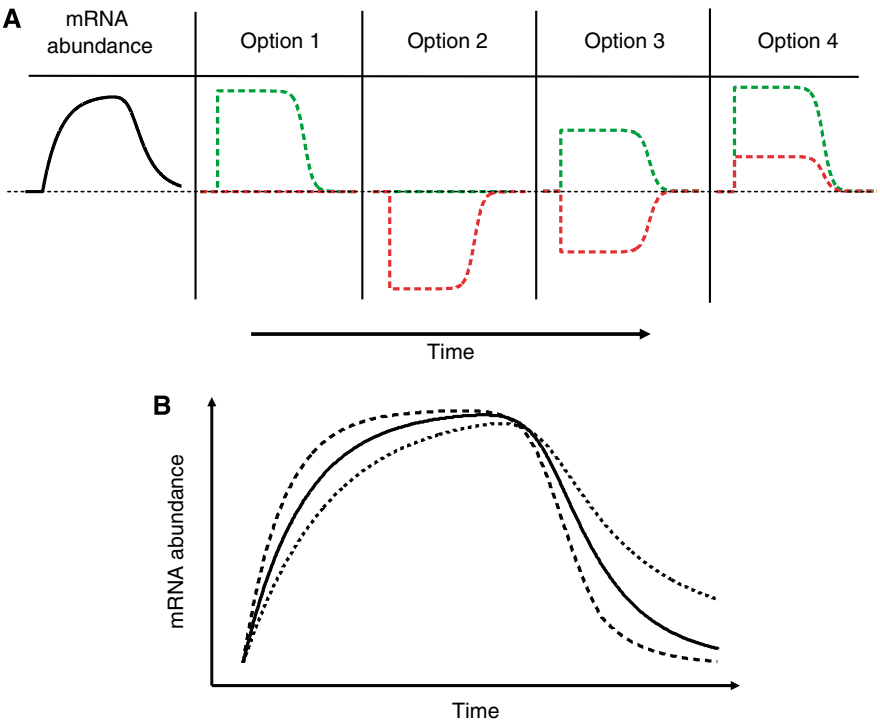
Introduction

In response to environmental stimuli, the mRNA abundance of a large fraction of the genome changes either by increasing or decreasing its levels (Gasch *et al*, 2000, 2001; Jelinsky *et al*, 2000; Gasch and Werner-Washburne, 2002). Clearly, to understand the state of the transcriptome under varying conditions, the role of both mRNA production and degradation must be examined. An increase in mRNA abundance in response to a stimulus may be achieved either by increasing the rate of transcription or by decreasing the rate of degradation. Likewise, a decrease in the transcript level can be achieved either by an increase in the rate of degradation or a decrease in the rate of production. More complex interplays between production and degradation are also possible. For instance, an increase in mRNA production rate might be accompanied by a decrease in degradation rate, leading to mRNA accumulation. Perhaps less intuitive is the possibility that an increase in transcript levels would be obtained by increasing both production and degradation rates, provided that the extent of production increase exceeds the elevation in the degradation rate (Box 1A). Whereas steady-state levels are simply determined by the ratio of production and degradation rates, the kinetic behavior is expected to be more complex, which is dependent on the actual rates, and hence different under the above regimens (Box 1B).

As a result, one may expect complex regimens of interplay between transcription induction and repression, and stabilization versus destabilization of mRNAs that will result in various effects on response kinetics (Perez-Ortin *et al*, 2007). Yet customary transcript abundance measurements, e.g. with microarrays, provide only the net values and do not provide information regarding the relative contribution of mRNA production and degradation.

Although still scarce when compared to transcription, the attention directed toward the control of mRNA degradation has

Box 1 Various potential relationships between mRNA production and degradation can produce similar mRNA abundance profiles but with distinct kinetics:



(**A**) A typical mRNA abundance profile plotted on the left (black) can be produced by utilizing changes in production and degradation in different ways. As an example, several options are plotted where the green and red lines represent the change in mRNA production and degradation, respectively, as a function of time. The first two options show that, at least hypothetically, either production or degradation alone can be modulated to produce an induction followed by relaxation of the mRNA abundance. The last two options show that a similar profile of mRNA abundance can be produced by changing both production and degradation either by counteracting each other or by working in the same direction. (**B**) Although induction can be achieved either by changes in production or degradation, the kinetics of induction and relaxation would depend on the levels of degradation. Following a simple kinetic model for changes in mRNA levels, which assumes zero-order production and first-order degradation: $dx/dt=\beta-\alpha x$, where x, β and α represent mRNA abundance, production rate and degradation rate, respectively, the change in mRNA abundance follows $\Delta x(t)=((\beta/\alpha)-x(0))(1-e^{-\alpha t})$, where $x(0)$ marks the level of x prior to induction. Thus, the response time, defined as the time at which half of the new steady-state level with respect to the former is achieved, is solely determined by the degradation rate, and is inversely proportional to it: $t_{1/2}=\ln(2)/\alpha$. Thus, increased decay rate is expected to result in a shorter response time. A reference mRNA abundance profile is plotted in a solid line, two profiles with increased and decreased decay rates are plotted in dashed and dotted lines, respectively. Production rates are changed accordingly to keep the same steady-state level. Increasing the decay rate results in faster kinetics both in the induction and relaxation phases, whereas a decrease in decay rate results in slower kinetics. The curves shown here were generated by solving the above differential equation. The dashed line thus represents an interesting case in which both production and degradation rates have increased, reconciling higher steady-state level with fast response time.

increased in recent years. Genome-wide measurements of mRNA half-lives have been conducted in several organisms, revealing interesting relationships between functional properties of genes and their decay rates (Bernstein *et al*, 2002; Wang *et al*, 2002; Yang *et al*, 2003; Grigull *et al*, 2004; Narsai *et al*, 2007). Genomic run-on methods were used to compare transcription rates to changes in mRNA abundance and concluded that changes in transcription rates are not sufficient to explain changes in mRNA abundances observed in response to stress (Fan *et al*, 2002; Garcia-Martinez *et al*, 2004). The regulation of mRNA stability by cis and trans-acting factors has also been investigated, both by extensive studies of specific examples (Xu *et al*, 1997; Wilusz *et al*, 2001; Parker and Song, 2004; Wilusz and Wilusz, 2004; Garneau *et al*, 2007) and also by genome-wide computational studies aiming to detect sequence features in 3′UTRs that are predicted to affect mRNA stability (Foat *et al*, 2005; Shalgi *et al*, 2005).

Despite this recent progress, the relationship between the control of mRNA production and degradation is poorly understood. Here, we investigate this interplay and study its effect on key kinetic parameters of the mRNA response to stress in the yeast *Saccharomyces cerevisiae*. We studied two intensively investigated conditions (Gasch *et al*, 2000, 2001; Jelinsky *et al*, 2000; Workman *et al*, 2006), an oxidative and a DNA-damaging stress selected such that they will differ in the kinetics of the mRNA response they induce. In oxidative stress, the majority of the responding genes show fast response followed by relaxation, whereas in the DNA damage experiment, the response is slow and long enduring. We found that changes in degradation counteract the changes in mRNA abundance for most of the genes under oxidative stress, whereas the opposite coupling, namely stabilization of induced genes, and destabilization of repressed genes, is observed for the majority of the transcripts during the

response to DNA damage. We show that the alternative types of interplays between mRNA production and degradation in each experiment shape the kinetic response of the transcriptome.

Results

The experimental design

We have chosen two environmental conditions, exposure to hydrogen peroxide (H_2O_2), which induces an oxidative stress, and exposure to methyl methanesulfonate (MMS), which induces DNA damage. We followed genome-wide response to these stresses using conventional microarray measurements of mRNA abundance (see Materials and methods). Under current parameters of each stress (see Materials and methods), the two conditions gave rise to distinct mRNA response kinetics. In oxidative stress, the majority of the responding genes showed, in agreement with previous measurements (Gasch *et al*, 2000), a fast transient response (see Supplementary information and Supplementary Figures S9–S11 for comparison with related experiments). In contrast, in the MMS experiment, the majority of the responding genes showed long enduring response with no relaxation within the measured time course, also in agreement with previous studies (Gasch *et al*, 2001) (Figure 1A). To decipher the ways in which modulation of mRNA decay rates is utilized by the cell to determine such different kinetic abundance profiles, we also performed in parallel mRNA decay experiments. In these experiments, the same stresses were applied, yet following the stress, we halted transcription in the cells and used microarrays to measure the decay kinetics of each gene under every condition. We realized that potential regulated change in mRNA stability in response to each stress may require active transcription of stability affecting genes. Thus, to build up their degradation control network, cells should be given a time interval between introduction of the stress and the transcriptional arrest. We therefore apply the stress and halt transcription after giving the cells a time interval to respond (see Materials and methods for details). We also carried out a reference decay experiment in which we measured the degradation kinetics of each gene after transcriptional arrest, yet without applying any of the above stresses (see schematic representation of the experiment design in Supplementary Figure S1).

The majority of the genes showed exponential decay in all three decay experiments, suggesting constant rate of degradation throughout time (Supplementary Figure S7). As carried out earlier for non-stressful conditions (Wang *et al*, 2002), we could calculate the half-life of each gene in each of the three conditions. Interestingly, the decay rates of some entire gene modules showed coherent change in a given stress when compared with the reference condition or with the other stress, indicating extensive condition-specific regulation of mRNA

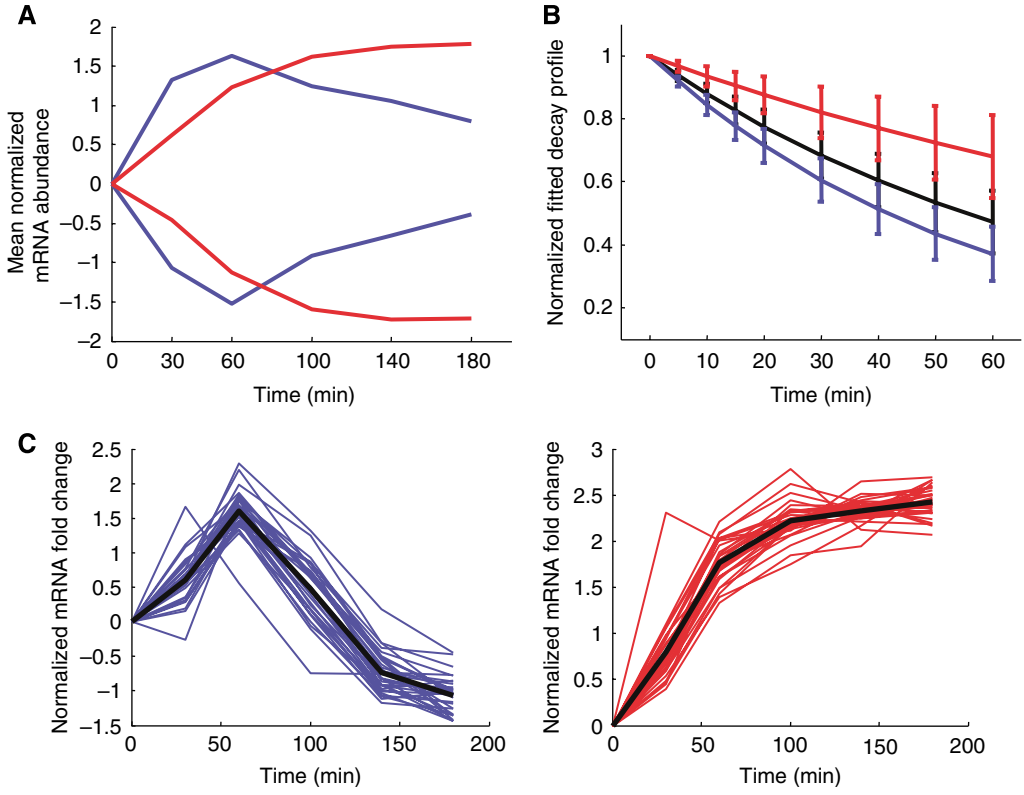


Figure 1 Distinct transcriptome responses at the two conditions. (**A**) Mean expression profile of all induced and repressed genes (fold change >2) in oxidative and MMS stress (blue and red curves, respectively). (**B**) The proteasomal genes as an example for a group of genes showing coherent change in mRNA stability in response to each stress. The mean of the fitted decay profiles is shown; black, blue and red represent the reference, oxidative stress and DNA damage conditions, respectively. (**C**) The mRNA abundance profiles of the proteasomal genes (after mean and variance normalization) are shown for the oxidative stress and DNA damage stress (left and right panels, respectively).

decay. Figure 1B shows an example of the average change in decay rate of the genes that encode the 26S proteasome. Although the genes show clear destabilization relative to the reference condition in response to the oxidative stress, they are significantly stabilized on treatment with MMS. Interestingly, although the stability of the proteasomal genes changes in opposite directions in the two conditions, their mRNA abundance profiles (measured in the experiment that did not involve transcriptional arrest) show a clear net induction in both the stabilizing and the destabilizing conditions. Yet the kinetics of the response is markedly different, featuring fast relaxation in the oxidative stress, and sustained response in MMS (Figure 1C). Although *a priori* the fast relaxation seen in oxidative stress could result from decrease in transcription alone, it may alternatively be also due to destabilization of the proteasomal genes in this condition. In the following, we take a systematic genome-wide view to study the interplay between the change in mRNA abundance and the change in mRNA stability, with the aim of deciphering the relative role of production and degradation in shaping transcript kinetics.

Alternative interplays between changes in mRNA stability and changes in mRNA abundance

On the basis of mRNA abundance measurements, we characterize each gene in every condition by the maximal fold change of the gene's mRNA abundance levels following the stress, and by the time at which such maximal fold change is attained (see Materials and methods). A third parameter that we use for this characterization is the log ratio of the half-life of the gene's mRNA in a given stress and it's half-life in the reference condition. This ratio serves as a measure of the stabilization/destabilization of the gene's mRNA in the stress (see Materials and methods). To investigate, in every condition, the interplay between changes in mRNA abundance and changes in mRNA degradation rates, we assessed the correlation between maximal fold change of each gene to the half-life ratio in a given condition (Figure 2). We found two opposing relationships in the two stressful conditions. Examining the response to MMS, we observe an opposite, and at first more intuitive trend. In this condition, induced genes show a tendency toward stabilization, whereas repressed genes show a tendency toward destabilization (Figure 2). In the oxidative stress, we found a surprising negative correlation—genes whose mRNAs are induced in response to the stress are typically destabilized, whereas repressed genes show a weaker, though still significant, tendency toward stabilization.

We further mined the data with respect to the dynamics of different functional categories using gene ontology (GO) (www.geneontology.org). We have first identified GO categories that are enriched among the induced and repressed genes (see Supplementary information), and found that these are in good agreement with classical analyses (Gasch *et al*, 2000). We then asked, separately for GO categories that are enriched among repressed or induced genes, whether the genes belonging to them are significantly stabilized or destabilized (Table I and Supplementary information). We used two separate statistical tests, one that asks whether enriched categories have a general tendency toward stabilization or destabilization, and a second test to check whether specific groups behave anomalously when compared to the general tendency of induced or repressed genes (see Materials and methods). In oxidative stress, induced genes have a strong tendency toward destabilization, whereas the opposite is observed following DNA damage, and indeed most enriched categories, particularly genes involved in response to stimuli, and oxidoreductase activity, behave according to this trend. Repressed genes in both conditions show a less clear trend, and enriched categories are found both stabilized and destabilized. Ribosomal proteins and rRNA metabolism proteins represent an interesting deviation from the (rather weak) genome-wide trend—these genes are repressed in response to oxidative stress, yet they undergo destabilization. This probably explains why these genes show less transient repression compared to the general repression in this condition.

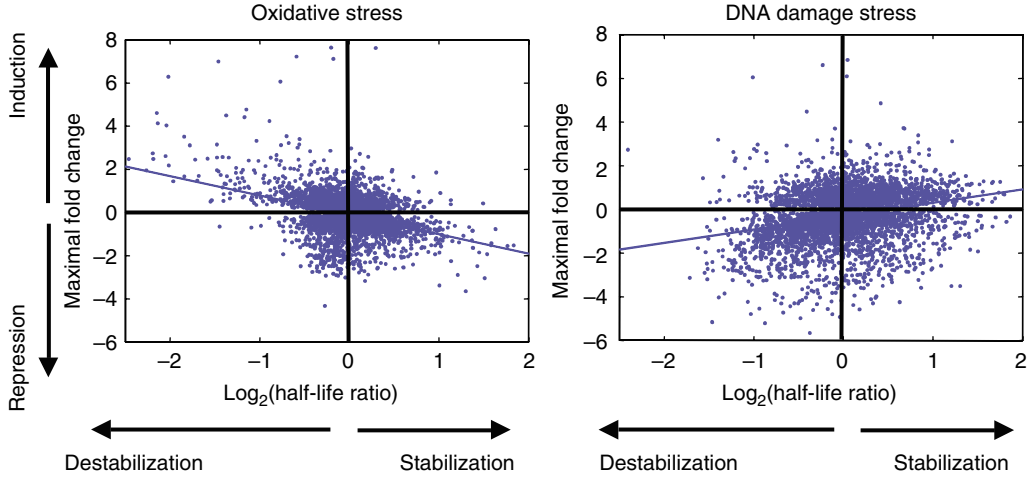


Figure 2 Distinct relationships between the change in mRNA abundance and the change in stability between the two conditions. For each stress, the change in mRNA stability relative to the reference state ($\log_2(t_{1/2\ \text{stress}}/t_{1/2\ \text{reference}})$) is plotted against the maximal fold change (defined as described in Materials and methods). Two different trends are observed, a negative trend ($R=-0.38$, p-value $< 3 \times 10^{-150}$) for the oxidative stress and a positive trend for the DNA damage stress ($R=0.27$, p-value $< 4.2 \times 10^{-70}$).

Table I Sample categories that are both enriched among the induced and repressed genes and show a significant tendency toward stabilization or destabilization

Induced genes						Repressed genes					
Destabilized (78%)			Stabilized (22%)			Destabilized (47%)			Stabilized (53%)		
Category name	*t*-test *P*-value	χ^2 *P*-value	Category name	*t*-test *P*-value	χ^2 *P*-value	Category name	*t*-test *P*-value	χ^2 *P*-value	Category name	*t*-test *P*-value	χ^2 *P*-value
Oxidative stress											
Mitochondrion	6.7E−10	0.056				Ribosome biogenesis	0.17	1.80E−04	Nuclease activity	0.011	4.90E−04
Response to stimulus	1.1E−5	0.58				Nucleolus	0.005	4.40E−06	Bud neck	1.40E−04	0.0079
Oxidoreductase activity	1.3E−4	0.75				Nuclear lumen	0.34	1.30E−04	Ribonuclease activity	0.0033	0.0024
Response to abiotic stimulus	3.2E−4	0.25				RNA metabolism	0.75	0.02	Biopolymer metabolism	0.0042	0.77
Cofactor metabolism	0.003	0.72				rRNA metabolism	0.71	0.05	Cytoplasmic exosome	0.0069	0.014
Destabilized (36%)			Stabilized (64%)			Destabilized (66%)			Stabilized (34%)		
Category name	*t*-test *P*-value	χ^2 *P*-value	Category name	*t*-test *P*-value	χ^2 *P*-value	Category name	*t*-test *P*-value	χ^2 *P*-value	Category name	*t*-test *P*-value	χ^2 *P*-value
DNA damage											
			Proteolysis	6.10E−05	0.0079	Translation	7.20E−56	6.20E−14	Nuclear part	0.045	3.30E−15
			Cellular macromolecule catabolism	1.30E−04	0.04	Ribosome	1.10E−52	7.30E−13	Nucleolus	0.0019	2.30E−15
			Carbohydrate metabolism	6.80E−04	0.02	Protein biosynthesis	1.00E−50	3.40E−13	RNA processing	0.0013	9.30E−14
			Protein catabolism	0.0012	0.038	Sterol metabolism	1.50E−07	0.0021	Biopolymer metabolism	0.8	2.70E−09
			Ubiquitin cycle	0.0056	0.038	Lipid biosynthesis	2.10E−07	0.0055	Methyltransferase activity	0.042	3.70E−04

All categories in this table have passed an initial hypergeometric test that indicated that they are enriched in the group of induced or repressed genes irrespective of their tendency toward stabilization or destabilization. Then two tests were performed on the half-life change of the group of genes that were both annotated by a specific category and induced or repressed: a one-sided *t*-test was performed to check whether the mean half-life change was significantly different from zero, indicating stabilization or destabilization. To look for groups of genes that behave anomalously, a χ^2 test was performed on the number of stabilized and destabilized genes in the intersection between the category genes and the group of induced or repressed genes when compared with all induced or repressed genes. The full list of enriched categories is given in Supplementary data set 2. Cases where both tests display a low *P*-value (as 'translation' in DNA damage repression and destabilization) indicate that the trend is stronger than the trend of the reference group but in the same direction.

Fast transcriptional responses are accompanied by an opposing force of degradation

We then investigated whether the correlation between the mRNA abundance fold change to the change in stability is different, within each condition, between genes that display different kinetics. For that, we grouped the genes in each condition according to the time of attainment of the maximal fold change. We then plotted again, for each group separately, the change in mRNA stability against the maximal fold change (also plotted are the corresponding mRNA abundance profiles in each set; Figure 3A). In both conditions, fast induced genes show strong destabilization, whereas fast repressed genes show stabilization. For instance, among the induced genes that peak in the first 40 min following the stress, a 2- to 4-fold reduction in half-life when compared with the reference

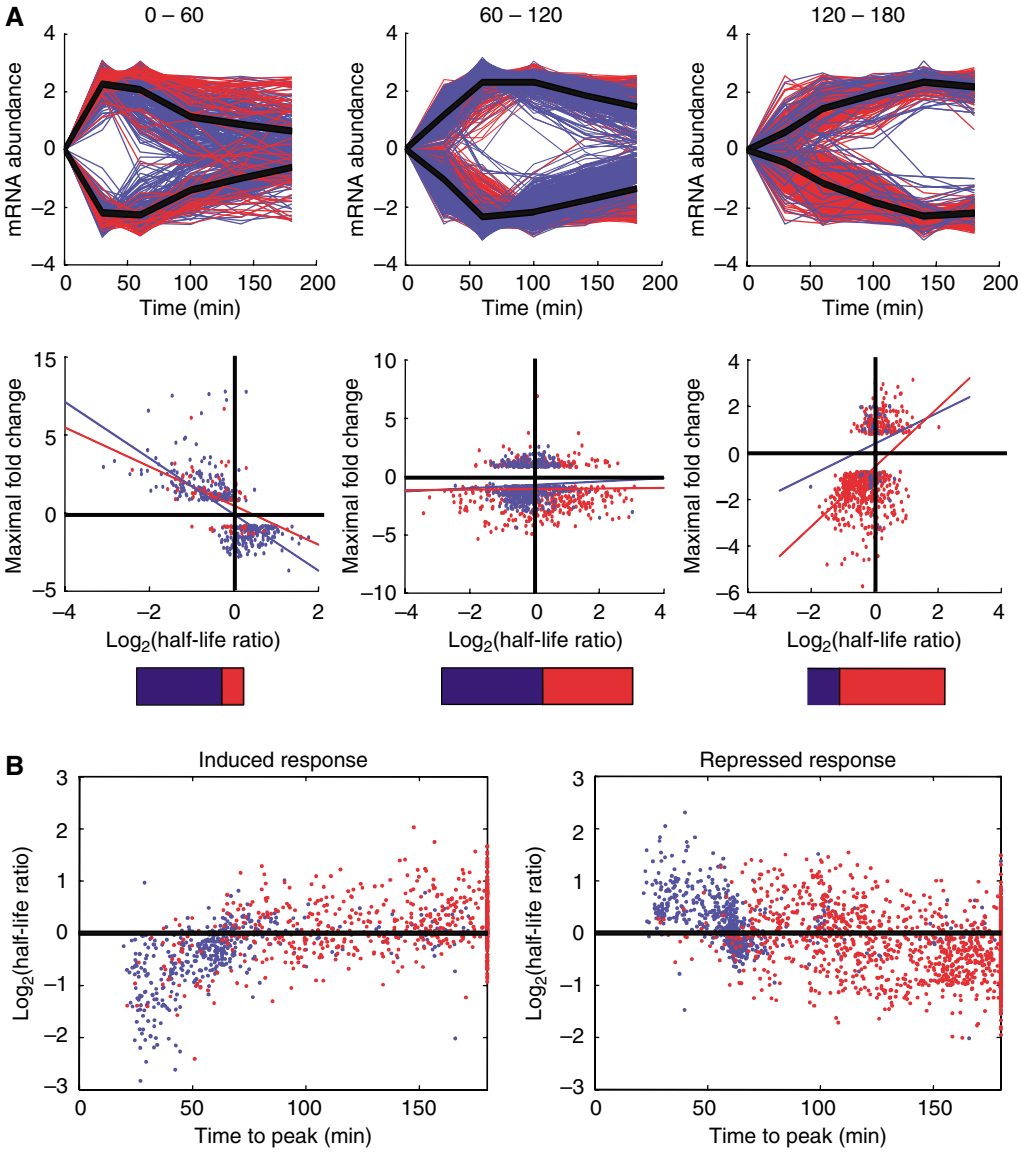


Figure 3 (**A**) Relationship between changes in mRNA abundance to changes in mRNA stability in different kinetic regimens. For both conditions, mRNA profiles are grouped according to the time point at which the maximal fold change is attained (0–60, 60–120 and 120–180 minutes). For each such group, we plot both normalized (mean and variance) mRNA abundance profiles (upper panels) and, as was done in Figure 2, the relationship between the maximal mRNA abundance fold changes to the changes in mRNA stability relative to the reference condition (lower panels). We joined in each plot profiles from the two conditions, blue and red correspond to oxidative and MMS stress, respectively. The first group of genes, which mostly consists of profiles in the oxidative stress condition, displays transient kinetics and an early peak, in the first 40 min following the stress. These genes display an opposite relationship between the changes in mRNA stability and the net change in mRNA abundance. As we progress to groups of genes that attained a later peak, the negative correlation is replaced with a positive correlation. The last group, which displays long enduring kinetics, shows a positive correlation between the changes in mRNA abundance to the change in mRNA stability. The bars below each group represent the relative amount of genes from each stress represented in each group. (**B**) Changes in mRNA stability determine response duration. The time at which the maximal fold change in mRNA abundance is attained against the half-life change is plotted separately for induced and repressed genes. The different colors represent the two conditions, blue and red for oxidative and DNA damage stress, respectively. Opposite trends are observed between induced and repressed genes. In addition, both conditions display similar trends with difference in the amount of genes that display transient versus endured response.

condition is common. In contrast, in both conditions, genes that display a slow long enduring response show a positive correlation between changes in the mRNA abundance level and changes in stability. Among these genes, induction is accompanied by stabilization, whereas repression is often seen among the destabilized genes. It is therefore apparent that in both conditions similar basic kinetic trends are obeyed: for genes in which a change in degradation rates counteracts the direction of change in mRNA abundance, a fast transient response is seen. In contrast, in both conditions, a long enduring response is seen if upregulated genes are also stabilized, and downregulated genes are destabilized. The two conditions thus only differ in the relative proportion of genes that display the transient versus the long enduring response as apparent from the horizontal bars at the bottom of Figure 3A.

To investigate quantitatively how changes in mRNA stability in response to stress affect the duration of mRNA abundance response, we plotted the half-life ratio against the time of maximal fold change attainment for induced and repressed genes separately (Figure 3B). We see a clear correspondence between the two quantities. For induced genes, in both conditions, short responses are obtained in proportionality to the extent of destabilization in the stress. In contrast, repressed genes in both conditions show the opposite trend—shorter times to maximal fold change are observed among the stabilized genes. We propose that destabilization of transient induced genes and stabilization of transient repressed genes accelerate the relaxation phase back to the (lower and higher, respectively) base level. In this respect, the picture with induced and repressed transient genes is consistent: in both cases, counteraction between production and degradation is suggested to facilitate the fast relaxation of the initial response.

These results suggest that mRNA decay is a key feature regulating the response duration of mRNA abundance.

As stated above, an increase in decay rate, as seen for the transient induced genes, is expected to shorten the response time (Box 1). Thus, it is possible that for these genes, the increase in decay rate serves not only to achieve fast relaxation but also to accelerate the response to the new stressful conditions. Reassuringly, for these fast transient induced genes the destabilization starts rather early, even before these genes attain their maximal level (Supplementary Figure S2), revealing a potential to affect the response time.

Discussion

We report an intriguing relationship between changes in mRNA abundance and changes in mRNA degradation in response to stress. We show that this relationship varies between conditions, and among genes with different kinetics within the same condition. In particular, among the transient genes, which attain maximal fold change, and begin relaxation within the first 60 min, it appears that the change in the mRNA abundance level is counteracted by the change in degradation rate.

Clearly, the destabilization of transiently induced genes has to be compensated by a large increase in transcription rates, otherwise a net increased level would not be obtained. Likewise, fast repressed transient genes must experience a decrease in production rate, as at the degradation level they are actually stabilized. Thus, although we deduce transcription rates indirectly, we can still conclude in this case, that transcription rate has increased among the induced destabilized genes and decreased among the repressed stabilized genes. A more direct indication that among destabilized genes transcription rates are predominantly increased, and that among stabilized genes transcription rates are most reduced, may be obtained from the measurements done recently by Molina-Navarro *et al* (2008). Reassuringly, re-analysis of their data strongly shows exactly that trend (Supplementary Figure S13). Among genes that respond either by induction or repression of above twofold change, genes which, by our measurements, are destabilized show a mean increase in transcription rate of above 60% in most time points, whereas genes that we found to be stabilized show a general decrease in transcription rate of about 40%. An interesting conclusion regarding the fast transient responding genes is that changes in transcription rates alone (increase or decrease) determine the *direction* of the response (induction or repression). Stability changes appear not to determine the direction of the net response as they actually occur in opposite directions. Yet our results ascribe a major role to changes in degradation—these, along with likely changes at the transcriptional level, appear to impact the speed and relaxation properties of the response. These conclusions are in line with the study of Perez-Ortin *et al* (2007) who showed for the *STL1* gene in yeast that the measured transcription rate profile was not sufficient to explain a transient mRNA abundance response after osmotic stress; an increase in degradation rate had to be assumed to achieve a rapid decrease back to the basal level. The combined increase in both production and degradation rates, observed at the transient induced genes, may thus represent an interesting strategy, as it allows the acceleration of the response, without compromise of the maximal expression level (see Box 1).

On the other hand, for genes that display a high endurance response, both up- and downregulated, it is possible that the direction of the response is both due to transcriptional and degradation effects, as for these genes both factors do not counteract each other. A potential explanation for the behavior seen among the long endurance responsive genes could be that changes in mRNA stability are used to maintain and enhance the changes in transcription, i.e. induced genes are further stabilized, whereas repressed genes are also degraded faster.

Increase in both production and degradation can account for a transient response if the increase in degradation is slower than the increase in production. Under these assumptions, for a short time period, production rate may be higher than degradation rate. In this period, mRNA level may overshoot when compared with the final steady state, and the relaxation is attained when degradation rate exceeds the rate of production (Supplementary Figure S3). Assuming that an external signal control both production and degradation rates, this model produces a relaxation of the response in mRNA even if the external signal (e.g. hydrogen peroxide in the present case) is still ON. Support to this idea is given by the fact that the levels of hydrogen peroxide are decreased only slightly throughout the experiment (not shown) in agreement with previous studies (Gasch *et al*, 2000), whereas the mRNA

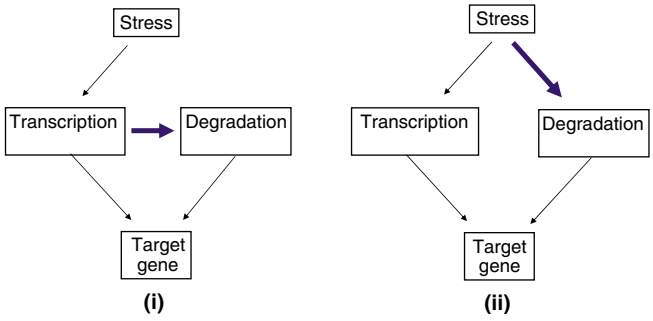


Figure 4 Two alternative models that might account for the observed coordination between transcription and degradation. Our results suggest coordination between changes in transcription to changes in mRNA degradation; this might be achieved by at least two alternative models: The first (i) suggests direct coupling between transcription and degradation meaning that mRNA degradation is directly affected by the rates of production. According to a possible alternative model (ii), the sensor of the stress activates a transcriptional response and, independently of that, it also induces a change in stability of the transcripts.

response begins to relax after an hour. Although concomitant increase in both production and degradation rates is an energetically costly solution, it might have an important advantage as it accelerates both the response to the stress and the relaxation back toward fast growing mode.

Finally, the idea of counteraction between production and degradation suggests coordination of transcriptional and mRNA degradation in the cell. Recently, experimental evidence suggested one potential mechanism that could account for this interplay. Lotan *et al* (2005, 2007) have shown that two subunits of RNA polymerase II, Rpb4p and Rpb7p, are involved in mRNA decay by enhancing both deadenylation and decapping. This suggests that counteraction could be achieved through direct coupling between transcription and degradation, namely that degradation rates are directly affected by changes in the rates of transcription. An alternative model is that the sensor of the stress activates a transcriptional response and, independently of that, it also induces a change in stability of the transcripts (Figure 4). This latter model has a topology reminiscent of a feed-forward loop, a recurring motif in many regulatory networks (Milo *et al*, 2002; Mangan *et al*, 2006), which was curiously found to accelerate response time to nutrient changes (Mangan *et al*, 2006). In this respect, it is noteworthy that recent observations made in the mammalian network spanned by microRNAs and transcription factors have proposed that similar coupling between transcription and post-transcription regulators may be implemented through similar in-coherent feed-forward loops (Shalgi *et al*, 2007; Sinha *et al*, 2008). Future work will be needed to determine the relative contribution of these models, or yet alternative ones, to the interplay seen here between the transcriptional and post-transcriptional regulation.

Materials and methods

Strains and growth conditions

Two types of experiments were conducted: experiments measuring mRNA abundance and mRNA decay experiments. Changes in mRNA abundance were measured in two separate experiments in response to both conditions described. mRNA decay experiments were performed in three different conditions: two for each of the stressful conditions and a reference condition, which was actually performed in two independent biological replicates as described below.

All experiments were carried out using the *S. cerevisiae* strain Y262 carrying a temperature-sensitive mutation in RNA polymerase II (Nonet *et al*, 1987) (*Matα ura3–52 his4–539 rpb1-1*). For all experiments, cells were grown in YPD medium (2% yeast extract, 1% peptone and 1% dextrose) at 26°C to the concentration of 2×10^7 cells/ml. Cells were then treated either by 0.1% MMS (Sigma-Aldrich) or 0.3 mM H_2O_2 (Frutarom Ltd). As a control experiment, no reagent was added. For mRNA abundance measurements, following each treatment, aliquots (15 ml) were removed at the following time points: 0, 30, 60, 100, 140 and 180 min and frozen in liquid nitrogen. RNA was extracted using MasterPure™ (Epicenter Biotechnologies). The quality of the RNA was assessed using the Bioanalyzer 2100 platform (Agilent); the samples were then processed and hybridized to Affymetrix yeast 2.0 microarrays using the Affymetrix GeneChip system according to manufacturer's instructions.

For measuring mRNA decay, a similar protocol was applied albeit with the following modifications: 25 or 40 min following addition of H_2O_2 or MMS (respectively), temperature was abruptly raised to the restrictive temperature of 37°C by the addition of an equal volume of a medium pre-warmed to 49°C. This step inactivated the RNA polymerase II and therefore stopped transcription. Following the temperature shift, aliquots were removed at the following time points: 0, 5, 10, 15, 20, 30, 40, 50 and 60 min and were processed and hybridized as described earlier. Time point zero in each of the four decay experiments was hybridized to arrays twice independently, representing two technical replicas. The purpose of this replica is two: first, it allows the validation of the normalization method—the ratio between the mRNA levels and spiked in RNA was found to be constant in replicates of the same time point (see Supplementary information for more details). In addition, the technical replica allows the assessment of the reproducibility of array hybridizations (Supplementary Figure S8). We also performed one complete biological replica of an entire decay experiment, at the reference condition. Although estimated half-life values can vary between replicas of the same condition, these analyses strongly suggested that the variation due to replicates is much smaller than the variation in estimated half-life across conditions (see Supplementary information).

Data preprocessing

Most preprocessing algorithms use a normalization step to bring all samples to have the same global distribution of intensity values. This is done under the assumption that the mean intensities of all samples should be similar, and deviations, between samples, represent technical artifacts that result from a difference in the processes that the samples undergo till hybridization and scanning. A unique aspect of mRNA decay measurements is that the above assumption, that there is no change in the total levels of mRNAs, is by definition not valid: due to transcription inhibition, we do expect a global decrease in the total amount of mRNA. Thus, for normalization between time points, we used an internal standard that was mixed with each RNA sample. This standard contained a pool of four—*in vitro* transcribed *Bacillus subtilis* RNAs ('spiked in'), each in a different concentration (poly(A) control kit supplied by Affymetrix). Each transcript was represented on the

microarrays by several probe sets. For a more detailed explanation on the preprocessing procedure, see Supplementary information and Supplementary Figures S4–S8.

Determination of mRNA half-lives

We represent the stability of an mRNA in each condition using a half-life that is derived from the measured decay profile in each condition. Each transcript decay profile is zero transformed by dividing each gene's decay profile by the mean measured expression value of the two replicates at time point zero, then assuming a constant decay rate throughout the course of the experiment the decay profile is fitted to a first-order exponential decay model, $y(t)=y(0)\cdot e^{-kt}$, from which the fitted decay constant k is used to calculate a gene-specific half-life in each condition, $t_{1/2}=\ln(2)/k$. Only genes for which a relatively good fit is achieved (R-square >0.7) are taken for further analysis ($\sim$70% of the genes).

Determination of the maximal fold change and the time at which it is attained

Responsive genes were defined as having an absolute fold change of above twofold for at least one time point out of the mRNA abundance measurements. These measurements were then used to approximate the full response by fitting the profiles of these genes to a cubic spline with breaks at each measured time point. The maximal point was taken as the point where both the spline derivatives were equal to zero and the fitted spline value reached the maximal absolute value (maximal for induced genes and minimal for repressed genes). For genes with a fitted spline that was constantly increasing or decreasing throughout the whole time course, the last point of the time course was considered the maximal point. It is important to note that the results presented in this paper will not change qualitatively if the data would be treated as a discrete time course containing only the measured time points.

Data mining

We have used two data sets, GO (http://www.geneontology.org/) and KEGG pathways (http://www.genome.ad.jp/kegg/pathway.html), to mine the data with respect to the dynamics of different functional categories. For each category, we took the intersection between the annotated genes to the induced/repressed genes in each condition and performed three separate statistical tests. First, we used the hypergeometric distribution to check which categories are enriched in the induced and repressed sets of genes using only the data on changes in mRNA abundance. Two additional tests were performed on the logarithm of the half-life ratio for each intersection. A one-sided t-test was performed to check whether it deviates significantly from zero, indicating a tendency for stabilization or destabilization. Additionally, to test whether some groups of genes behave anomalously when compared with the general trend of induced or repressed genes, we used a χ^2 test on the number of destabilized/stabilized genes in the intersection when compared with the expected number based on the percentage of destabilized genes in all induced or repressed genes. A representative sample of the results is given in Table I and the full data set is provided as part of the Supplementary information.

Single-gene measurements of decay and mRNA abundance with RT–PCR

We performed real-time PCR experiments on selected genes to verify the array-derived decay profiles and changes in mRNA abundance (see Supplementary Figure 14A–D). For mRNA quantification, a 2 μg aliquot of total RNA was reverse transcribed using random primers. RT–PCR was performed using 480 SYBR Green I Master (Roche) Reagent on LightCycler 480 Real-Time PCR system (Roche).

Expression data

The entire expression data set, including decay profiles in the reference condition, and under oxidative and MMS stresses, along with mRNA abundance profiles in each of the stress conditions, is deposited in the GEO databases (GSE12222). The data set and additional supplementary information is also available on line at http://longitude.weizmann.ac.il/pub/papers/Shalem2008_mRNAdecay/suppl/, and at MSB website.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

We thank The Pilpel and Segal labs for fruitful discussions, especially Bella Groisman and Zohar Bloom for help with experiments. We thank Shirley Horn-Saban and Ron Ophir from the microarray unit of the Weizmann Institute and Amos Grundwag from Affymetrix for excellent support. We thank Richard Young for kindly providing the RNA polymerase II temperature-sensitive strain. ES is supported by an NIH grant HG004361-01 and is the incumbent of the Soretta and Henry Shapiro career development chair. YP is an incumbent of the Rothstein Career Development Chair in Genetic Diseases. YP acknowledges support from EMBRACE—an EU-funded network of excellence and from the 'Ideas' grant of the European Research Council. The research leading to these results has received funding from the European Research Council's Seventh Program (FP7/2007–2013).

Conflict of interest

The authors declare that they have no conflict of interest.

References

Bernstein JA, Khodursky AB, Lin PH, Lin-Chao S, Cohen SN (2002) Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc Natl Acad Sci USA* **99:** 9697–9702

Fan J, Yang X, Wang W, Wood III WH, Becker KG, Gorospe M (2002) Global analysis of stress-regulated mRNA turnover by using cDNA arrays. *Proc Natl Acad Sci USA* **99:** 10611–10616

Foat BC, Houshmandi SS, Olivas WM, Bussemaker HJ (2005) Profiling condition-specific, genome-wide regulation of mRNA stability in yeast. *Proc Natl Acad Sci USA* **102:** 17675–17680

Garcia-Martinez J, Aranda A, Perez-Ortin JE (2004) Genomic run-on evaluates transcription rates for all yeast genes and identifies gene regulatory mechanisms. *Mol Cell* **15:** 303–313

Garneau NL, Wilusz J, Wilusz CJ (2007) The highways and byways of mRNA decay. *Nat Rev Mol Cell Biol* **8:** 113–126

Gasch AP, Huang M, Metzner S, Botstein D, Elledge SJ, Brown PO (2001) Genomic expression responses to DNA-damaging agents and the regulatory role of the yeast ATR homolog Mec1p. *Mol Biol Cell* **12:** 2987–3003

Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO (2000) Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell* **11:** 4241–4257

Gasch AP, Werner-Washburne M (2002) The genomics of yeast responses to environmental stress and starvation. *Funct Integr Genomics* **2:** 181–192

Grigull J, Mnaimneh S, Pootoolal J, Robinson MD, Hughes TR (2004) Genome-wide analysis of mRNA stability using transcription inhibitors and microarrays reveals posttranscriptional control of ribosome biogenesis factors. *Mol Cell Biol* **24:** 5534–5547

Jelinsky SA, Estep P, Church GM, Samson LD (2000) Regulatory networks revealed by transcriptional profiling of damaged *Saccharomyces cerevisiae* cells: Rpn4 links base excision repair with proteasomes. *Mol Cell Biol* **20:** 8157–8167

Lotan R, Bar-On VG, Harel-Sharvit L, Duek L, Melamed D, Choder M (2005) The RNA polymerase II subunit Rpb4p mediates decay of a specific class of mRNAs. *Genes Dev* **19:** 3004–3016

Lotan R, Goler-Baron V, Duek L, Haimovich G, Choder M (2007) The Rpb7p subunit of yeast RNA polymerase II plays roles in the two major cytoplasmic mRNA decay mechanisms. *J Cell Biol* **178:** 1133–1143

Mangan S, Itzkovitz S, Zaslaver A, Alon U (2006) The incoherent feed-forward loop accelerates the response-time of the gal system of *Escherichia coli*. *J Mol Biol* **356:** 1073–1081

Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U (2002) Network motifs: simple building blocks of complex networks. *Science* **298:** 824–827

Molina-Navarro MM, Castells-Roca L, Belli G, Garcia-Martinez J, Marin-Navarro J, Moreno J, Perez-Ortin JE, Herrero E (2008) Comprehensive transcriptional analysis of the oxidative response in yeast. *J Biol Chem* **283:** 17908–17918

Narsai R, Howell KA, Millar AH, O'Toole N, Small I, Whelan J (2007) Genome-wide analysis of mRNA decay rates and their determinants in *Arabidopsis thaliana*. *Plant Cell* **19:** 3418–3436

Nonet M, Scafe C, Sexton J, Young R (1987) Eucaryotic RNA polymerase conditional mutant that rapidly ceases mRNA synthesis. *Mol Cell Biol* **7:** 1602–1611

Parker R, Song H (2004) The enzymes and control of eukaryotic mRNA turnover. *Nat Struct Mol Biol* **11:** 121–127

Perez-Ortin JE, Alepuz PM, Moreno J (2007) Genomics and gene transcription kinetics in yeast. *Trends Genet* **23:** 250–257

Shalgi R, Lapidot M, Shamir R, Pilpel Y (2005) A catalog of stability-associated sequence elements in 3′ UTRs of yeast mRNAs. *Genome Biol* **6:** R86

Shalgi R, Lieber D, Oren M, Pilpel Y (2007) Global and local architecture of the mammalian microRNA-transcription factor regulatory network. *PLoS Comput Biol* **3:** e131

Sinha S, Adler AS, Field Y, Chang HY, Segal E (2008) Systematic functional characterization of cis-regulatory motifs in human core promoters. *Genome Res* **18:** 477–488

Wang Y, Liu CL, Storey JD, Tibshirani RJ, Herschlag D, Brown PO (2002) Precision and functional specificity in mRNA decay. *Proc Natl Acad Sci USA* **99:** 5860–5865

Wilusz CJ, Wilusz J (2004) Bringing the role of mRNA decay in the control of gene expression into focus. *Trends Genet* **20:** 491–497

Wilusz CJ, Wormington M, Peltz SW (2001) The cap-to-tail guide to mRNA turnover. *Nat Rev Mol Cell Biol* **2:** 237–246

Workman CT, Mak HC, McCuine S, Tagne JB, Agarwal M, Ozier O, Begley TJ, Samson LD, Ideker T (2006) A systems approach to mapping DNA damage response pathways. *Science* **312:** 1054–1059

Xu N, Chen CY, Shyu AB (1997) Modulation of the fate of cytoplasmic mRNA by AU-rich elements: key sequence features controlling mRNA deadenylation and decay. *Mol Cell Biol* **17:** 4611–4621

Yang E, van Nimwegen E, Zavolan M, Rajewsky N, Schroeder M, Magnasco M, Darnell Jr JE (2003) Decay rates of human mRNAs: correlation with functional characteristics and sequence attributes. *Genome Res* **13:** 1863–1872

OPEN ACCESS Freely available online

PLoS GENETICS

Transcriptome Kinetics Is Governed by a Genome-Wide Coupling of mRNA Production and Degradation: A Role for RNA Pol II

Ophir Shalem, Bella Groisman, Mordechai Choder, Orna Dahan[9], Yitzhak Pilpel[9]*

Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel

Abstract

Transcriptome dynamics is governed by two opposing processes, mRNA production and degradation. Recent studies found that changes in these processes are frequently coordinated and that the relationship between them shapes transcriptome kinetics. Specifically, when transcription changes are counter-acted with changes in mRNA stability, transient fast-relaxing transcriptome kinetics is observed. A possible molecular mechanism underlying such coordinated regulation might lay in two RNA polymerase (Pol II) subunits, Rpb4 and Rpb7, which are recruited to mRNAs during transcription and later affect their degradation in the cytoplasm. Here we used a yeast strain carrying a mutant Pol II which poorly recruits these subunits. We show that this mutant strain is impaired in its ability to modulate mRNA stability in response to stress. The normal negative coordinated regulation is lost in the mutant, resulting in abnormal transcriptome profiles both with respect to magnitude and kinetics of responses. These results reveal an important role for Pol II, in regulation of both mRNA synthesis and degradation, and also in coordinating between them. We propose a simple model for production-degradation coupling that accounts for our observations. The model shows how a simple manipulation of the rates of co-transcriptional mRNA imprinting by Pol II may govern genome-wide transcriptome kinetics in response to environmental changes.

Citation: Shalem O, Groisman B, Choder M, Dahan O, Pilpel Y (2011) Transcriptome Kinetics Is Governed by a Genome-Wide Coupling of mRNA Production and Degradation: A Role for RNA Pol II. PLoS Genet 7(9): e1002273. doi:10.1371/journal.pgen.1002273

Editor: Gregory S. Barsh, Stanford University School of Medicine, United States of America

Received May 13, 2011; **Accepted** July 19, 2011; **Published** September 8, 2011

Funding: We thank the ''Ideas'' program of the European Research Council and the Ben May Charitable Trust for grant support. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: pilpel@weizmann.ac.il

[9] These authors contributed equally to this work.

Introduction

The dynamics of the transcriptome in response to environmental changes is chiefly governed by two opposing processes – RNA production, namely transcription, and RNA degradation. Despite this fact, most of the attention has been given to the study of transcription. Recently genome-wide techniques have been established that allow to measure separately the contribution of mRNA degradation [1–4] and transcription [5–8] to the balanced mRNA levels in the cell. Such studies revealed extensive regulation on both production and degradation rates. In particular, it became apparent that mRNA degradation is heavily regulated - genes that belong to the same complexes or gene modules, such as the ribosomal proteins or the proteasome, were shown to be co-degraded in several conditions and are considered to be part of the same decay regulon [2,9,10]. In addition, the decay rates of some genes across various growth conditions showed extensive variation, featuring stabilization in some conditions and de-stabilization in others [9,11,12].

Yet the emerging picture from many of these studies is that in addition to heavy regulation on both levels of production and degradation, there is often a correlation between the regulation of the two levels. In particular several studies have shown a ''counter-action'' mode of coupling between the two levels of control. In this mode of coupling genes that are induced at a given situation undergo, somewhat surprisingly, de-stabilization. The outcome of such type of coupling appears to be a fast transient change in mRNA abundance. This notion was demonstrated in the yeast *Saccharomyces cerevisiae* [9,10] in *Saccharomyces pombe* [13] and in mammals [14]. Interestingly, counter-action is not the only mode of coupling of production and degradation of mRNAs. Genes that show a slow, and sustained dynamic change in mRNA levels in response to a certain stimulus typically display the opposite and more intuitive correlation by which changes in mRNA abundance and stability are in the same direction [5,9]. The above results indicate that regulation of mRNA stability and its mode of coupling with transcription have a major contribution to the shape and dynamics of the transcriptome response to the environment [9,13,15,16].

A major question is thus – is there a molecular mechanism in cells which ensures the coupling between transcription and mRNA decay. A potential mechanistic basis for such coupling can be suggested based on detailed biochemical analysis of individual genes in the yeast *S. cerevisiae*. It was recently shown that, in addition to their genome wide role in transcription [17,18], two subunits of the RNA polymerase II (Pol II), Rpb4 and Rpb7, which belong to the holo-enzyme, not the core enzyme, can associate with the transcript during transcription and later chaperone it to the cytoplasm [19–22]. Crystal structure analysis of Pol II supports this notion since Rpb7 was shown to interact

Author Summary

Organisms alter genes expression programs in response to changes in their environment. Such programs can specify fast induction, slow relaxation, oscillations, etc. Conceivably these kinetic outputs may depend on proper orchestration of the various phases of gene expression, including transcription, translation, and mRNA decay. In particular, in the transcriptomes of a broad range of species, fast mRNA "spikes" appear to result from surprisingly "pressing the gas and the brakes" together, i.e. by activating both transcription and degradation of same transcripts. A recently discovered molecular mechanism, in which subunits of RNA polymerase II (Pol II) associate to mRNAs during transcription and control their decay, could explain how such transcription-decay counter-action works. Yet, how such potential coupling responds to physiological conditions and how it shapes transcriptome kinetics remain unknown. Here we used a minimalist mutation in yeast RNA Pol II that is defective in the above mechanism in order to show that Pol II governs the ability of the cell to modulate mRNA decay in stress and, most importantly, that Pol II is essential for appropriate coupling between mRNA production and degradation. We further show that this transcription-decay coupling is responsible for shaping the transcriptome kinetic profiles under changing environmental conditions.

with the nascent transcript as it emerges from the core polymerase [23]. In addition, *in vitro* studies, using proteins extracted from human cells, demonstrated that hsRpb7p interacts with the transcript as it emerges from Pol II [24]. Additional support for the role of Rpb4 in post-transcription control came from the realization that it recruits to the mRNA 3′ processing and polyadenylation enzymes [25]. Chaperoning of the transcript to the cytoplasm by the two polymerase subunits may affect a diversity of post-transcriptional process including translation and mRNA degradation [22]. These two subunits of Pol II may thus implement a simple means of coupling between transcription and mRNA decay. To show that the cytoplasmic role of Rpb4/7 depends on its nuclear association with Pol II core subunits in the nucleus, Goler-Baron et al [19] used a mutant in a subunit of the core polymerase, Rpb6. In this mutant, a glutamine at position 100 in Rpb6 was replaced with an arginine. This mutation displayed reduced ability to recruit Rpb4/7 both to the core polymerase [26] and to mRNA transcripts [19,22] which resulted in impaired production and degradation for a selected group of genes. The mutant thus seem to hold a key to understanding the coupling since it represents a minimalist perturbation that could de-couple genome-wide transcription from degradation, while maintaining intact Rpb4 and Rpb7 in the cell.

Here we used this minimalist perturbation to reveal a role of Pol II in affecting mRNA degradation and coordinating this process with transcription on a genome wide level. For that we compared gene expression, and decay rates in optimal growth conditions and in stress in the rpb6 mutant and a wild-type strain that is otherwise 100% identical to the mutant throughout the genome. We found that the rpb6 mutant is compromised in its ability to module mRNA decay rates, in addition to impaired transcription, in response to stress. As a result, while the wild-type features a counter-action coupling between production and degradation of mRNA, in the mutant this coupling is lost. This loss of negative correlation between changes in mRNA abundance to changes in stability in response to stress, results in an impaired temporal mRNA abundance response both in magnitude and kinetics. We thus conclude that in addition to its prime role in transcription, Pol II, also affects mRNA decay genome-wide, in a way that provides coupling of the two processes. This coupling appears to shape the kinetics of the transcriptome in response to stress. We propose a simple model which explains this coupling and its effect on the transcriptome.

Results

The experimental design

To check whether Rpb4/7 has a role in the previously observed counter-action coupling of mRNA production and degradation in stress [9] we used the rpb6^{Q100R} mutant previously described [26] and compared it to its parental isogenic wild-type strain that is other-wise 100% identical throughout the genome. Pol II containing this mutant subunit has a reduced ability to recruit Rpb4/7 subunits to transcripts [19,22]. The experimental setting is described schematically in Figure 1. Briefly, for both the wild type and rpb6^{Q100R} mutant strain we applied oxidative stress to two cultures, one for a conventional mRNA abundance profiling and one in which transcription was inhibited using 1,10-phenanthroline [1], 5–7 minutes after applying the stress, for mRNA decay measurement (see Materials and Methods). Following the addition of the drug mRNA levels were measured in five time points using the same microarray procedure as for the mRNA abundance profiling which did not involve transcription arrest. In addition we also measured for each strain mRNA stability without an oxidative stress, producing a reference decay profile for each gene. Decay profiles were fitted to an exponential decay model from which half-lives were calculated for each gene in each strain and condition (see Materials and Methods). The majority of the genes in the genome (>90%) showed a good fit to an exponential decay model (see Materials and Methods).

In another stress condition, the DNA-damaging drug MMS, which we have tuned before to invoke a qualitatively different transcriptome kinetics compared to the response to the current oxidative stress [9] we also compared the mutant and the wild-type though only in conventional mRNA profiling, without transcriptional arrest.

Rpb6 mutant is defective in counter-acting changes in mRNA production and degradation in response to stress

We computed for each gene in each strain the mRNA abundance maximal fold change in response to stress, a measure that can be affected both by changes in transcription and degradation. In parallel, to isolate response to stress at the degradation level, we characterize each gene in each strain by the ratio between its half-life in oxidative stress to its half-life in the reference condition. Figure 2A summarizes the data for the wild type strain: we observed a negative correlation between mRNA fold change and stability changes, i.e. induced genes show a tendency towards destabilization while repressed genes show a tendency to be stabilized in response to oxidative stress. This negative correlation is in agreement with our previous finding using slightly different experimental setup and a different genetic background [9]. In Figure 2B the same analysis is presented for the mutant on top of the wild type. This comparison reveals an almost complete elimination of the counter-action coupling that we observed in the wild type between mRNA abundance fold-change and stability change; this same result is presented in a more quantitative way in Figure 2C by calculating the observed vs. expected association between changes in production to changes in degradation assuming independence between these two measures (fold enrichment). In addition, the mutant shows a reduced

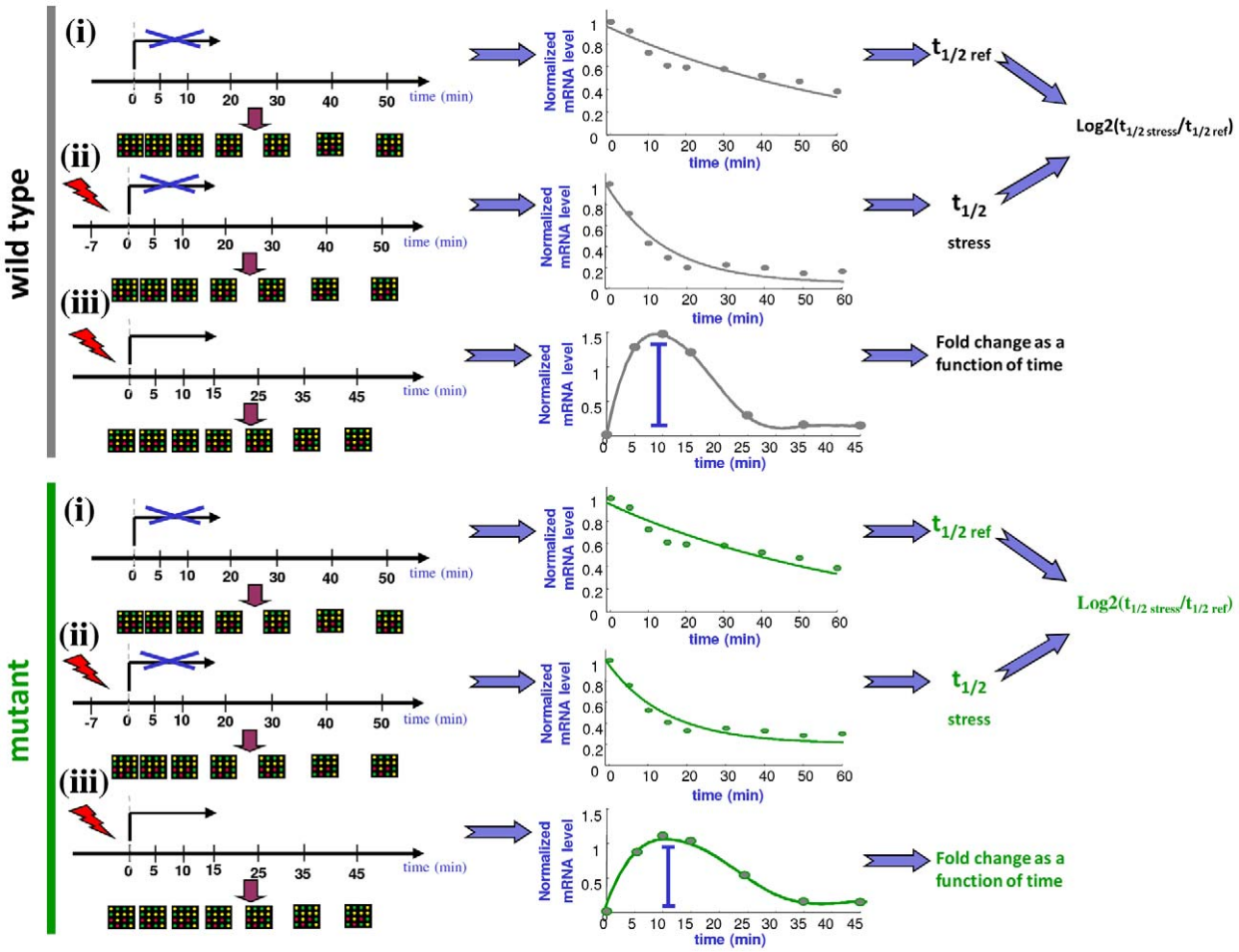


Figure 1. A schematic illustration of the experimental procedure. For each strain three types of experiments were conducted: (i) a reference decay experiment where decay kinetics was measured after transcription inhibition without applying additional stress. (ii) A stress followed by transcription inhibition to measure condition specific decay kinetics, and (iii) A conventional microarray experiment where mRNA abundance was measured following the perturbation.
doi:10.1371/journal.pgen.1002273.g001

capacity to change mRNA stability in response to stress, both extreme stabilization and extreme de-stabilization is not seen in the mutant (Figure 2B). Such a reduced capacity to modulate stability due to a mutation in Pol II is remarkable and by itself intriguingly suggests an effect of RNA Pol II on mRNA degradation.

These results are an indication that the counter-action mode of coupling may indeed require coordination between transcription and degradation and they further suggest that recruitment of Rpb4/7 to Pol II is important for the coupling mechanism.

Rpb6 mutant displays an impaired mRNA abundance temporal response to stress

We suggested earlier that the counter-action mode of coupling between production and degradation may be responsible for the spiked and fast relaxing dynamics on both induced and repressed genes in response to oxidative stress [9]. Rpb6 mutant provides a good opportunity to examine whether lack of coupling results in a less spiked and slower relaxing temporal mRNA abundance profiles. Yet, since the mutant is defective both in mRNA synthesis and stability we expect a more complex change in expression profiles, i.e. both in the magnitude and kinetics of the response.

To show the effect of the mutation on the entire genome and on different groups of genes we clustered concatenated wild-type and mutant mRNA abundance profiles for the stress-responsive genes (see Materials and Methods). As shown from the dendrogram in Figure 3, most of the stress-responsive genes reside within clusters that show a difference between the wild type and mutant (1 and 3). As expected the difference is displayed both in the magnitude of the response and in its kinetics. Reassuringly, for the induced genes that show a difference between the wild-type and the mutant, the difference in the kinetics matches with the expectations given the reduced destabilization: we see a less transient response in the mutant, which might be the effect of delayed de-stabilization-dependent relaxation of the early response and longer response time. To emphasize the differences in kinetics we also plot the mRNA abundance profiles of cluster 1 normalized by the standard deviation for each gene (Figure S1) which reveals more clearly the shift toward a slower response for induced genes. For repressed genes too most genes show reduced repression and a slight shift in response kinetics towards a more sustained and less transient behavior. In the case of repression the difference is more apparent in the magnitude and less in the kinetics. A possible explanation might be that reduced repression is a result of the lower basal level for these genes, see below.

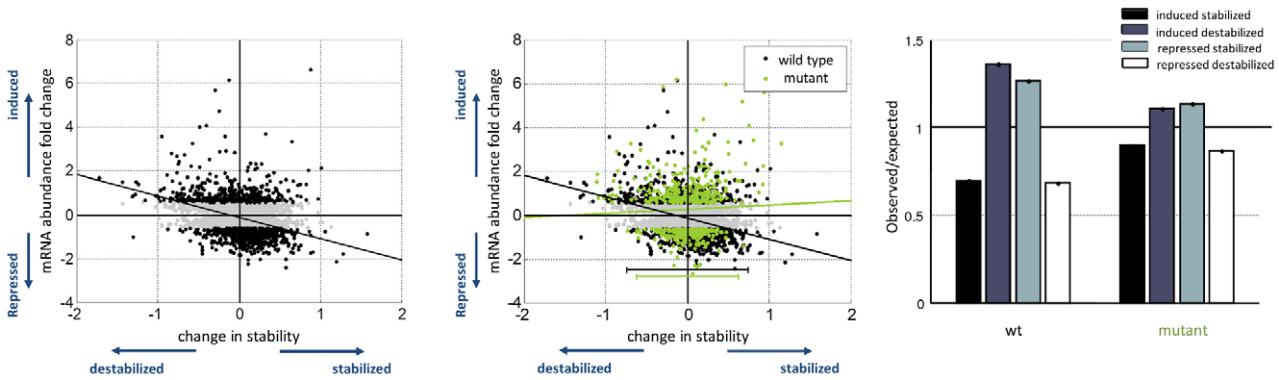


Figure 2. Reduced coupling in Rpb6 mutant strain. The change in mRNA stability relative to the reference state ($\log_2(t_{1/2oxidative} / t_{1/2reference})$) is plotted against the maximal fold change, defined as the maximal change in mRNA abundance for each gene across the time course. (A) Shows the wild type strain where black dots marks genes which respond to the stress. A negative correlation bywhich induced genes are destabilized is illustrated by the plotted least square line ($R=-0.23$, $-\log_{10}$(p-value)>58). Fitted line $y=-0.97x-0.12$ with (-1.14, -0.7947) and (-0.1875, -0.05864) 95% confidence interval for each parameter. (B) The mutant measurements are plotted on top of the wild type. The negative correlation observed in the wild type, reflected by the black least straight line, is almost completely eliminated, displayed in the green least square line ($R=-0.06$, $-\log_{10}$(p-value)<6). Fitted line $y=0.1852x+0.27$ with (-0.1723, 0.5426) and (0.1596, 0.3805) confidence interval for each parameter. Also the width of the distribution across the x-axis is slightly narrower for the mutant strain an indication for reduced ability to modulate mRNA stability. (C) Fold enrichment for the change in stability in induced and repressed genes. The number of observed stabilized and destabilized genes within both induced and repressed groups of genes divided by the expected number, assuming no correlation. Expected number is calculated as the percentage of stabilized/destabilized genes in the genome times the induced/repressed group size.
doi:10.1371/journal.pgen.1002273.g002

It is important to mention, that actually the fastest responding and relaxing induced genes show the least degree of change due to the mutation (cluster 2 in Figure 3). This, at- first- surprising, observation could be explained by the nature of the mutation which only reduces, but does not eliminate, the recruitment of Rpb4/7 to transcripts. According to this possibility the fastest responding genes have the highest ability to recruit Rpb4/7, and hence is their ability to still recruit these two subunits in the mutant too. Clearly an alternative explanation might be that for these genes the transient behavior is independent of Rpb4/7 (see also Discussion). We performed a GO enrichment analysis to characterize the functional association of the genes in each of the clusters shown in Figure 3 (Table S1). Reassuringly we find a typical response to stress by which stress-related genes are induced while growth-related genes are repressed.

In order to show that reduced ability to modulate mRNA decay in the mutant is indeed associated with, and perhaps even causative of, the observed shift in kinetics, we compared for each cluster the change in stability in response to stress between the two strains (right panels of Figure 3). A slope smaller than unity, in the linear fit, represent an impaired ability to modulate stability in the mutant compared to the wild-type. Reassuringly, comparison of the two induced clusters shows a significantly reduced slope for the cluster that shows the difference in mRNA abundance temporal profiles between the strains. This result shows how the reduced ability of the mutant to destabilize induced genes might in part be responsible for the shift of these genes from spiked to slower response and relaxation kinetics. A reduced change in mRNA stability is also observed in the cluster that shows similar mRNA abundance profiles between the wild-type and the mutant. In this case, in addition to the above explanations, compromised transcription might compensate for the change in stability.

We also experimented with an additional stress, the DNA damaging agent MMS, which in our earlier work showed sustained and long enduring expression profiles [9]. In the present experiment the mutant and the wild-type showed very similar expression profile, with only a small reduction in the magnitude of the response in the mutant (not shown, and data deposited publicly, see Materials and Methods). This observation could indicate that the long enduring response displayed in this condition is independent of a coupling mechanism between production and degradation (or at least that it is not dependent on Rpb4/7 in that condition). We thus decided not to carry out decay experiments in the mutant under this condition.

Reduced recruitment of Rpb4/7 in the mutant results in a reduction in both production and degradation in basal conditions

A difference between the two strains can also be observed at the reference un-stressed conditions. To study the basal difference in mRNA abundance between the two strains in optimal conditions the mRNA abundance levels at time point zero (before addition of the stress) in the two strains where normalized using spiked-in RNA internal standard. A unique normalization procedure is required here because standard microarray normalization methods assume a constant global distribution of intensity values between samples preserving ranking differences between genes but eliminating global differences (see Materials and Methods). While gene expression levels strongly correlate between the two stains, we observe an overall reduction in mRNA levels in the mutant vs. the wild-type, as expected from a mutant in Pol II (Figure 4A). This observation extends previous reports of reduced production in this strain [27], showing this effect on a genome wide level.

We next asked how the difference in steady state mRNA abundance relates to differences in stability by plotting the two parameters against each other. Figure 4B shows how genes with reduced mRNA abundance in the mutant actually become more stable, opposite of a potentially intuitive relationship but in line with the combined effect of Rpb4/7 on both production and degradation [19,21]. The large reduction in mRNA abundance, despite the elevation in stability, indicates that a larger decrease in production is taking place - otherwise the net decrease in mRNA levels would not be observed. Genes displaying increased destabilization due to the mutation are expected to show a

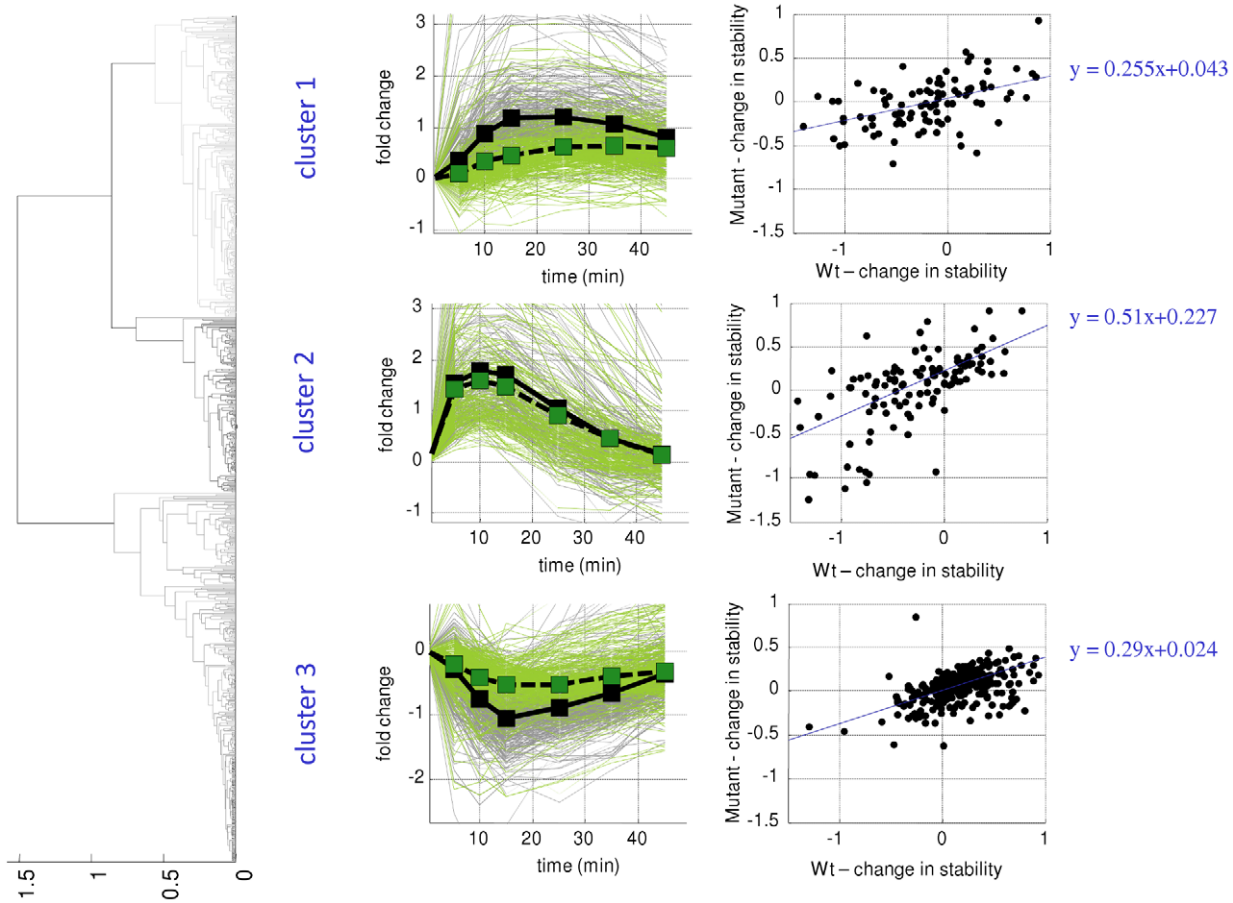


Figure 3. Impaired response of Rbp6 mutant strain. Hierarchical clustering of mRNA abundance temporal profiles. The three clusters marked in the dendrogram by different gray colors correspond to the three clusters in the middle and right panels. Middle panels show the fold change as a function of time for both strains, gray for wild type and green for mutant, with thick lines representing the mean of each strain. Left panels show the difference in stability in response to stress, calculated as the log2 ratio of the stress half-life by the reference half-life. The wild type value is plotted against the mutant with a blue least square line showing the general trend for each cluster. To show the slope difference between the two induced clusters fit parameters are also displayed for all clusters.
doi:10.1371/journal.pgen.1002273.g003

reduction in mRNA abundance which is not observed in this plot. The reason resides in the normalization procedure which artificially centers the differences in stability around zero, thus correcting a probable shift of the values towards less degradation in the mutant (see Materials and Methods).

We thus conclude from this plot that the genes which are most affected by the mutation at optimal conditions are the genes which show a decrease in mRNA abundance coupled with stabilization. We find that those genes are highly enriched with ribosome biogenesis GO functionalities and other specific functional groups (Figure 4C), in line with previous observations relating stability of mRNAs encoding ribosome biogenesis factors with rpb4/7 functions [20]. This observation goes along with a slower growth rate for this mutant of about 40% relative to the wild-type.

Discussion

Previous studies of Rbp4/7 proposed for the first time a biochemical mechanism in which Pol II affects mRNA degradation by imprinting mRNA transcripts with two of the polymerase subunits that escort mRNAs to the cytoplasm [28]. In this work we show that the process of coupling transcription and mRNA decay is a genome-wide, regulated process that depends on the physiology of the cell and on the environment. Further, we show that this mechanism shapes the temporal kinetic response of mRNA abundance to changing external conditions.

To explain how Rpb4/7 has a genome-wide effect on mRNA stability, and in opposite directions for induced and repressed genes, we propose a simple molecular model, described schematically in Figure 5. Previous molecular work shows that an mRNA molecule can be exported into the cytoplasm either associated (imprinted) or not associated with Rpb4/7 and that imprinting will result in an increased probability to be degraded [19,21]. Let us assume that the general factors responsible for mRNA decay in the cytoplasm are in limited amount and are distributed across the transcripts according to each transcript's ability to recruit these factors. Because Rpb4/7 imprinted mRNAs have increased ability to recruit general decay machinery, the distribution of the general decay machinery across different genes will be in proportion to the fraction of Rpb4/7-imprinted mRNA molecule copies for each gene. Assume that in basal conditions we have some probability (e.g. 0.3) for an mRNA to be exported into the cytoplasm imprinted by Rpb4/7. This will give rise to a similar percentage of Rpb4/7 imprinted mRNA molecules per gene in the cytoplasm which, assuming no other effects, will result in a uniform distribution of decay resources over all genes (Figure 5A).

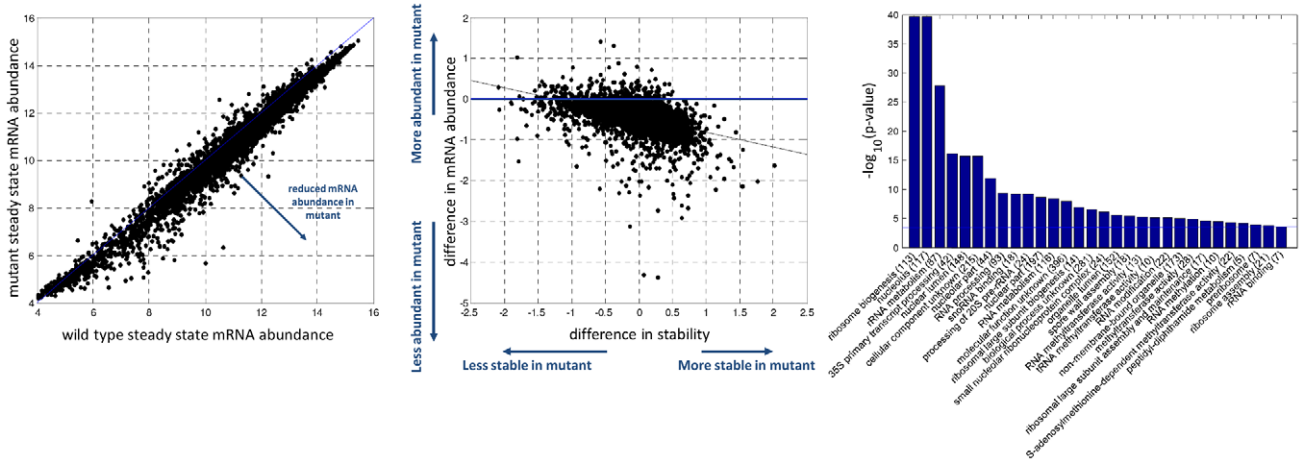


Figure 4. Wild type and mutant differences in basal conditions. (A) correlation of steady state mRNA abundance measurements between the wild type and mutant strain. A global reduction in mRNA levels is observed in the mutant. Data is plotted in log2 scale. (B) Basal difference in stability (log 2 ratio of the reference half-lives of the mutant divided by the wild type) is plotted against the difference in mRNA abundance (log 2 ratio of basal mRNA abundance in the WT and mutant). Most changing genes show both a reduction and stabilization in the mutant strain. The black least square line shows the negative correlation between these two parameters. Fitted line $y = -0.3986x - 0.4557$ with $(-0.4183, -0.3789)$ and $(-0.4636, -0.4477)$ 95% confidence intervals for each parameter. Blue line represents no change in mRNA abundance. (C) Results of a gene ontology enrichment analysis for the genes that show the largest reduction in mRNA abundance in the mutant strain.
doi:10.1371/journal.pgen.1002273.g004

Now let us assume that in stress the probability of imprinting by Rpb4/7 is increased (e.g. $p = 0.8$) per individual transcript molecule. This can be easily achieved by increasing the concentration of any cofactor contributing to Rpb4/7 imprinting. Following this change, genes which are induced due to the stress will quickly increase the percentage of imprinted mRNA molecules, thus gaining a higher affinity to general decay machinery which will result in faster decay for these genes (gene A in Figure 5B). Such genes would usually be also present in lower copies in basal conditions as depicted in Figure 5A. Genes which are not induced at the transcription level in response to the stress, yet are still transcribed at lower rates, will slowly increase the percentage of imprinted mRNA molecules resulting in intermediate decay (gene B). Repressed genes, for whose transcription is decreased, will keep the basal percentage of imprinted molecules and will even experience a decrease in this percentage due to the faster decay of imprinted molecules, which will result in general slow decay for these genes and a transient decrease in numbers. Thus, by only increasing the probability of Rpb4/7 imprinted export we get redistribution of the general decay machinery in favor of induced genes. Increasing it during a response to stress will result in coupling between mRNA production and degradation and a genome wide fast transient adaptive response. In support of this model, previous works have indeed shown increased Rpb4/7 mediated export in stressful conditions [29,30].

While a lot is known about the control of each stage of gene expression in isolation, recent accumulating knowledge suggests that many of the stages are coupled [31–33]. Using the rpb6 mutant it was shown that reduced ability to recruit rpb4/7 to the core polymerase results in impaired production and decay for selected genes [19–21]. Here we show that this single amino acid substitution in Pol II, which likely decouples transcription from mRNA decay, has a genome wide effect on gene expression both under optimal conditions and in stress. This mutant is defective both in mRNA synthesis and decay and has a reduced ability to modulate mRNA decay under stress conditions. The impaired ability to affect production and degradation is reflected in the mRNA abundance temporal response to stress by which most responsive genes, induced and repressed, show a difference in the magnitude and the kinetics of the response – towards lower fold change difference and slower relaxing kinetics. While a genome-wide decrease in production due to a mutation in Pol II is expected, the large effect that this mutation has also on mRNA stability and the loss of correlation between the changes in stability and production is intriguing. This, together with the accumulating molecular evidence [19–21], indicates that Rpb4/7 not only affects different regulation levels, as production and degradation, but also serves to coordinate between these levels as has been recently proposed [22]. We note though that we cannot exclude an alternative in which due to a defect of the mutant in transcription the mutant does not transcribe machinery that is required for normal decay of mRNAs.

It should be noted that while most responding genes were affected by the mutation, some did not, especially the most fast-relaxing transient genes, genes that also show a strong counter-action between mRNA abundance levels and decay (Figure 3). This could indicate that these genes are not clients of the *RPB4* and *RPB7* chaperoning mechanism and may owe their strong counter action and fast response to another mechanism. Alternatively the lack of sensitivity of the most rapidly responding genes to the mutation could simply indicate that the Q100R mutation did not abolish the normal dynamics of these genes, presumably because these genes represent the strongest-affinity clients of this coupling machinery, this possibility should be examined experimentally for individual genes.

Tirosh et al. in a recently published paper [34], compared mRNA abundance and stability between two yeast species and interestingly found a similar result by which differences in mRNA abundance were accompanied by opposite differences in stability, e.g. increase in mRNA abundance in one species was accompanied by decrease in stability in the same species. This might suggests that the negative correlation that we observe in transient responses to stress is part of a more universal feedback mechanism by which changes in stability compensate for changes in production.

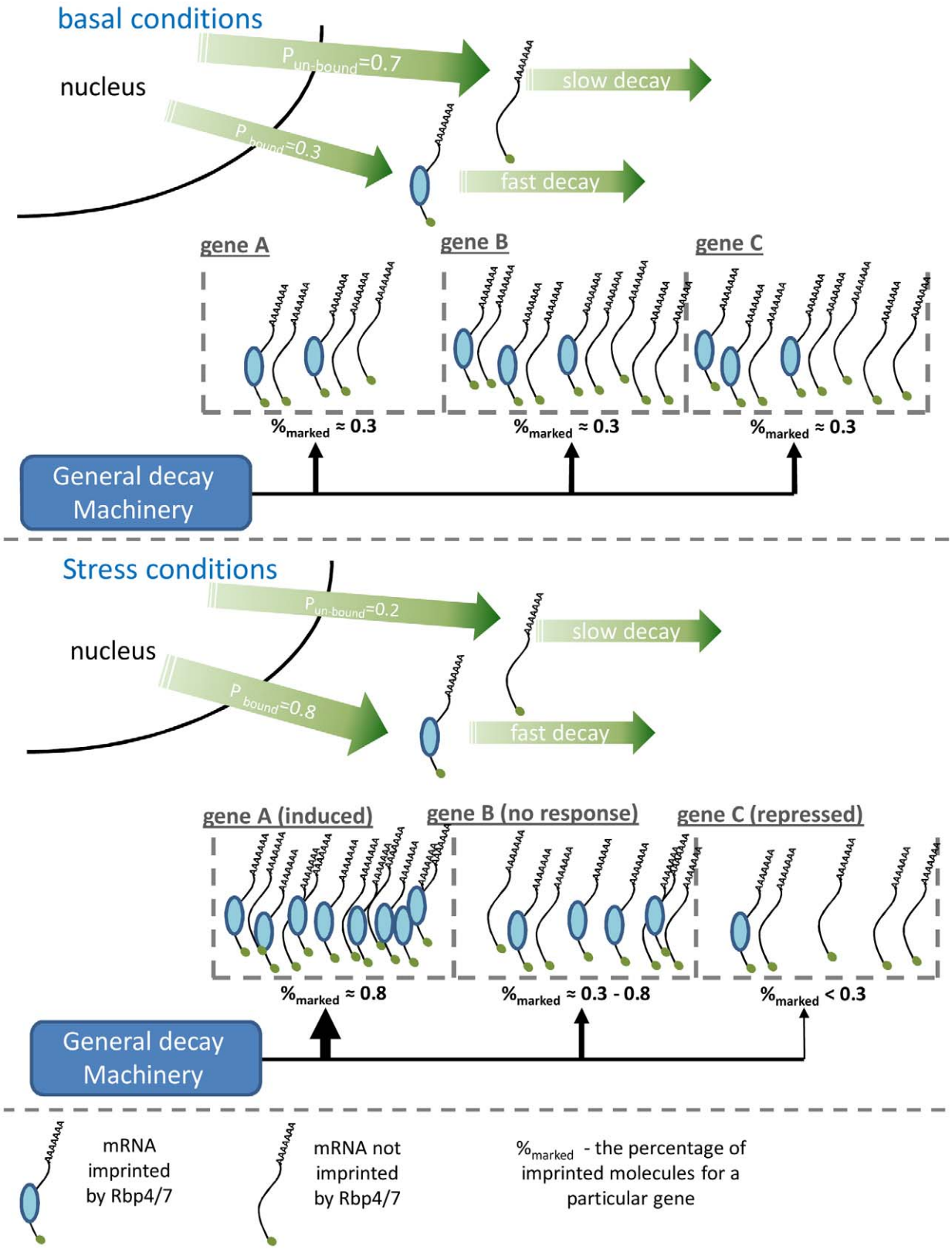


Figure 5. A schematic illustration of the suggested model explaining the genome-wide effect of Rpb4/7. The figure shows how an increase in the association probability following stress causes a global redistribution of the general decay machinery resulting in genome wide coupling of changes in mRNA abundance and stability. Theoretical probability for a transcript to be either imprinted or non-imprinted by Rpb4/7 during transcription in basal conditions (A) and for stress (B). A more detailed explanation is given in the main text.
doi:10.1371/journal.pgen.1002273.g005

Coupling of transcription and post-transcriptional regulation, and in particular by a counter action mode, is now known to be obtained by additional completely different mechanisms. One interesting mode of coupling is based on coordinated regulation by transcription factors and microRNAs which regulate shared targets [35,36], which may allow to combine the activation of a gene with its inhibition. Thus completely different "hardware" may implement the same "software" of counter-acting regulation. Inspection of kinetics in such microRNA-transcription factor combined regulation even reveals a similar effect of the mode of coupling and the response behavior [37].

We suggest a simple molecular model which is sufficient to explain how changes in mRNA abundance are negatively coupled at the genome-wide level to change in stability in response to stress. This model is especially appealing because it shows how by a very simple manipulation, namely increasing the general imprinting probability of Rpb4/7 to transcripts, cells can achieve a genome-wide re- distribution of the general decay machinery which will result in fast relaxation of the mRNA abundance response and faster response time. Such a fast relaxation, which for induced genes, results from the coordinated increase in production and degradation, might probably lead to faster relaxation than what would be possible by increasing and then decreasing transcription, while keeping degradation constant. Our model suggests that complicated cellular responses, such as genome-wide transient mRNA changes at different temporal dynamics, can actually be controlled by simple manipulations of global cellular parameters. Of course the exact molecular details of such model remain to be validated.

Materials and Methods

Strains and growth conditions

All experiments were carried out using the two strains, The *rpb6*^{Q100R} strain (*MATalpha ura3-52*, *his3delta200*, *lys2delta201*, *ade2*, *RPB6delta::HIS3* p*RPB6/CEN/LEU2*) [26] and its isogenic parental WT strain (WY37) (*(MATalpha ura3-52*, *his3delta200*, *lys2-delta201*, *ade2*). Two types of experiments were conducted: experiments measuring mRNA abundance and experiments measuring mRNA decay. Both types of experiments were carried out for both strains (Figure 1 for a detailed description). For all experiments cells were grown in YPD medium (2% yeast extract, 1% peptone, 1% dextrose) at 30°C to the concentration of $1*10^7$ cells/ml. Due to a slower growth rate of the mutant, of about 40%, starters for the two strains started with slightly different concentration to ensure that both strain experiments start at the same cell concentration. To measure response to an environmental stress cells were then treated with 0.3 mM Hydrogen peroxide (H_2O_2) (Frutarom LTD.), or 0.1% MMS (Sigma-Aldrich) (data not shown). Survival experiments were also performed resulting in similar survival rates for both strains. For mRNA abundance measurements, aliquots (15 ml) were removed following each treatment, in the following time points: 0, 5, 10, 15, 25, 35 and 45 minutes and frozen in liquid nitrogen. RNA was extracted using MasterPure (EPICENTER Biotechnologies). The quality of the RNA was assessed using the BIOANALYZER 2100 platform (AGILENT); the samples were then processed and hybridized to Affymetrix yeast 2.0 microarrays using the Affymetrix GeneChip system according to manufacture instructions. For measuring mRNA decay a similar protocol was applied albeit with the following modifications: transcription was inhibited by adding 1,10-phenanthroline (before each experiment we prepared a fresh stock of 40 mg/ml in ethanol which was then diluted to a working concentration of 100 ug/ml) to the media 7 minutes following the addition of Hydrogen peroxide. Then aliquots were removed in the following time points: 0, 5, 10, 20, 30, 40 and 50 and were processed and hybridized as previously described. In addition we added a complete biological replicate for all mutant measurements, including half-lives in reference and oxidative stress and mRNA abundance changes in response to oxidative stress. The replicate reproduces well the presented results (Figures S2, S3, S4, S5).

Microarray processing, determination of half-life, and mRNA abundance fold changes

All microarray samples were processed with the RMA preprocessing algorithm. In this paper unlike the work presented in our previous paper [9], transcription inhibition was achieved using the drug 1,10-phenanthroline which inhibits transcription not just from Pol II but also from Pol I, and Pol III . Thus the spike-in normalization procedure, which was used in previous decay experiments [2,9] could not be used in this case. To get a decay profile for each gene we assumed that the global mRNA abundance decays similarly between strains and conditions and our previous measurements and scaled the data accordingly. In detail, we used as a reference the mean decay profile over the entire genome from our previous decay measurements [9]. Then each time point in the current experiment was scaled such that mean intensity over the time course would decay according to the expected reference profile. Data were then zero transformed to the first time point and exponential fit was performed on these normalized profiles. If the data was not scaled, then standard normalization procedures would result in an mRNA abundance increase for the more stable genes and a decrease for the less stable genes along the time course. It is important to note that if the slope of this change would be taken as a proxy to the genes relative stability, still the results presented in this paper will qualitatively remain the same. Also, although both of these procedures will not detect a global difference in the genome decay between the two strains, the results presented in the paper would not change because they are based on the relative rank of each gene.

Following scaling, assuming a constant decay rate throughout the course of the experiment, decay profiles were fitted to a first order exponential decay model, $y(t)=y(0)\cdot e^{-kt}$, from which the fitted decay constant k was used to calculate a gene specific half-life in each condition and strain, $t_{1/2}=\ln(2)/k$. Only genes for which a high goodness of fit is achieved (R-square>0.9) were taken for further analysis. About 90% of the genes in each strain and each condition display an R-square above this cutoff indicating that this model is a good approximation to describe the decay kinetics in this data set.

For the mRNA abundance profiles, responsive genes were defined as having an absolute fold change of above 1.75 for at least one time point out of the wild type mRNA abundance measurements. These measurements were then interpolated to get a continuous response profile using a cubic spline. The maximal point was taken as the point where both the spline derivative was equal to zero and the fitted spline value reached the maximal absolute value (maximal for induced genes and minimal for repressed genes). The qualitative results presented remained robust to different response and R-square cutoffs.

For clustering the responsive mRNA abundance profiles the wild type and mutant profiles were concatenated and clustered. Distance matrix between genes was calculated using spearman rank correlation which was followed by average linkage clustering. Separation to the three clusters was done by eye using the presented dendrogram (MATLAB implementation).

All data is available at the GEO public data base under accession GSE26829.

Microarray preprocessing for basal mRNA abundance comparisons

Standard mRNA abundance preprocessing algorithms assume a constant global distribution of mRNA levels across samples, thus any global differences between samples, e.g. an increase in the genome mean mRNA level, would not be detected. In order to compare basal mRNA abundance levels between the two strains, we normalized two zero time points, of the reference decay time courses, using spiked in RNA which followed the standard RMA normalization. We used the same normalization method described previously to normalize decay experiments[9]. Using this method mRNA levels are scaled according to an internal standard, thus a global change in the mean mRNA level would be detected. This normalization is valid here because Rpb4 and 7 are specific to Pol II which is responsible for the production of a very small amount (2–5%) of the total RNA in the cell. Thus when RNA is extracted a global difference in mRNA abundance level will be reflected in the ratio between the internal standard and the mean microarray intensity signal. As explained above, decay measurements using 1,10-phenanthroline are not suitable to such normalization because the effect is not specific to Pol II. We thus normalize all decay data sets to have the same global decay rate which results in a distribution of decay differences which is centered on zero as shown in Figure 4B. As a result, global differences between strains or conditions, as a general reduction in degradation capacity in the mutant, would be missed. Yet, the results presented in this paper will not change due to any global shifts in half-life distributions in any of the data sets.

Supporting Information

Figure S1 Standard deviation normalized mRNA abundance profiles of cluster 1. mRNA abundance profiles from cluster 1 of Figure 3. Data is divided by the standard deviation for each gene such that the resulting profiles would have the same standard deviation. Such normalization eliminates differences in magnitude to emphasize the difference in kinetics between the two strains. Wild type mean is shown in black and the mutant in green.
(TIF)

Figure S2 Biological replicate for Figure 2B and 2C. Repetition of Figure 2B and 2C using an independent biological replicate for the mutant measurements. Data for the wild type (right panel) is taken from the first measurements.
(TIF)

Figure S3 Repetition of Figure 4A. Repetition of Figure 4A using an independent biological replicate for the mutant measurements.
(TIF)

Figure S4 Repetition of Figure 4B. Repetition of Figure 4B using an independent biological replicate for the mutant measurements.
(TIF)

Figure S5 Mutant half-life comparisons between repeats. Correlation between the half-live measurements between the two mutant biological replicates.
(TIF)

Table S1 GO enrichment analysis for clusters in Figure 3. GO categories which passed hyper geometric test with FDR ($q = 0.05$) for the three clusters in Figure 3.
(XLS)

Acknowledgments

We thank the Pilpel lab and Itay Tirosh for stimulating discussion and Reut Shalgi for critically reading the MS. We thank Dr, Shirley Horn-Saban of the DNA chip unit at the Weizmann Institute for devoted and professional service.

Author Contributions

Conceived and designed the experiments: OS BG MC OD YP. Performed the experiments: OS BG. Analyzed the data: OS BG. Contributed reagents/materials/analysis tools: OS BG MC OD YP. Wrote the paper: OS OD YP.

References

1. Grigull J, Mnaimneh S, Pootoolal J, Robinson MD, Hughes TR (2004) Genome-wide analysis of mRNA stability using transcription inhibitors and microarrays reveals posttranscriptional control of ribosome biogenesis factors. Molecular and cellular biology 24: 5534–47.
2. Wang Y, Liu CL, Storey JD, Tibshirani RJ, Herschlag D, et al. (2002) Precision and functional specificity in mRNA decay. Proceedings of the National Academy of Sciences of the United States of America 99: 5860–5.
3. Bernstein JA, Khodursky AB, Lin P-H, Lin-Chao S, Cohen SN (2002) Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. Proceedings of the National Academy of Sciences of the United States of America 99: 9697–702.
4. Narsai R, Howell KA, Millar AH, O'Toole N, Small I, et al. (2007) Genome-wide analysis of mRNA decay rates and their determinants in Arabidopsis thaliana. The Plant cell 19: 3418–36.
5. García-Martínez J, Aranda A, Pérez-Ortín JE (2004) Genomic run-on evaluates transcription rates for all yeast genes and identifies gene regulatory mechanisms. Molecular cell 15: 303–13.
6. Pelechano V, Chávez S, Pérez-Ortín JE (2010) A complete set of nascent transcription rates for yeast genes. PloS ONE 5: e15442. doi:10.1371/journal.pone.0015442.
7. Miller C, Schwalb B, Maier K, Schulz D, Dümcke S, et al. (2011) Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. Molecular systems biology 7: 458.
8. Molina-Navarro MM, Castells-Roca L, Bellí G, García-Martínez J, Marín-Navarro J, et al. (2008) Comprehensive transcriptional analysis of the oxidative response in yeast. The Journal of biological chemistry 283: 17908–18.
9. Shalem O, Dahan O, Levo M, Martinez MR, Furman I, et al. (2008) Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation. Molecular systems biology 4: 223.
10. Molin C, Jauhiainen A, Warringer J, Nerman O, Sunnerhagen P (2009) mRNA stability changes precede changes in steady-state mRNA amounts during hyperosmotic stress. RNA (New York, N.Y.) 15: 600–14.
11. Fan J, Yang X, Wang W, Wood WH, Becker KG, et al. (2002) Global analysis of stress-regulated mRNA turnover by using cDNA arrays. Proceedings of the National Academy of Sciences of the United States of America 99: 10611–6.
12. Foat BC, Houshmandi SS, Olivas WM, Bussemaker HJ (2005) Profiling condition-specific, genome-wide regulation of mRNA stability in yeast. Proceedings of the National Academy of Sciences of the United States of America 102: 17675–80.
13. Amorim MJ, Cotobal C, Duncan C, Mata J (2010) Global coordination of transcriptional control and mRNA decay during cellular differentiation. Molecular systems biology 6: 380.
14. Elkon R, Zlotorynski E, Zeller KI, Agami R (2010) Major role for mRNA stability in shaping the kinetics of gene induction. BMC genomics 11: 259.
15. Pérez-Ortín JE, Alepuz PM, Moreno J (2007) Genomics and gene transcription kinetics in yeast. Trends in genetics: TIG 23: 250–7.
16. Romero-Santacreu L, Moreno J, Pérez-Ortín JE, Alepuz P (2009) Specific and global regulation of mRNA stability during osmotic stress in Saccharomyces cerevisiae. RNA (New York, N.Y.) 15: 1110–20.
17. Jasiak AJ, Hartmann H, Karakasili E, Kalocsay M, Flatley A, et al. (2008) Genome-associated RNA polymerase II includes the dissociable Rpb4/7 subcomplex. The Journal of biological chemistry 283: 26423–7.
18. Pillai B, Verma J, Abraham A, Francis P, Kumar Y, et al. (2003) Whole genome expression profiles of yeast RNA polymerase II core subunit, Rpb4, in stress and nonstress conditions. The Journal of biological chemistry 278: 3339–46.
19. Goler-Baron V, Selitrennik M, Barkai O, Haimovich G, Lotan R, et al. (2008) Transcription in the nucleus and mRNA decay in the cytoplasm are coupled processes. Genes & development 22: 2022–7.

20. Lotan R, Bar-On VG, Harel-Sharvit L, Duek L, Melamed D, et al. (2005) The RNA polymerase II subunit Rpb4p mediates decay of a specific class of mRNAs. Genes & development 19: 3004–16.
21. Lotan R, Goler-Baron V, Duek L, Haimovich G, Choder M (2007) The Rpb7p subunit of yeast RNA polymerase II plays roles in the two major cytoplasmic mRNA decay mechanisms. The Journal of cell biology 178: 1133–43.
22. Harel-Sharvit L, Eldad N, Haimovich G, Barkai O, Duek L, et al. (2010) RNA polymerase II subunits link transcription and mRNA decay to translation. Cell 143: 552–63.
23. Bushnell DA, Kornberg RD (2003) Complete, 12-subunit RNA polymerase II at 4.1-A resolution: implications for the initiation of transcription. Proceedings of the National Academy of Sciences of the United States of America 100: 6969–73.
24. Ujvári A, Luse DS (2006) RNA emerging from the active site of RNA polymerase II interacts with the Rpb7 subunit. Nature structural & molecular biology 13: 49–54.
25. Runner VM, Podolny V, Buratowski S (2008) The Rpb4 subunit of RNA polymerase II contributes to cotranscriptional recruitment of 3′ processing factors. Molecular and cellular biology 28: 1883–91.
26. Tan Q, Prysak MH, Woychik NA (2003) Loss of the Rpb4/Rpb7 subcomplex in a mutant form of the Rpb6 subunit shared by RNA polymerases I, II, and III. Molecular and cellular biology 23: 3329–38.
27. Sheffer A, Varon M, Choder M (1999) Rpb7 can interact with RNA polymerase II and support transcription during some stresses independently of Rpb4. Molecular and cellular biology 19: 2672–80.
28. Choder M (2011) mRNA imprinting: Additional level in the regulation of gene expression. Cellular logistics 1: 37–40.
29. Selitrennik M, Duek L, Lotan R, Choder M (2006) Nucleocytoplasmic shuttling of the Rpb4p and Rpb7p subunits of Saccharomyces cerevisiae RNA polymerase II by two pathways. Eukaryotic cell 5: 2092–103.
30. Farago M, Nahari T, Hammel C, Cole CN, Choder M (2003) Rpb4p, a subunit of RNA polymerase II, mediates mRNA export during stress. Molecular biology of the cell 14: 2744–55.
31. Halbeisen RE, Gerber AP (2009) Stress-Dependent Coordination of Transcriptome and Translatome in Yeast. PLoS Biol 7: e1000105. doi:10.1371/journal.pbio.1000105.
32. Proshkin S, Rahmouni AR, Mironov A, Nudler E (2010) Cooperation between translating ribosomes and RNA polymerase in transcription elongation. Science (New York, N.Y.) 328: 504–8.
33. Shen Z, St-Denis A, Chartrand P (2010) Cotranscriptional recruitment of She2p by RNA pol II elongation factor Spt4-Spt5/DSIF promotes mRNA localization to the yeast bud. Genes & development 24: 1914–26.
34. Dori-Bachash M, Shema E, Tirosh I (2011) Coupled Evolution of Transcription and mRNA Degradation. PLoS Biol 9: e1001106. doi:10.1371/journal.pbio.1001106.
35. Shalgi R, Lieber D, Oren M, Pilpel Y (2007) Global and local architecture of the mammalian microRNA-transcription factor regulatory network. PLoS Comp Biol 3: e131. doi:10.1371/journal.pcbi.0030131.
36. Tsang JS, Ebert MS, van Oudenaarden A (2010) Genome-wide dissection of microRNA functions and cotargeting networks using gene set signatures. Molecular cell 38: 140–53.
37. Avraham R, Sas-Chen A, Manor O, Steinfeld I, Shalgi R, et al. (2010) EGF decreases the abundance of microRNAs that restrain oncogenic transcription factors. Science signaling 3: ra43.

The effect of 3'UTR sequences on gene expression in yeast

Ophir Shalem, Danny Zeevi, Eilon Sharon, Leeat Keren, Adina Weinberger, Orna Dahan, Yitzhak Pilpel* and Eran Segal*

Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100 Israel
Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 7610 Israel

*Corresponding author at the above address

December 12, 11

Abstract:

A full understanding of gene regulation requires an understanding of the contributions that the various regulatory regions has on gene expression. Here, we study the independent effect of 3' UTR regions on expression by constructing a library of yeast reporter strains that differ only by a 3' UTR sequence integrated downstream to a fluorescent reporter. Our library spans a continuous range of expression values of more than 10 fold with each 3' UTR having a unique and accurately reproducible effect on expression. These measurements correlate with endogenous mRNA levels, suggesting that 3' UTR sequences contribute to the determination of mRNA levels. Using sequence analysis, we found that increased A/T content upstream of the polyadenylation (3' UTR end) site correlates with higher expression, highlighting the importance of this genomic region and suggesting that differences in transcription termination efficiencies might be partly responsible for the observed effect. Finally, using single cell fluorescence measurements, we show that the dynamics by which different 3'UTR sequences modulate protein expression is distinct from expression differences mediated by changes in promoter activation, suggesting that 3' UTR sequences affect expression by changing protein burst size and not the frequency at which bursts occur. Together, our study demonstrates the role of synthetic genetic libraries in complementing information that originates from intact genome analyses in deciphering the codes that govern gene expression.

Introduction

Studies aimed at understanding the determinants of gene expression have traditionally focused on promoter sequences. However, regulatory information is also encoded in other genomic regions such as the 5' and 3' untranslated regions (UTRs) and may even be embedded within the coding regions themselves[1-3]. Since measurements of endogenous expression levels of mRNAs[4-9] and proteins[10-12] represent the net effect of all regulatory regions and regulatory layers (e.g., transcription, translation and degradation), it is difficult to use such data to dissect the relative contribution of every genomic region to the overall measured levels. Thus, if the expression level of one gene is higher than another, we cannot tell which regulatory region or combination thereof causes this behavior. The situation becomes even more complicated when considering the recent observations that suggest that the different regulatory layers often affect each other[13-20]. In the context of transcription initiation, the challenge of deciphering the regulatory code that maps sequence into expression levels was addressed by separately fusing the promoter of different genes to a fluorescent protein reporter, integrating the resulting constructs into the same genomic location, and then comparing the levels of the reporters for different promoters[21-27]. Since strains for different genes in such synthetic libraries differ only in the promoter sequence that is fused to the reporter, this setting allows a direct comparison of the promoter effect across genes, providing important insights into cis-regulatory mechanisms and principles of promoter activation. However, mRNA and protein levels are also affected by post-transcriptional regulation and some of the sequence signals that govern such processes are present in the 3'UTRs of genes. Cases of 3' UTR mediated regulation in S. *cerevisiae* are known[28-30], but the general contribution of 3' UTR sequences to protein expression is unclear.

Here, we applied a similar approach to that taken for promoters to study the effect of 3' UTR regions on expression levels, by constructing a library of yeast strains that differ only by a 3'UTR sequence integrated immediately downstream to a yellow florescence protein (YFP) that serves as a reporter gene. Transcription of all genes in the library is driven by a constant promoter -the well-studied Gal1/10 promoter[31-33]. This

promoter was chosen as it can be induced to different levels by changing the concentration of Galactose in the growth media. Thus, at a given Galactose concentration, the contribution of the promoter region to the overall YFP level is the same across all strains such that any measured differences in YFP expression level between strains are attributable only to differences in their integrated 3' UTR sequences.

We constructed and measured the expression of 87 different 3' UTRs taken from metabolic and ribosomal protein yeast genes. We found that the effect of 3'UTR sequences on YFP expression spans more than 10 fold and that this effect correlates with measurements of mRNA abundance, suggesting that endogenous mRNA levels are determined in part by 3' UTRs. We unraveled some of the sequence features that may be responsible for the differential effect of the 3' UTRs and found evidence that these features may also account for a small yet significant fraction of the variability in genome-wide measurements of mRNA abundance. The region in the 3' UTR containing these features resides in close proximity to the polyadenylation site highlighting the importance of the nucleotide composition of this genomic region to protein levels. We thus suggest that transcription termination efficiency might vary between genes and exert a considerable effect on gene expression. Further characterization of these sequence features will be needed to prove this and also to identify other novel mechanisms through which 3' UTRs affect expression levels. This will pave the way to a more complete understanding of gene regulation that also incorporates the effect of 3' UTR regions. Finally, we also show that the dynamics by which changes across our library occur differ from changes mediated by promoter activation, with respect to the relationship between mean expression to noise measurements, suggesting that 3' UTRs affect protein expression after transcription initiation.

Results

Constructing a library of 3' UTR yeast reporter strains

To measure the effect of 3' UTR sequences on expression levels, we adapted an approach that we previously devised for measuring the transcriptional effect of promoter regions[27](**Fig. 1**). To this end, we engineered a single master strain to have a genomically integrated YFP reporter downstream to the Gal1/10 inducible promoter, and at the same genomic location, we also integrated an mCherry fluorescent protein reporter downstream to the promoter of TEF2, a transcriptional elongation factor. A truncated URA3 selection marker, without a promoter and a start codon, was also integrated downstream to the YFP gene to increase transformation specificity. Next, we used PCR to amplify 87 3' UTR regions from 49 ribosomal protein genes and 38 metabolic genes from the yeast genome, attached each selected region to the promoter and start codon of the URA3 selection marker using extension PCR, and separately integrated each of the resulting constructs into the master strain, downstream to the YFP reporter. Since transcription end sites are not accurately defined, we selected the genomic region amplified from each gene to be the region located between its translation end site and the end of its downstream neighboring gene up to 1kb.

We validated the resulting clones in several different ways. First, since the same constant promoter drives the mCherry reporter across all strains, its expression should be highly similar across strains, and we thus used it to filter out clones with global transcriptional and growth deficiencies. Second, we used colony PCR and sequencing to validate the integrity of the integrated sequences. Finally, for each 3' UTR, we measured the expression of three different clones, and only took clones for which at least two independent strain constructions yielded highly similar YFP measurements (**Fig. S1**).

Thus, aside from a possible effect that the region downstream to the selected 3' UTR may have on YFP transcription, differences in YFP levels across strains are solely attributable to differences in the 3' UTRs, since the YFP transcripts generated from all library strains are otherwise identical and are all driven by the same Gal1/10 promoter.

To test whether our experimental system can indeed be used to measure post-transcriptional effects, we cloned the 3' UTR of COX17, as its transcript is known to be destabilized by the Puf3p RNA binding protein through interaction of with the COX17

3'UTR[34-36]. As a reference, we also cloned the 3' UTRs of MFA2 and RPS6. We found that the COX17 3' UTR indeed resulted in significantly lower YFP levels compared to clones of MFA2 and RPS6. Further, the YFP accumulation profiles of several different clones for each of these three distinct 3' UTRs were highly similar to one another (**Fig. 2A**).

Native 3'UTRs span a wide and continuous range of expression values which correlate to endogenous mRNA abundances

We first examined the overall effect of the integrated 3'UTR sequences on YFP expression. To this end, we grew the library strains in a 96 well plate for a period of two days to deep stationary phase in Raffinose, a carbon source in which the Gal 1/10 promoter is inactive. We then inoculated a small volume of stationary cells into fresh media containing various concentrations of Galactose in order to activate the Gal1/10 promoter and induce YFP expression. We measured the growth (measured as optical density (OD)), mCherry and YFP levels over time using a robotically automated plate fluorometer in a 20 minute time resolution until all wells reached stationary phase. As a single value per induction curve, we took the amount of YFP fluorescence produced during the exponential growth phase divided by the integral of the OD curve during the same time period. This results in a measure of the average rate of YFP production per cell per second during exponential phase[27] (**Fig. 2D**).

As expected, the profiles of both OD and mCherry were highly similar across the library strains, such that any measured differences in YFP expression should indeed be attributable to the different 3'UTR regions integrated (**Fig. 2A,B**). Notably, in contrast to the highly similar OD and mCherry measurements across strains, we found that the YFP expression levels spanned a wide and continuous range of expression values (**Fig. 2C**) across several different Galactose concentrations (**Fig. 2D**). For example, at the highest Galactose concentration and thus Gal1/10 promoter induction, we found a 16-fold difference between the YFP levels between the highest and lowest library strains. Thus, our results demonstrate that 3' UTRs can greatly affect expression levels. Moreover, since our library represents only a small sample of 87 3' UTRs from the yeast genome, it

is likely that the effect of 3' UTRs on expression may be greater than the one we observed. We also checked whether the effect of our tested 3'UTRs depends on the specific growth condition or is constitutive and will be observed in other growth conditions as well by measuring the library in a few other media. We found a similar effect in a few other growth conditions, indicating that it is similar at least in the tested conditions (**Fig. S2**).

To check whether our measurements may explain endogenous mRNA levels, we compared the YFP expression of each 3'UTR library strain to the mRNA level of the respective endogenous gene measured by RNA-seq[4] (**Fig. 3B**). We find the correlation between these two measurements to be relatively high (R=0.34) given that we measure the effect of only the 3'UTR separated from its native genomic region and promoter. As a reference, we also compared the same RNA-Seq mRNA abundance values to measurements of a promoter YFP library[27] (and unpublished data) for the same genes (with a constant 3' UTR), and found, as expected, a higher correlation (**Fig 3C**, R=0.82).

3' UTRs with higher A/T content right upstream to the poly-adenylation site have higher expression in the library strains and in genome-wide mRNA abundance measurements

Since differences in the YFP expression levels of strains in our library are attributable to differences in the integrated 3'UTR sequences, we next sought to identify sequence features that may underlie the measured differences in YFP levels across our library. Unlike promoter sequences, where many regulatory elements such as transcription factor binding sites are known, relatively little is known about regulatory elements in 3' UTRs, and so far a few putative regulatory motifs were predicted[1,37]. We thus searched for general features that may explain the measured expression variability. We first examined general features of the 3' UTRs which included their length, enriched k-mers, and minimal free energy of the predicted secondary structure of the 3'UTR, but found no significant correlation between these measures and the measured YFP levels.

We next examined the G/C content of the entire 3' UTRs and found a low but significant negative correlation (R=-0.21) to the measured YFP levels. To further refine this correlation, we performed the same analysis but where the G/C content was computed from sliding windows of various lengths and various locations relative to the 3' UTR end (**Fig. S3**). Using this analysis, we found that the highest correlation between the YFP expression levels and the G/C content is generated from relatively short windows (~15-25bp) starting ~15-20 base pairs upstream to the poly adenylation site (3' UTR end) (R=-0.42) (**Fig. 4A,B**). Since higher G/C content generally results in more closed RNA secondary structures, we hypothesized that the lower G/C content of the more highly expressed strains may result in more open secondary structures that may allow more efficient translation, but a similar correlation analysis that uses predicted free energies instead of G/C content resulted in significantly lower correlations compared to our above analysis based solely on G/C content (**Fig. S4**).

To test whether G/C content may indeed be predictive of the transcriptional effect of 3' UTRs in the entire genome, and not just a result of the relatively small number of strains that we used in our analysis, we performed a similar analysis but using all yeast genes and correlating the G/C content of their 3' UTRs to genome-wide RNA-Seq measurements of mRNA abundance[4]. Since endogenous mRNA abundance levels represent the combined effect of all regulatory layers and not just that of 3' UTRs as in our library, we do not expect a high correlation even if 3' UTRs contain true regulatory elements that affect mRNA levels. Despite this caveat, we indeed found a low yet highly significant anti-correlation between the G/C content upstream of the transcription termination site and mRNA abundance at the genome-wide scale (R=-0.24, $P=10^{-45}$), but in contrast to the library this correlation is observed only when G/C content is calculated in larger windows of 70-90bp. (**Fig. 5A, Fig S5**) Thus, although we cannot pinpoint the exact region of the 3' UTR that exhibits the highest correlation, our results suggest a significant association between high expression and high A/T content in the vicinity of the transcription end site.

We next sought to check whether the A/T content signature hides an underlying, more refined, nucleotide composition signal that displays a higher correlation with expression values. First, we identified prominent sequence features around the poly adenylation site using genome-wide sequence data and then checked the correlation, in sliding windows, of these features with YFP or genome wide mRNA levels. We calculated nucleotide and di-nucleotide composition in sliding windows around the poly adenylation site and found a significant increase in both A and T composition but also in the AT or TA dinucleotide, as was previously noted[38-40] (**Fig S6**). We then ran a similar sliding window analysis calculating these features together with A/T content and plotted the correlation as a function of the window location (**Fig S7**). Here too, we found that of all of the above features, A/T content remains the most highly correlated feature with expression values both in the library and genome wide.

Other sequence features, which are frequently found right upstream of transcription termination sites, are two known transcription termination motifs, usually termed the "efficiency" and "positioning" elements[40-43]. We examined whether the occurrence of these motifs also correlates with expression and found a slight increase in the occurrence of these motifs in genes with high expression both in the library and genome wide (**Fig S8,9**). However, since these elements are AT rich, and were probably defined imperfectly, it is hard to tell whether this signal is distinct from the general correlation with higher AT content for highly expressed genes.

3' UTRs likely affect expression by changing the burst size and not the burst frequency

We next asked whether we could gain insights into the way in which 3' UTRs affect YFP expression. Previously protein expression was shown to occur in bursts which are characterized by frequency and size, i.e. number of bursts in unit time and number of proteins produced per burst[44-46]. We reasoned that an increase in YFP expression can arise from more frequent bursts or from a larger number of proteins produced from each burst, or from a combination thereof. For example, increasing the concentration of a transcription factor is expected to increase more the burst frequency than the size of

the burst in the activated promoter state. In contrast, since we expect the effect of 3' UTRs on YFP expression to be mediated by changes to YFP transcription termination or degradation rates, or by changes to YFP translation rates, we hypothesized that 3' UTRs should mainly affect the average size of protein bursts but not the burst frequency.

Under certain assumptions, burst frequency and burst size can be extracted from single cell protein expression measurements of an isogenic population[47,48]. We used this approach to measure the effect of 3' UTRs on burst frequency and burst size. For comparison, we also modulated transcription activation by varying galactose levels and measured changes in burst size and frequency for each gene across various activation levels. To this end, we used flow cytometry to obtain single cell YFP fluorescence measurements of all of our 87 strains under four increasing galactose concentrations, resulting in increasing Gal4 activity[49]. Thus, since all of our library strains are driven by the same Gal4 target promoter (Gal1/10), comparing the same strain across increasing galactose concentrations allows us to examine the effect of increasing expression by increasing transcription factor activity, while comparing the different strains to each other at a fixed galactose concentration allows us to examine the effect of the various 3' UTR.

We devised an automated pipeline to extract the mean and variance of a cell population from flow cytometry data, while controlling for variance in physiological cell parameters (see Methods). From these measurements, we could then compute the burst frequency of each strain at every galactose concentration as the inverse of the transcriptional noise (variance divided by mean squared), and the burst size as the noise strength (variance divided by mean)[48]. Notably, consistent with our hypothesis, we found that increasing expression by increasing galactose concentration and thus transcription factor concentration results in increased burst frequency but has little effect on burst size, while at any fixed galactose and thus fixed factor concentration, the expression of strains with different 3' UTRs is correlated with burst size while burst frequency remains constant (**Fig. 6**). Even if the assumptions under which our calculations correspond to burst size and frequency do not hold, it is clear that the two

different strategies that we examined here for changing expression, namely changing transcription factor activity or changing 3' UTRs, have opposing effects on noise and noise strength.

Discussion

To assess the quantitative effect of different genomic regions it is essential to establish experimental systems that separate these regions from their native genomic context and measure their direct effect. Here, we developed such an experimental procedure to build yeast reporter libraries that differ by a 3' UTR region, integrated downstream to an inducible YFP gene. Using this method we show that native 3' UTR sequences span a broad range of expression values of greater than 10-fold, with each 3' UTR displaying a unique and highly reproducible effect on protein expression. By correlating our measurements with endogenous mRNA levels, we suggest that endogenous mRNA levels are determined in part by information in the 3' UTR sequences.

Our measurements represent the sole effect of each strain's integrated 3' UTR sequence and thus serve as a good platform for sequence analysis. Indeed, we found a positive relationship between increased AT content upstream to the polyadenylation signal and YFP expression. Although we cannot say whether the AT content itself causes higher expression or whether it is a proxy for a more specific signal, our results highlight the 3'UTR end as a genomic region which may have a significant effect on mRNA levels. A lower yet highly significant similar correlation is also observed genome-wide, further supporting a role for AT content around 3'UTR ends in determining mRNA levels.

Our results raise the question of which molecular mechanism may be responsible for the observed correlation. Due to the proximity of the predictive area in the 3'UTR to the polyadenylation site and known termination motifs, we speculate that increased AT content may result in more efficient transcription termination. Efficienct termination could then give rise to elevated protein expression. Such an effect could be a result of either stronger AT rich transcription termination motifs[41,42] or an efficient

release of RNA polymerase after termination due to a looser elongating 8bp RNA/DNA duplex. Efficient release of the polymerase could results in more efficient recycling of the transcription initiation machinery for an additional round of transcription[50-53] within the current burst, hence is the effect of burst size. Additional potential means by which efficient transcription termination could give rise to higher protein expression comes from a recent work in mammalian cells[54], which suggested that with more efficient termination, more transcripts escape from nuclear surveillance, resulting in more mature mRNA molecules exported into the cytoplasm. The connection between improper termination and nuclear degradation that was recently shown[55] might also explain our results. We thus suggest that native 3'UTR sequences encode variable efficiency levels of transcription termination and that this variability is responsible in part for the observed variability in expression. However, we cannot exclude other potential explanations, such as effects on mRNA stability and translation.

Our study demonstrates the strength of a synthetic approach in establishing a causal link between sequence features and their outputs. Observing correlations in the genome, e.g. between sequence features in the 3' UTR and expression levels could always be explained by indirect non-causal effects. For example, one could argue that the genes with certain UTR features may also have strong promoters. Observing such connections in a setup such as the current library in which the effect of 3' UTR sequences is measured in isolation removes many such potential confounders.

Finally, we showed that the observed span of YFP values in our library, mediated by the different 3' UTR sequences affect population noise in a very distinct way compared to expression changes that are mediated by differential promoter activations. Our results thus put 3'UTR sequences as appealing candidates for the design of specific circuits in which changes in the mean expression level of a population are needed with little effects on noise.

Materials and Methods:

Construction of master strain and library strains

The first step in the construction of the 3'UTR synthetic library was to build a master strain into which the different 3'UTR sequences would be integrated. We built on a plasmid a construct containing mCherry under the control of TEF1 promoter and terminated by the ADH1 terminator sequence, and a non-terminated YFP gene under the control on the Gal1/10 promoter. Downstream of the YFP gene we also integrated a truncated URA3 gene lacking the promoter and start codon (see supplementary data for the exact sequence). The whole construct was then lifted from the plasmid by PCR and integrated into yeast strain Y8205 (courtesy of Charlie Boone) at the HIS deletion locus (chromosome 15, location 721987-722506bp). Desired downstream intergenic regions, containing the tested 3'UTRs were then lifted from the genome of BY4741 yeast strain. Forward primers had a 3'end matching the sequence starting immediately downstream to the ORF related to the tested 3'UTR and a 5'end matching the end of the YFP gene. We planned the library such that the integrated sequences will start from the stop codon till the next ORF (but up to 1kb if the intergenic region was longer). The reverse primers had a 3'end matching this location and a 5'end matching the URA3 promoter which was lifted from a plasmid with the start codon and a few base pairs into the URA3 promoter to match the sequence in the master strain (**Fig 1**). Following PCR amplification of the library from the genome the URA3 promoter and start codon were attached to the PCR products using extension PCR and the intact sequences were then integrated into the master strain using homologous recombination (REF) to create the library strains. All steps were automated and preformed in a 96 well plate except for the final plating and selection of the final clones which was done in 6 well agar (SCD-URA) plates.

From each transformation 3 clones where manually picked and grown on selective media (SC-URA). Clones for the final library were chosen under the following criteria: (1) At least two clones gave the exact same expression values (**Fig S1**). (2) integrated sequence length was validated by colony PCR. (3) OD and mCherry

measurements were used to ensure no growth of general transcription deficiencies. (4) A few selected inserts were validated by sequencing. Final library strains were grown in 96 well plates containing YPD as a growth medium for two days into deep stationary phase and frozen by adding glycerol to a final concentration of 25%.

Acquisition of bulk time course OD and florescence measurements

Cells were inoculated from stocks of -80^{o}C into SC + 2% Raffinose (180ul, 96 well plate) and left to grow at 30^{o}C for 48 hours, reaching deep stationary phase. Next, 5ul were passed into a fresh medium (175ul SC + 2% Raffinose) supplemented with the varying amounts of Galactose. Measurements were carried out every ~20 minutes using a robotic system (Tecan Freedom EVO) with a plate reader (Tecan Infinite F500). Each measurement included optical density (filter wavelengths 600nm, bandwidth 10nm), YFP fluorescence (excitation 500nm, emission 540nm, bandwidths 25/25nm accordingly) and mCherry fluorescence (excitation 570nm, emission 630nm, bandwidths 25/35nm accordingly). Measurements were replicated three times revealing high correlation between independent measurements. We also compared the measurements of the final library plate to the initial measurements of the different clones, representing same strain measured on a different geographical location within the 96-well plate, showing a minimal geographic effect on YFP measures.

Data processing and calculation of YFP production per cell

We used the same processing pipeline used by Zeevi at al.[27] to quantify promoter activity values from the exact same florescence and optical density measurements. Briefly, background levels are subtracted from OD, mCherry and YFP curves using media measurements, strain with only YFP reporter and strain with only mCherry reporter respectively. Next outlier removal was done on the measurement points which compose each individual curve, removing points which deviate considerably from their neighboring points. We validated by eye that all strains have same OD and mCherry curves (**Fig 2**). To calculate one value per induction curve an

automated procedure divides the OD curve into four growth phases: lag phase, exponential phase, linear phase and stationary phase. Then YFP accumulation was calculated by dividing the total amount of YFP produced during exponential phase by the integral of the OD curve during this time. Because both YFP and mCherry are very stable proteins, this measure represents the amount of YFP or mCherry produced per cell over this time course.

Flow cytometry measurements and extraction of transcriptional burst size and frequency measurements

Similar to the bulk measurements, cells were inoculated from stocks of -80^{o}C into SC + 2% Raffinose (180ul, 96 well plate) and left to grow at 30^{o}C for 48 hours, reaching deep stationary phase. Next, 5ul were passed into a fresh medium (175ul SC + 2% Raffinose) supplemented with the desired amount of galactose and grown with shaking at 30 ^{o}C . Six hours following dilution into a fresh medium containing galactose, plates were subjected to flow cytometry measurements using an LSRII flow cytometry machine supplements with an High Throughput Sampler (HTS) to measure 96 well plates.

To control for extrinsic variation we select a sub-population of cells with similar size and physiological status using an automatic gating procedure, which removes cells with spores or with outlier forward and side scatter parameters. Following gating, mean and standard deviation of YFP values was calculated for each strain in each galactose concentration and were used to calculate burst size and frequency following Friedman et al.[48] The general trends presented in this paper were robust to changing the gating parameters.

Figure legends:

Figure 1. Illustration of the master strain and library construction procedure. A master strain was constructed such that it will contain two main constructs in the HIS deletion

locus: a constant control construct with mCherry driven by the TEF2 promoter and terminated by a constant ADH1 terminator; and a test construct with a YFP gene driven by Gal1/10 promoter.Following master strain construction, a library of PCR products containing the downstream intergenic regions of 87 tested genes was amplified from the genome by PCR and extended to also contain the URA3 promoter and start codon. This library of DNA sequences was then integrated into the master strain such that only integrations in the exact genomic location would result in an intact selection marker. This created a library of yeast reporter strain that differ only by the 3'UTR region downstream to the YFP gene.

Figure 2. Controls and experimental variability of our system. (A) YFP measurements of clones with three different 3'UTRs. Shown are YFP measurements of three different strains, each with a unique 3' UTR sequence. Lines of the same color represent measurements of different clones from the same type of 3' UTR sequence, demonstrating that the effect of the different 3'UTRs on expression is above the variability of our experimental system. The lowest expressing strain (red) contains the COX17 3'UTR and serves as a positive control for our experimental system. **(B,C,D) Plate fluorometer measurements over time.** Following inoculation of the cells in a fresh media containing galactose, optical density (OD), mCherry and YFP are measured over time (B,C and D respectively). Note that as expected, OD and mCherry measurements remain highly similar between different library strains, while YFP expression varies considerably.

Figure 3. The effect of 3' UTRs on expression is large and is correlated with endogenous mRNA levels. (A) Dynamic range of YFP levels of library 3' UTR strains at different galactose induction levels. YFP production per cell per second was measured and calculated in different Galactose concentrations resulting in different promoter activation levels for all library strains at every galactose concentration. Shown are YFP measurements of the 3' UTR library strains. Note that the ratio between the highest and

lowest strain at the highest induction level (0.1% galactose) shows a fold difference of more than 10-fold. **(B)** Comparison of YFP levels and endogenous mRNA levels. Shown is a comparison of the YFP levels of our 3' UTR library (y-axis) and endogenous mRNA levels (x-axis) measured by RNA-Seq[4]. The Pearson correlation and associated p-value are given (inset). **(C)** Same as (B) but for a different YFP library in which different promoters are fused to a YFP reporter. Shown are promoters of the respective genes from which our 3' UTR library was constructed.

Figure 4. higher A/T content right upstream of 3' UTR end is associated with higher YFP expression. (A) Shown is the correlation (top graph) between the A/T content and the expression of the library 3' UTR strains as a function of the distance from the transcription end site (marked at 0bp). Each point represents the correlation between the average A/T content in a window of 20bp centered on that point. For each point, the bottom graph shows the (negative log) of the p-value of the associated correlation coefficient. **(B)** For the window of -20bp to -40bp from (A), shown is the actual G/C content of each library strain (x-axis) against its measured YFP production per cell per second. The blue line connects the mean of each G/C content bin and displays the standard deviation of the corresponding points of each bin.

Figure 5. Higher A/T content of the 3' UTR is associated with higher endogenous mRNA levels genome-wide. (A) Same as Fig. 4A, but using the 3' UTRs of all 3'UTRs in the yeast genome with mapped transcription end sites[6] and endogenous mRNA levels measured using RNA-seq[4]. Shown is the correlation (top graph) between the G/C content and the mRNA levels of all 3' UTRs as a function of the distance from the transcription end site (marked at 0bp). Each point represents the correlation between the average G/C content in a window of 80bp centered on that point. For each point, the bottom graph shows the (negative log) of the p-value of the associated correlation coefficient. **(B)** Shown (top graph) is the average G/C content of three sets of genes

groups by their mRNA expression levels (0.2 percentile of the lowest and highest expressing genes) as a function of the distance from mapped transcription end sites, in windows of 20bp centered around each point. Also shown is a histogram (bottom graph) of the expression values of each of the three sets.

Figure 6. YFP expression is correlated with noise strength. For several different galactose induction levels (represented by different colors), shown is the YFP expression of each 3' UTR library strain (y-axis) versus its noise strength (y-axis, left graph, expression variance divided by mean expression) and noise (y-axis, right graph, expression variance divided by mean squared). Each point represents the noise strength and noise computed from single cell flow cytometry measurements of the corresponding 3' UTR strain. Note the correlation between noise strength and YFP expression but not between noise and YFP expression.

Supplementary figure legends:

Figure S1. High reproducibility of our experimental system. Following the transformation of the 3'UTR sequences into the master strain, we selected three clones for each sequence, and measured their OD, mCherry and YFP values over time during a batch growth experiment. As part of the clone validation process we also compared the YFP levels (after 10 hours) of pairs of clones from the same transformation and selected only pairs of clones with very similar YFP expression (up to 15%) that thus display highly reproducible expression. Red dots mark clones that were not taken for the final library.

Figure S2. Expression of 3' UTR library is highly similar across condition. Shown is the YFP production per cell per second for several growth conditions (y-axis) against a reference growth condition (SC + 2% Galactose). The effect of different 3'UTRs remainshighly similar across the tested conditions.

Figure S3 Correlation between the AT content and YFP expression for different window sizes and different locations. Each row represents a window size and each column represents the beginning of the window continuing upstream into the 3'UTR. The red vertical line shows the 3'UTR end. Colors represent the Pearson correlation (upper panel) and –log10 of the p-value (lower panel). So the values in location (i,j) represent the correlation (or –log1(p-value)) between library YFP expression to the GC content in a window with coordinates [j-i+1 j] relative to the 3'UTR end.

Figure S4. Lower G/C content is more associated with YFP expression than is the folding energy. Same as Fig. 4A, but also including a comparison of the predicted free energy of each 20bp window, shown as either the the free energy of the best folding structure (MFE, green) and the weighted sum of the ensemble energy using the Boltzman distribution to calculate the probability of each structure (FE, blue). The correlation is also shown for windows of 100bps.

Figure S5 Correlation between the GC content and genome-wide mRNA levels for different window sizes and different locations. The analysis of Figure S3 is repeated here using genome-wide 3'UTR mapping and mRNA levels measured by RNA-seq. It shows that a similar result as in figure S3 is seen genome wide only with a lower yet much more significant correlation and is achieved in larger window sizes (around 100bp) which start at around -80bp and end around 20bp relative to the 3'UTR end.

Figure S6. Composition of mono- and di-nucleotides around transcription end sites. Shown are the mono- (upper panel) and di-nucleotide (lower panel) composition in sliding windows of 20bp centered around each point as a function of the distance from the transcription end site.

Figure S7 Correlation between expression measurements to different sequence composition features: Similar analysis to figure 4A (lower panel) and figure 5A (upper panel) calculating additional sequence composition features and correlating them also to expression (either RNA-seq or YFP production for the library). It shows that the highest correlation is achieved with AT content composition.

Figure S8. The occurrence of known transcription termination motifs is not significantly correlated with the expression level of the corresponding 3' UTR strain. For a sorting of the 3' UTR YFP expression, shown is their YFP production per cell per second (right); a coloring of the nucleotides of each 3' UTR (aligned by its transcription end site) according to the G/C content (gray – AT, white – GC) and markings of known transcription termination motifs (green and red lines mark termination efficiency and positioning elements, respectively[42]). The left panel shows a histogram of locations of transcription termination motifs as a function of their distance to the transcription end site, where the histograms correspond to the library strains that are within the top third expression values (top), middle third (middle), and lowest expression values (bottom).

Figure S9. Occurrence of known termination motifs in different expression groups using genome wide data. Using genome-wide expression data we divide the genes to three equally sized groups (shown in the histogram in the lower panel) and show the occurrence location histogram for efficiency (right panel) and positioning (left panel) elements showing the increase in these elements for highly expressed genes.

References:

1. Shalgi, R., Lapidot, M., Shamir, R. & Pilpel, Y. A catalog of stability-associated sequence elements in 3' UTRs of yeast mRNAs. *Genome biology* **6**, R86 (2005).

2. Xie, X. et al. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* **434**, 338-45 (2005).

3. Itzkovitz, S., Hodis, E. & Segal, E. Overlapping codes within protein-coding sequences. *Genome research* **20**, 1582-9 (2010).

4. Yassour, M. et al. Ab initio construction of a eukaryotic transcriptome by massively parallel mRNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 3264-9 (2009).

5. Spellman, P.T. et al. Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. *Molecular biology of the cell* **9**, 3273-97 (1998).

6. Nagalakshmi, U. et al. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science (New York, N.Y.)* **320**, 1344-9 (2008).

7. Lipson, D. et al. Quantification of the yeast transcriptome by single-molecule sequencing. *Nature biotechnology* **27**, 652-8 (2009).

8. Gasch, A.P. et al. Genomic expression programs in the response of yeast cells to environmental changes. *Molecular biology of the cell* **11**, 4241-57 (2000).

9. Holstege, F.C. et al. Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**, 717-28 (1998).

10. de Godoy, L.M.F. et al. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* **455**, 1251-4 (2008).

11. Newman, J.R.S. et al. Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* **441**, 840-6 (2006).

12. Ghaemmaghami, S. et al. Global analysis of protein expression in yeast. *Nature* **425**, 737-41 (2003).

13. Dahan, O., Gingold, H. & Pilpel, Y. Regulatory mechanisms and networks couple the different phases of gene expression. *Trends in genetics : TIG* **27**, 316-22 (2011).

14. Dori-Bachash, M., Shema, E. & Tirosh, I. Coupled evolution of transcription and mRNA degradation. *PLoS biology* **9**, e1001106 (2011).

15. Harel-Sharvit, L. et al. RNA polymerase II subunits link transcription and mRNA decay to translation. *Cell* **143**, 552-63 (2010).

16. Lasko, P. & Sonenberg, N. Coordinated transcriptional and translational control in metabolic homeostasis in flies. *Genes & development* **21**, 235-7 (2007).

17. Shalem, O. et al. Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation. *Molecular systems biology* **4**, 223 (2008).

18. Shalem, O., Groisman, B., Choder, M., Dahan, O. & Pilpel, Y. Transcriptome kinetics is governed by a genome-wide coupling of mRNA production and degradation: a role for RNA Pol II. *PLoS genetics* **7**, e1002273 (2011).

19. Zeisel, A. et al. Coupled pre-mRNA and mRNA dynamics unveil operational strategies underlying transcriptional responses to stimuli. *Molecular systems biology* **7**, 529 (2011).

20. Halbeisen, R.E. & Gerber, A.P. Stress-Dependent Coordination of Transcriptome and Translatome in Yeast. *PLoS biology* **7**, e105 (2009).

21. Cox, R.S., Surette, M.G. & Elowitz, M.B. Programming gene expression with combinatorial promoters. *Molecular systems biology* **3**, 145 (2007).

22. Gertz, J., Siggia, E.D. & Cohen, B.A. Analysis of combinatorial cis-regulation in synthetic and genomic promoters. *Nature* **457**, 215-8 (2009).

23. Kalir, S. et al. Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria. *Science (New York, N.Y.)* **292**, 2080-3 (2001).

24. Ligr, M., Siddharthan, R., Cross, F.R. & Siggia, E.D. Gene expression from random libraries of yeast promoters. *Genetics* **172**, 2113-22 (2006).

25. Murphy, K.F., Balázsi, G. & Collins, J.J. Combinatorial promoter design for engineering noisy gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 12726-31 (2007).

26. Zaslaver, A. et al. Just-in-time transcription program in metabolic pathways. *Nature genetics* **36**, 486-91 (2004).

27. Zeevi, D. et al. Compensation for differences in gene copy number among yeast ribosomal proteins is encoded within their promoters. *Genome research* (2011).doi:10.1101/gr.119669.110

28. Hook, B.A., Goldstrohm, A.C., Seay, D.J. & Wickens, M. Two yeast PUF proteins negatively regulate a single mRNA. *The Journal of biological chemistry* **282**, 15430-8 (2007).

29. Olivas, W. & Parker, R. The Puf3 protein is a transcript-specific regulator of mRNA degradation in yeast. *The EMBO journal* **19**, 6602-11 (2000).

30. Wickens, M., Bernstein, D.S., Kimble, J. & Parker, R. A PUF family portrait: 3'UTR regulation as a way of life. *Trends in genetics : TIG* **18**, 150-7 (2002).

31. Lohr, D., Venkov, P. & Zlatanova, J. Transcriptional regulation in the yeast GAL gene family: a complex genetic network. *The FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **9**, 777-87 (1995).

32. Giniger, E., Varnum, S.M. & Ptashne, M. Specific DNA binding of GAL4, a positive regulatory protein of yeast. *Cell* **40**, 767-74 (1985).

33. Giniger, E. & Ptashne, M. Cooperative DNA binding of the yeast transcriptional activator GAL4. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 382-6 (1988).

34. Jackson, J.S., Houshmandi, S.S., Lopez Leban, F. & Olivas, W.M. Recruitment of the Puf3 protein to its mRNA target for regulation of mRNA decay in yeast. *RNA (New York, N.Y.)* **10**, 1625-36 (2004).

35. Olivas, W. & Parker, R. The Puf3 protein is a transcript-specific regulator of mRNA degradation in yeast. *The EMBO journal* **19**, 6602-11 (2000).

36. Foat, B.C., Houshmandi, S.S., Olivas, W.M. & Bussemaker, H.J. Profiling condition-specific, genome-wide regulation of mRNA stability in yeast. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 17675-80 (2005).

37. Foat, B.C., Houshmandi, S.S., Olivas, W.M. & Bussemaker, H.J. Profiling condition-specific, genome-wide regulation of mRNA stability in yeast. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 17675-80 (2005).

38. van Helden, J., del Olmo, M. & Pérez-Ortín, J.E. Statistical analysis of yeast genomic downstream sequences reveals putative polyadenylation signals. *Nucleic acids research* **28**, 1000-10 (2000).

39. Graber, J.H., Cantor, C.R., Mohr, S.C. & Smith, T.F. In silico detection of control signals: mRNA 3'-end-processing sequences in diverse species. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 14055-60 (1999).

40. Ozsolak, F. et al. Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. *Cell* **143**, 1018-29 (2010).

41. Guo, Z. & Sherman, F. Signals sufficient for 3'-end formation of yeast mRNA. *Molecular and cellular biology* **16**, 2772-6 (1996).

42. Guo, Z. & Sherman, F. 3'-end-forming signals of yeast mRNA. *Trends in biochemical sciences* **21**, 477-81 (1996).

43. Zhao, J., Hyman, L. & Moore, C. Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiology and molecular biology reviews : MMBR* **63**, 405-45 (1999).

44. Cai, L., Friedman, N. & Xie, X.S. Stochastic protein expression in individual cells at the single molecule level. *Nature* **440**, 358-62 (2006).

45. Yu, J., Xiao, J., Ren, X., Lao, K. & Xie, X.S. Probing gene expression in live cells, one protein molecule at a time. *Science (New York, N.Y.)* **311**, 1600-3 (2006).

46. Li, G.-W. & Xie, X.S. Central dogma at the single-molecule level in living cells. *Nature* **475**, 308-15 (2011).

47. Paulsson, J. & Ehrenberg, M. Random signal fluctuations can reduce random fluctuations in regulated components of chemical regulatory networks. *Physical review letters* **84**, 5447-50 (2000).

48. Friedman, N., Cai, L. & Xie, X.S. Linking stochastic dynamics to population distribution: an analytical framework of gene expression. *Physical review letters* **97**, 168302 (2006).

49. Traven, A., Jelicic, B. & Sopta, M. Yeast Gal4: a transcriptional paradigm revisited. *EMBO reports* **7**, 496-9 (2006).

50. Kuehner, J.N., Pearson, E.L. & Moore, C. Unravelling the means to an end: RNA polymerase II transcription termination. *Nature reviews. Molecular cell biology* **12**, 283-94 (2011).

51. Komissarova, N., Becker, J., Solter, S., Kireeva, M. & Kashlev, M. Shortening of RNA:DNA hybrid in the elongation complex of RNA polymerase is a prerequisite for transcription termination. *Molecular cell* **10**, 1151-62 (2002).

52. Kireeva, M.L., Komissarova, N., Waugh, D.S. & Kashlev, M. The 8-nucleotide-long RNA:DNA hybrid is a primary stability determinant of the RNA polymerase II elongation complex. *The Journal of biological chemistry* **275**, 6530-6 (2000).

53. Mapendano, C.K., Lykke-Andersen, S., Kjems, J., Bertrand, E. & Jensen, T.H. Crosstalk between mRNA 3' end processing and transcription initiation. *Molecular cell* **40**, 410-22 (2010).

54. West, S. & Proudfoot, N.J. Transcriptional termination enhances protein expression in human cells. *Molecular cell* **33**, 354-64 (2009).

55. Amrani, N. et al. Aberrant termination triggers nonsense-mediated mRNA decay. *Biochemical Society transactions* **34**, 39-42 (2006).

Figure 1

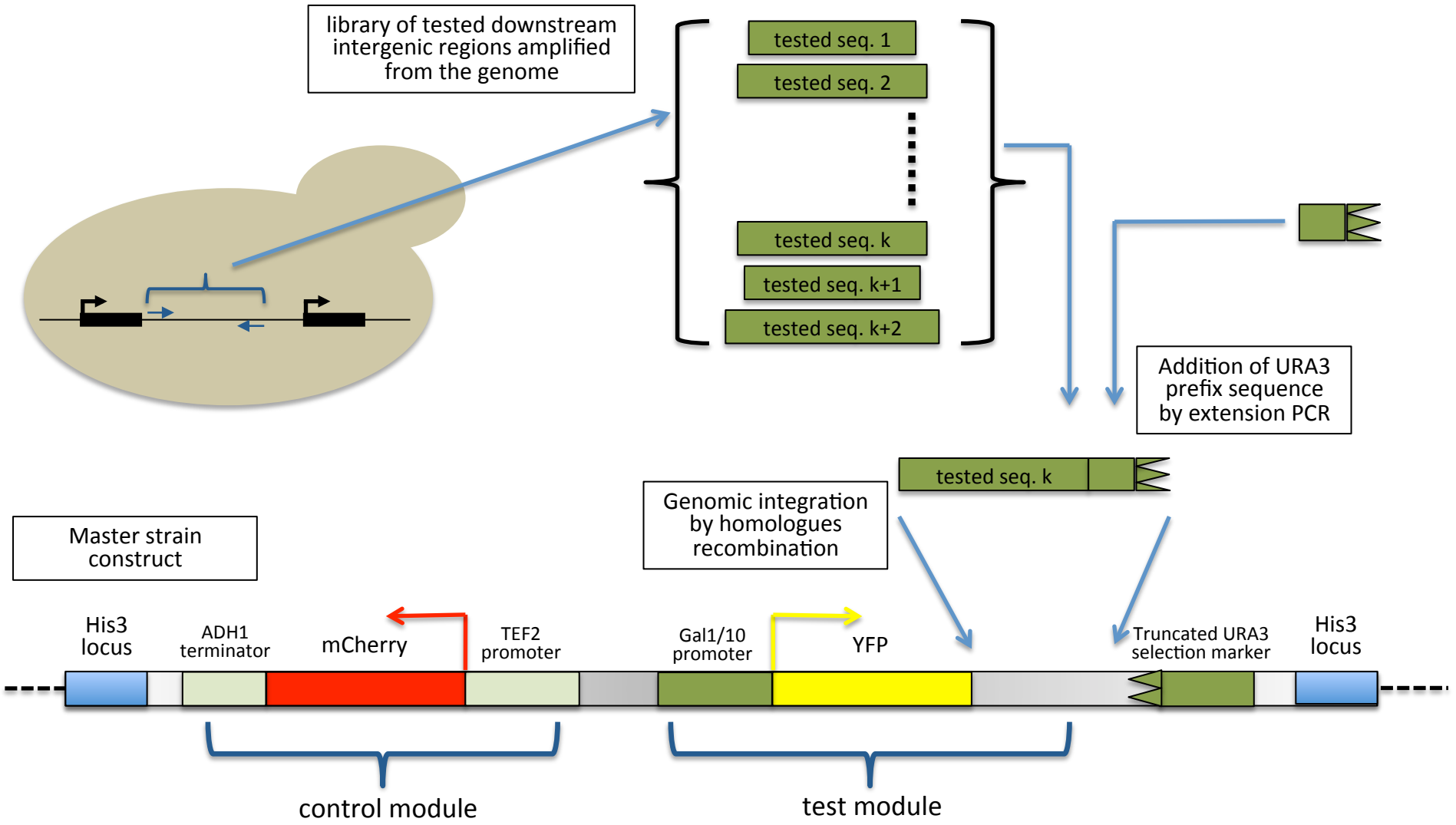

library of tested downstream intergenic regions amplified from the genome
tested seq. 1
tested seq. 2
tested seq. k
tested seq. k+1
tested seq. k+2
Addition of URA3 prefix sequence by extension PCR
tested seq. k
Genomic integration by homologues recombination
Master strain construct
His3 locus
ADH1 terminator
mCherry
TEF2 promoter
Gal1/10 promoter
YFP
Truncated URA3 selection marker
His3 locus
control module
test module

Figure 2A

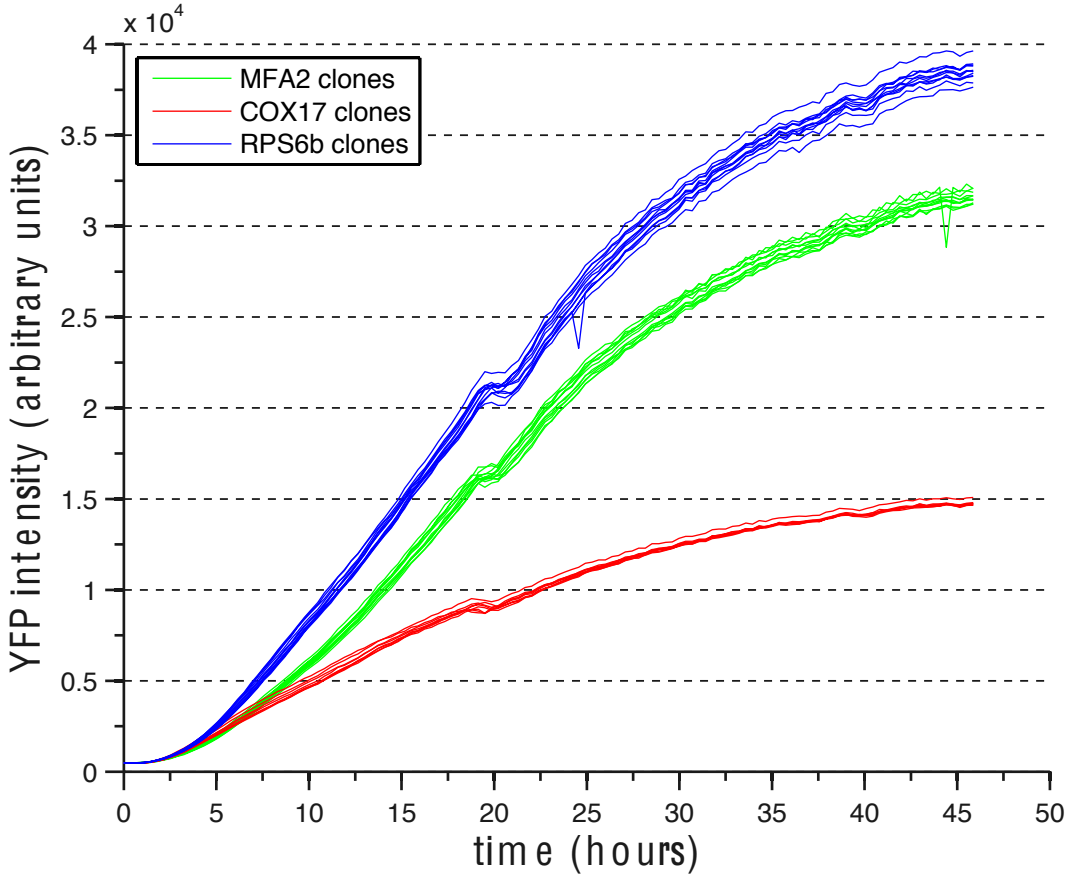

x 10^4
4
3.5
3
2.5
2
1.5
1
0.5
0
YFP intensity (arbitrary units)
MFA2 clones
COX17 clones
RPS6b clones
0
5
10
15
20
25
30
35
40
45
50
time (hours)

Figure 2B

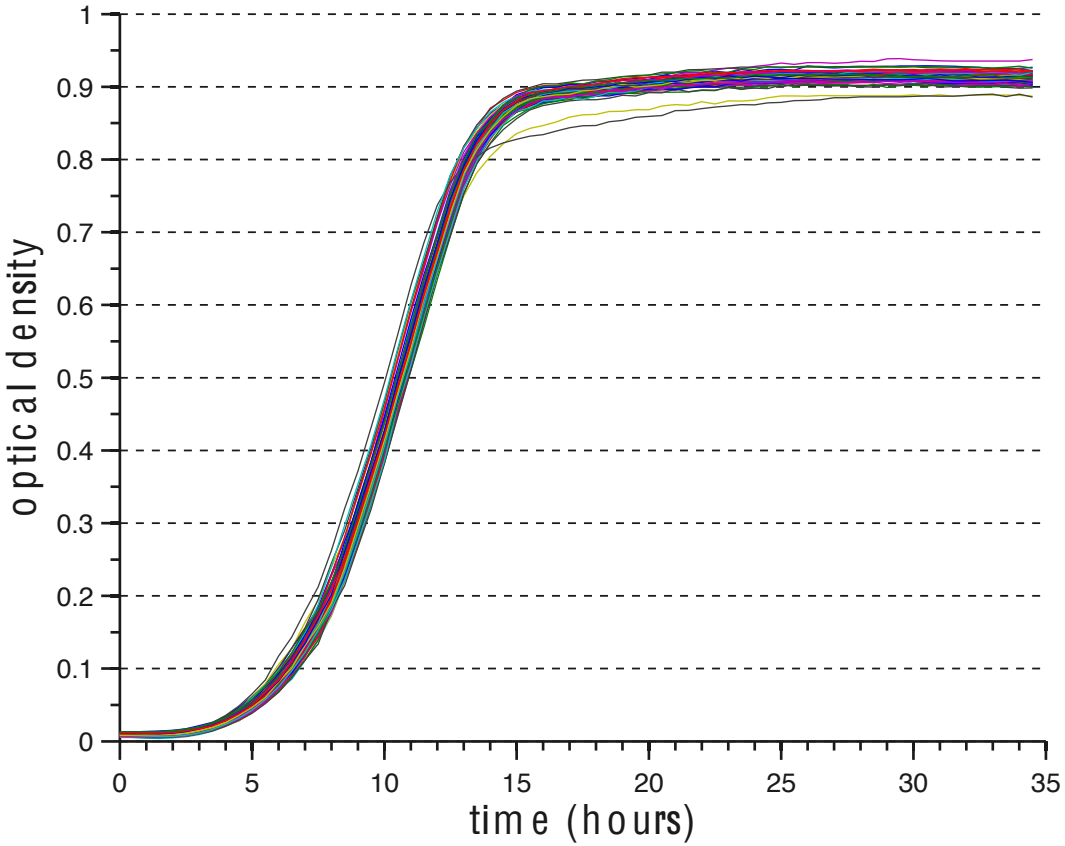

optical density
time (hours)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
5
10
15
20
25
30
35

Figure 2C

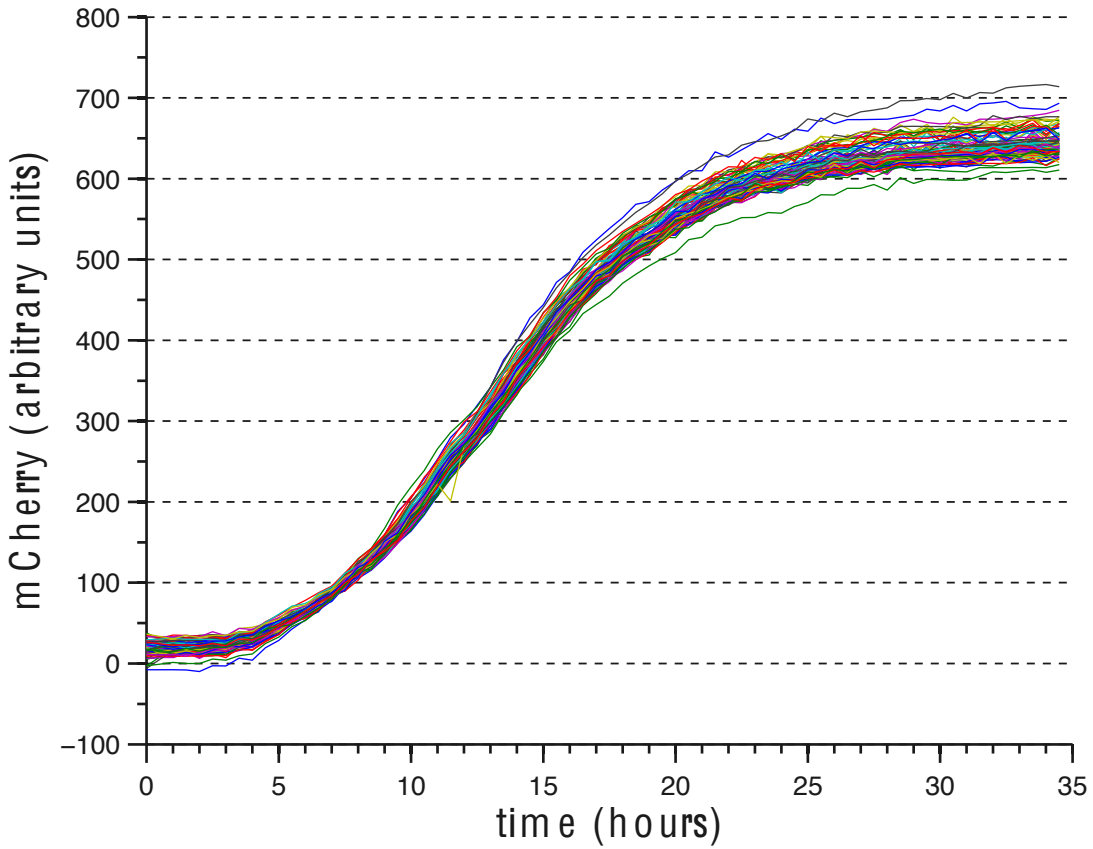

800
700
600
500
400
300
200
100
0
–100
0
5
10
15
20
25
30
35
mCherry (arbitrary units)
time (hours)

Figure 2D

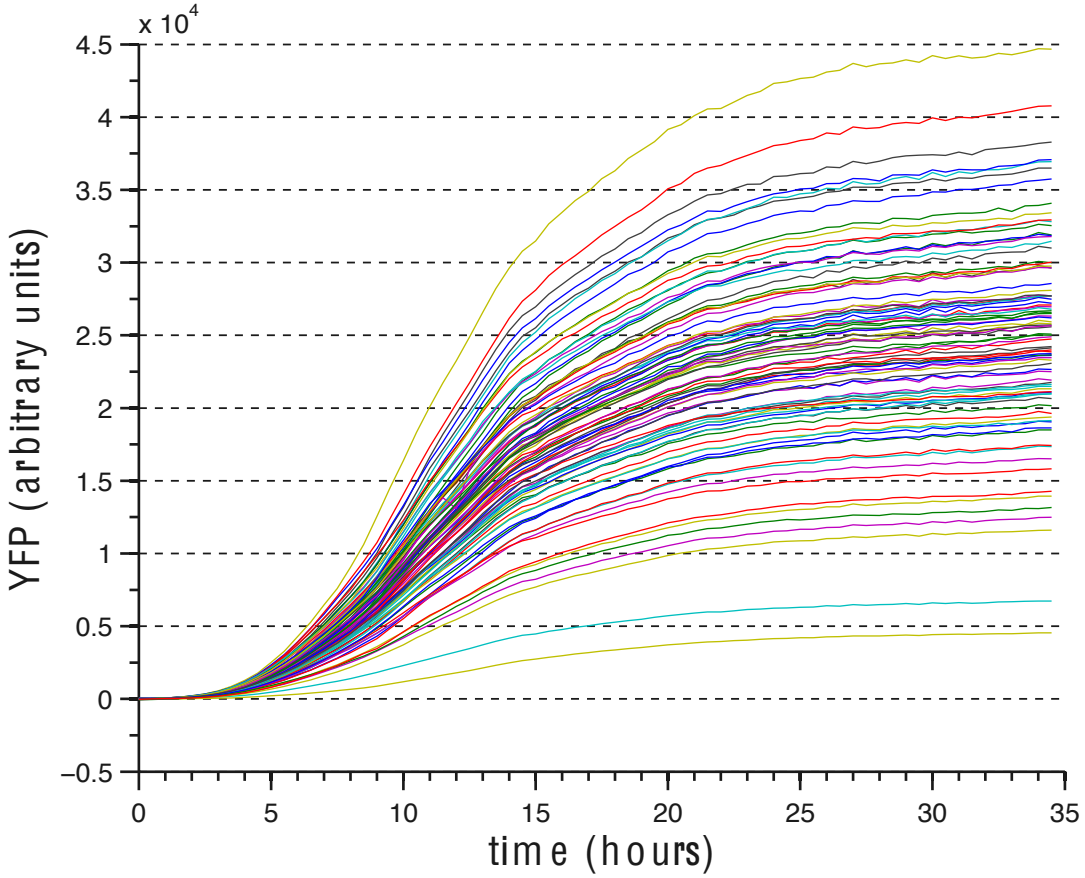

x 10^4
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
−0.5
YFP (arbitrary units)
0
5
10
15
20
25
30
35
time (hours)

Figure 3A

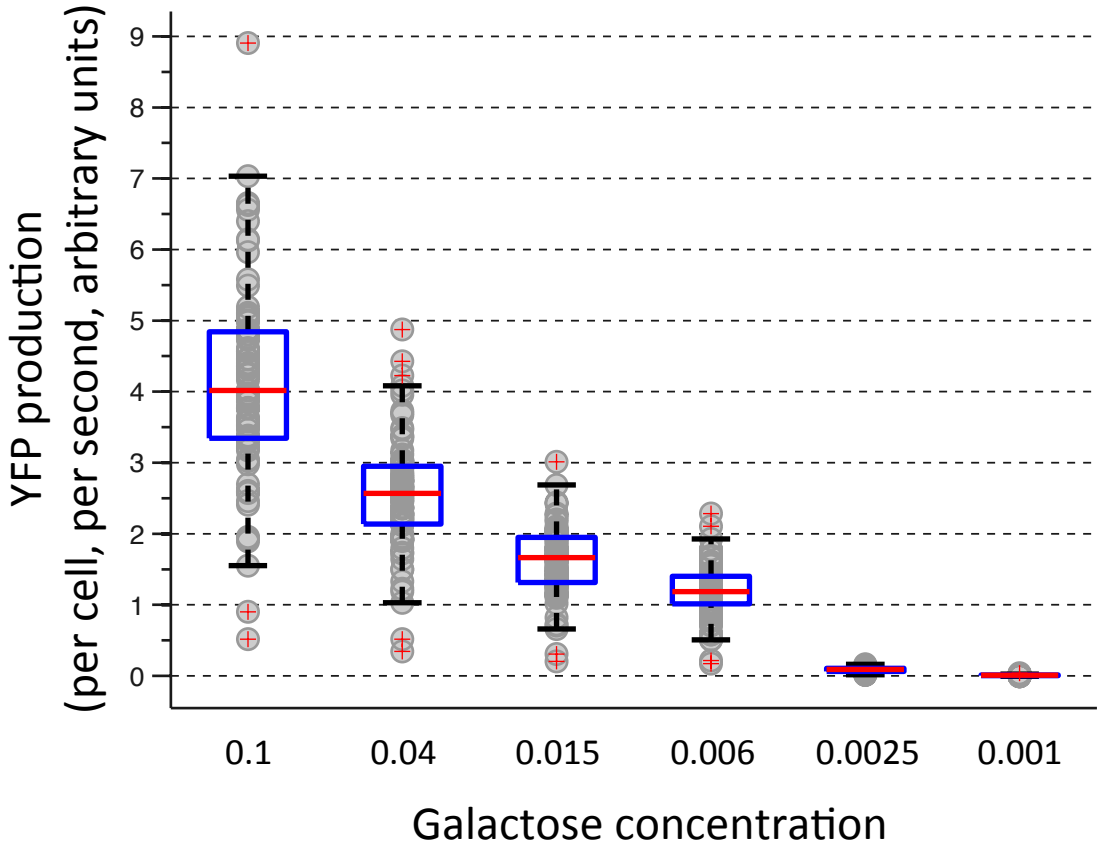

YFP production
(per cell, per second, arbitrary units)
9
8
7
6
5
4
3
2
1
0
0.1
0.04
0.015
0.006
0.0025
0.001
Galactose concentration

Figure 3B

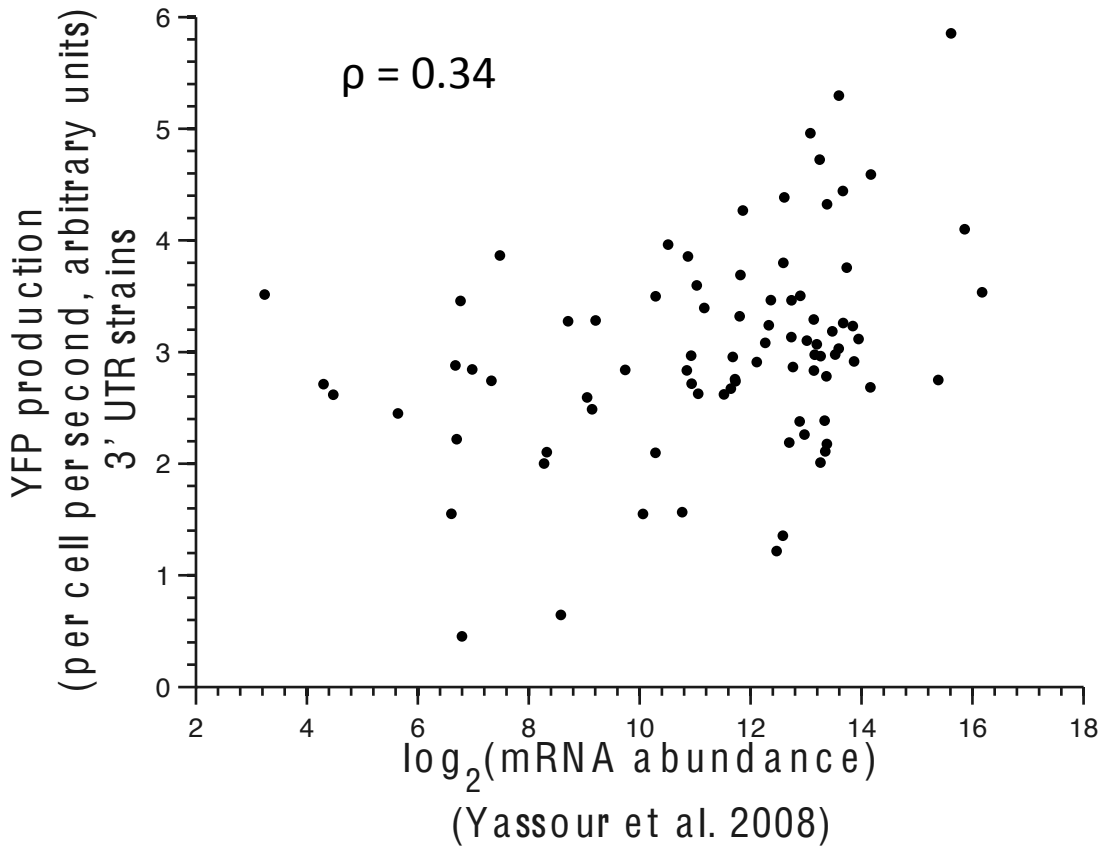

ρ = 0.34
YFP production
(per cell per second, arbitrary units)
3' UTR strains
log2(mRNA abundance)
(Yassour et al. 2008)
0
1
2
3
4
5
6
2
4
6
8
10
12
14
16
18

Figure 3C

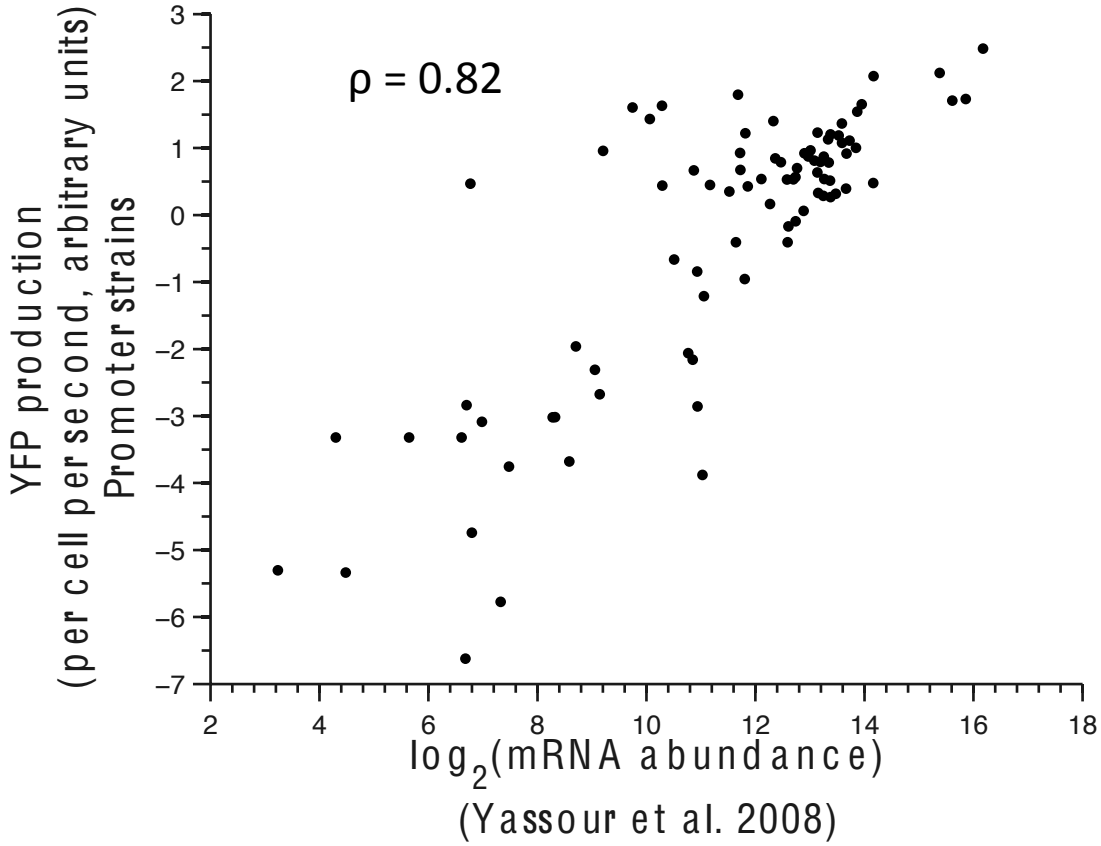

ρ = 0.82
YFP production
(per cell per second, arbitrary units)
Promoter strains
log2(mRNA abundance)
(Yassour et al. 2008)
3
2
1
0
−1
−2
−3
−4
−5
−6
−7
2
4
6
8
10
12
14
16
18

Figure 4A

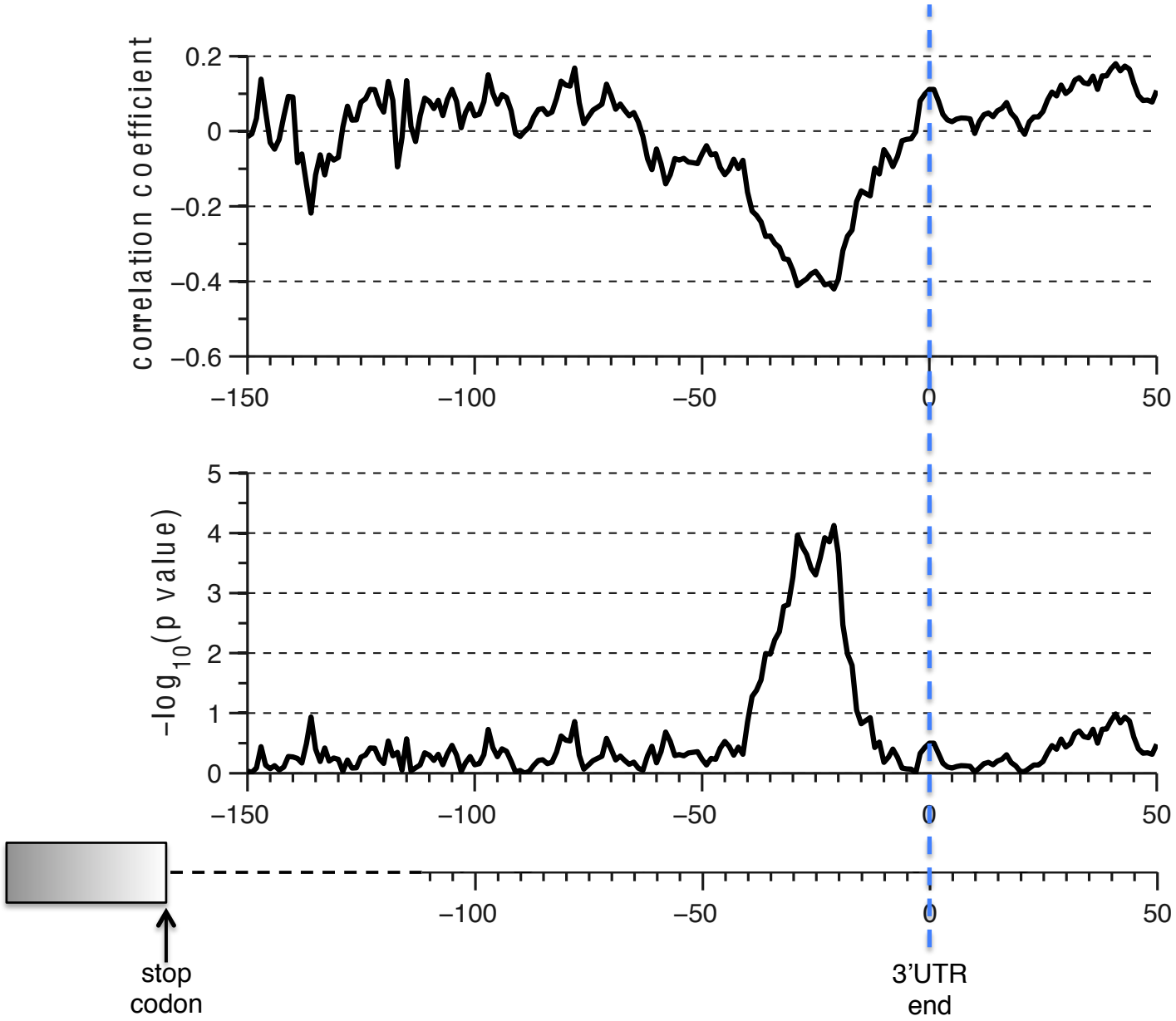

correlation coefficient
0.2
0
−0.2
−0.4
−0.6
−150
−100
−50
0
50
−log10(p value)
5
4
3
2
1
0
−150
−100
−50
0
50
−100
−50
0
50
stop
codon
3'UTR
end

Figure 4B

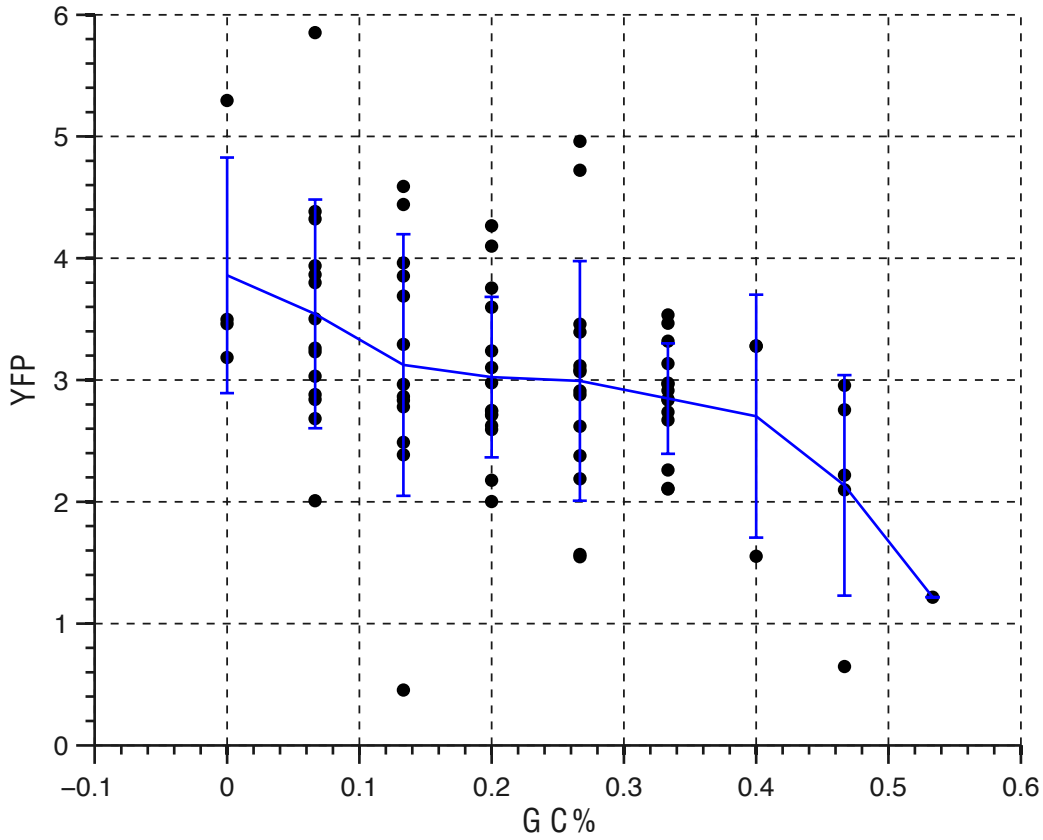


Figure 5A

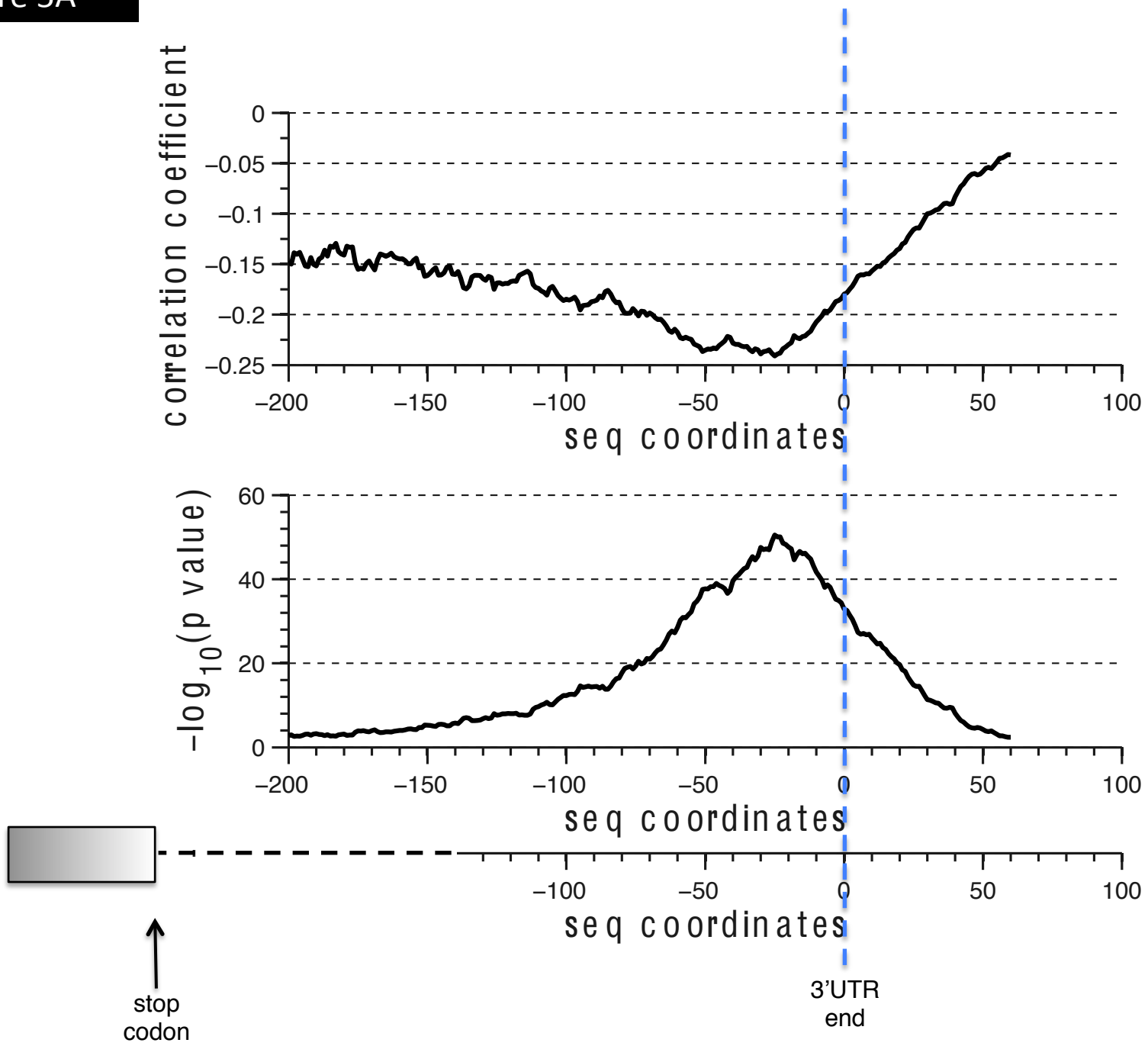

correlation coefficient
0
−0.05
−0.1
−0.15
−0.2
−0.25
−200
−150
−100
−50
0
50
100
seq coordinates
−log10(p value)
60
40
20
0
−200
−150
−100
−50
0
50
100
seq coordinates
−100
−50
0
50
100
seq coordinates
stop codon
3'UTR end

Figure 5B

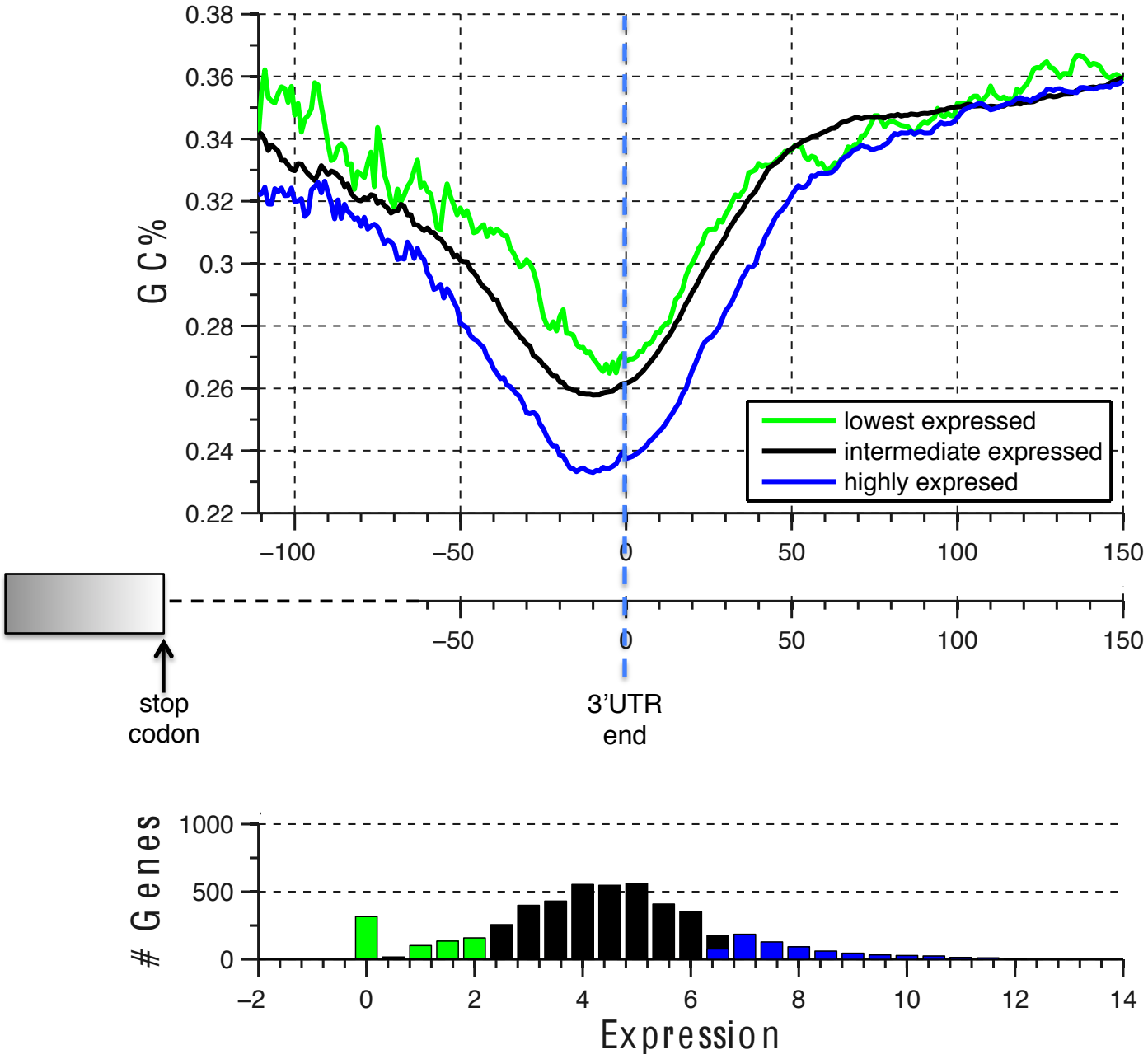
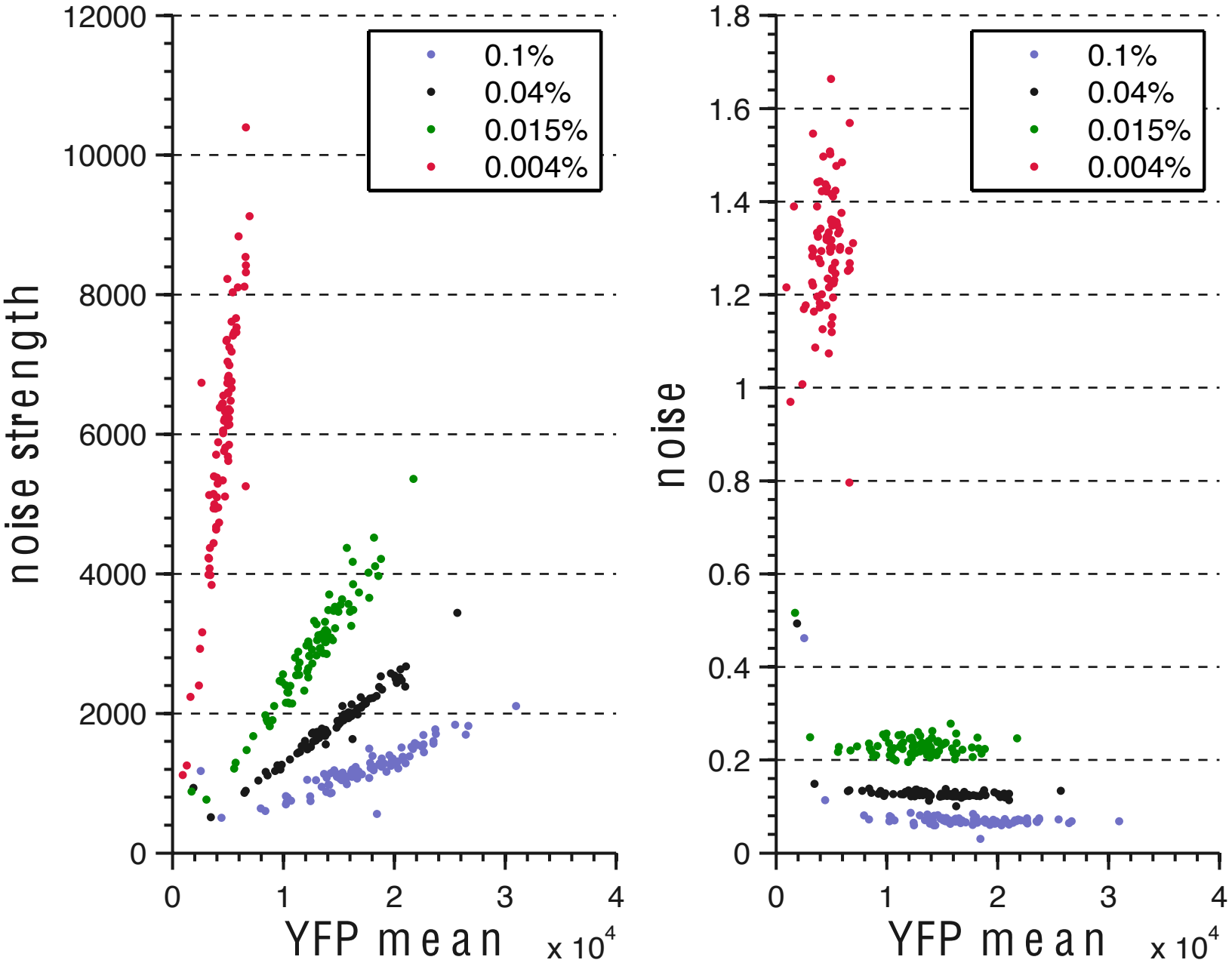

0.38
0.36
0.34
0.32
0.3
0.28
0.26
0.24
0.22
GC%
lowest expressed
intermediate expressed
highly expresed
−100
−50
0
50
100
150
stop
codon
3'UTR
end
Genes
1000
500
0
−2
0
2
4
6
8
10
12
14
Expression

Figure 6


noise strength
12000
10000
8000
6000
4000
2000
0
0.1%
0.04%
0.015%
0.004%
0
1
2
3
4
YFP mean
x 10^4
noise
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0.1%
0.04%
0.015%
0.004%
0
1
2
3
4
YFP mean
x 10^4

Discussion

We have initially set to study the dynamic properties of mRNA decay and found that on top of extensive condition specific modulation, changes in stability are coordinated with changes in production to control the temporal kinetics of mRNA abundance in response to changing conditions[22]. Following our publication similar observations were made in different conditions[37,38], different model systems[39,40] and using less perturbing ways to follow genome wide mRNA stability[41,42]. Our work makes an important contribution to the growing understanding thatfor acomprehensive view of gene expression regulation an integrated view is needed, in which the coordinated regulation of the different layers is taken into account, together with the propagation of the effect from one layer to the other. Indeed more studies examining regulatory coupling between the different stages in gene expression are quickly accumulating. Such studies show coupling effects ranging from transcription and pre-mRNA processing all the way to translation[43–49].

Despite the increasing understanding that regulatory layers are interconnected the molecular basis for those connections, especially on a genome-wide level, are lacking. In the work that followed the initial publication[22] we proposed a mechanism for the genome-wide coordination of mRNA production and degradation. We used a mutant strain in another Pol II subunit, Rpb6, bearing a single amino acid substitution which was shown to result in reduced recruitment of Rpb4/7 to the core polymerase[50]. Using measurements of mRNA abundance and decay in stress and permissive conditions we show that this molecular mechanism, involving co-transcriptional imprinting by Rpb4 and 7, can account for our previously observed genome-wide coordination and its effect on the temporal kinetics of mRNA abundance in response to stress. We suggest a simple mechanistic model, which is sufficient to explain these results. This model is especially appealing since it shows how a seemingly complicated task, of controlling the temporal kinetics of a genome wide transcriptome response, can be achieved by simple manipulations of the general co-transcriptional imprinting

probability, leading to the redistribution of the decay machinery across mRNAs in the cytoplasm.

Our above results are based on measurements of mRNA decay using transcriptioninhibition, achieved by either inhibiting RNA Pol II, using a temperature sensitivemutant, or by chemical inhibition of all RNA polymerization using the drug 1,10-Phenantroline. Complete shutoff of transcription has been widely used in recent years, both to study gene specific cases of mRNA stability[51] and to follow genome wide decay kinetics[3,8]. The different drugs used to chemically inhibit transcription were compared and shown to have similar genome-wide effect compared to inhibition using RNA Pol II temperature sensitive mutant[6]. Specifically the effect of 1,10-Phenantroline on mRNA decay was most similar to that obtained with the temperature-sensitive mutant and hence we chose to use this drugfor the characterization of the mutant (and for an additional characterization of the wild-type in order to facilitate comparison of the strains). In addition transcription-arrest based methods were also compared to less pervasive methods such as the traditional pulse-chase experiments revelingsimilar results on a few tested genes[3,51].
Despite the wide use of the transcription arrest research paradigm it has some crucial downfalls that cannot be ignored in futureworksaiming to study the dynamics of gene expression regulation using measurements of mRNA stability. Transcription inhibition, either by temperature sensitive mutant or chemical inhibition, involves a major stress to the cells, meaning that using these methods, stability cannot be measured in real permissive conditions. While the cells cannot have a full transcriptional response, they probably react to the treatment post-transcriptionally affecting decay measurements by some unknown way. It is also likely that transcription inhibition by itself poses a serious stress to the cells to which it might react by some feedback regulatory mechanism.
An additional confounding factor of our experimental systemis residual transcription that might be still left even after the attempt to arrest transcription. In case of incomplete shut down of transcription, itmight have a variable effect across genes and conditions affecting our conclusions. A detailed explanation on the ways these confoundingfactors are handled is given in the

methods and discussion sections in the relevant papers. Briefly, to deal with the first concern we only use the ratio of half-livesbetween pairs of conditions, usually with and without a treatment. In this case every effect we see is additional to the effect of transcription inhibition, shared by the two conditions in which we measure stability. Regarding residual transcription, we would expect genes suffering from such effect to have variable degradation rate throughout the measured time course. In particular such genes may show a characteristic exponential decay only from a late time point after transcriptional arrest, and may not show a good fit to an exponential decay model from the initial time points. We thus only consider for further analyses genes that display a good linear fit of the log2 mRNA profile as a function of time following transcription inhibition, using a stringent R^2 cutoff.We verified that our qualitative results are insensitive to the choice of R^2 cutoff.
In light of these concerns, to complement direct measurements, recent methods have been proposed to allow for the dissectionof changes in mRNA levels into their production and degradation components in a less perturbing way. This is done either by incorporation of metabolically labeled nucleotides to follow newly synthesized mRNAs[42,52] or by measuring pre-mRNA and mRNA together followed by kinetic modelingof the transcriptome response[53].These methods will probably be the basis for future works aiming to study additional coupling mechanisms.

In a second line of research I study 3' UTR sequences by creating an experimental pipeline to measure the independent contribution of these regions on protein expression. Cases of 3' UTR mediated regulation are known, yet the general constitutive contribution of these sequences to differences in protein expression is still unclear. We were surprised to find out that each 3' UTR sequence had a unique and reproducible effect on gene expression. This effect spansa continues range of expression values of up to 15 fold difference between the highest to the lowest expressing strain. Because we sample such small set of 3' UTR sequences, without any prior information regarding their expected effect, it is likely that the extent of genome-wide effect of 3' UTRs on gene expression is larger.

Using simple sequence analysis we find a signature, based on nucleotide composition that correlates significantly with expression measurements both in the library and the genome. Despite its vague description, it highlights the importance of the sequences upstream to the polyadenylation site for protein expressionsuggesting that part of the observed expression differences in our library results from differences in transcription termination efficiency. Transcription termination is a highly complex process, involving large protein complexes, physically linked to both transcription elongation and export[54].Despite the importance of this process to the maturation of each mRNA, it is not usually considered to exert any effect on protein levels. Our work joins a few recent studies[55,56], which put forth transcription termination, together with its controlling sequence elements, as important determinants of protein levels. AT rich sequences at the 3' UTR end may enable more efficient transcription termination by either introducing stronger termination motifs, known to be A/T rich[57–59], or by an efficient release of RNA polymerase after termination due to a looser elongating 8bp RNA/DNA duplex[60,61]. Efficient transcription termination has been proposed to affect protein levels by two means: The first suggests that with less efficient termination more elongating RNA polymerases read-though the termination sites producing mRNA molecules with aberrantly long 3' UTRs[55]. These aberrant transcripts are then degraded in the nucleus by mRNA surveillance mechanisms resulting in less mature mRNA molecules in the cytoplasm per transcription initiation event. A second way by which efficient termination might enhance protein levels is by recycling of the RNA polymerase complex proteins. Recent works show that when known termination motifs are mutated protein levels significantly decrease. This decrease was shown to be a result of a decrease in the number of initiation events at the promoter suggesting a direct link between transcription termination and initiation[56]. A possible mechanistic explanation could be that changes in the genome structure during active transcription bring the 3' and 5' ends of the transcribing open reading frame in close proximity[54].

Further characterization of the sequence features that we found by mutational analysis, and examination of additional 3' UTR sequences will be needed in order to identify additional mechanism by which 3' UTR sequences exert their effect on

protein expression. Such further analysis will pave the way to a more complete understanding of gene expression, which also incorporates the effect of the 3' UTR.

This study demonstrates a novel research paradigm in which isolated genomic regions that are constructed in the context of synthetic libraries can allow to dissect the regulatory effects of specific regions on the genome. Such detailed examination is not possible using genomic data sets of mRNA and protein expression since natural genomic data represent the cumulative effect of all regulatory regions.

On top of what we have shown, this synthetic approach enables the establishment of a causal link between sequence features and their outputs.Observing correlations in the genome, e.g. between sequence features in the 3' UTR and expression levels could always be explained by indirect non-causal effects. For example, one could argue that the genes with certain UTR features may also have strong promoters. Also, efficient transcription termination in highly expressed should not necessarily affect expression,it can be that non efficient termination exerts a cost on the cell which results in higher optimizationof this process for highly expressed genes. Observing such connections in thecontext of our synthetic library, in which the effect of 3' UTR sequences is measured in isolation, removes many such potential confounders.

The two works described in this thesis share a simple notion. In both projects we study the complexity of gene expression by dissecting it into its basic components. In the first line of work this idea is obtained on the level of the different regulatory processes, specifically mRNA production and decay, by inhibiting one process we can infer the other. In the second work separation between the effects of various regulatory regions is achieved by synthesizing a genetic construct that consists of only one variable part of the regulatory input (the 3' UTR) in addition to a constant part (the promoter).

Interestingly, both works reveal a related principle,that different regulatory processes are coupled, sharing regulatory mechanisms and having effects that propagate from one layer to the other. This is immediately apparent from the results of the first study, suggestingthat mRNA decay is co-transcriptionally

regulated. The second studymight also reveal a type of coupling: If indeed the differences in florescence levels in our library result from differences in transcription termination efficiency, it draws a direct line between this process to the downstream events leading to protein translation. More intriguingly if this effect is due to re-initiation through recycling of RNA polymeraseit shows how termination at the 3' end is directly coupled to initiation at the promoter.
All together, these studies expand our understanding of the basic processes of gene expression, showing interdependencies between the different layers of regulation and putting forth an additional layer which is usually not considered to have a significant effect on protein levels.

Bibliography

1. Garneau, N.L., Wilusz, J. & Wilusz, C.J. The highways and byways of mRNA decay. *Nature reviews. Molecular cell biology***8**, 113-26 (2007).

2. Drinnenberg, I.A. *et al*. RNAi in budding yeast. *Science (New York, N.Y.)***326**, 544-50 (2009).

3. Wang, Y. *et al*. Precision and functional specificity in mRNA decay. *Proceedings of the National Academy of Sciences of the United States of America***99**, 5860-5 (2002).

4. Sheth, U. & Parker, R. Decapping and decay of messenger RNA occur in cytoplasmic processing bodies. *Science (New York, N.Y.)***300**, 805-8 (2003).

5. Wilusz, C.J., Wormington, M. & Peltz, S.W. The cap-to-tail guide to mRNA turnover. *Nature reviews. Molecular cell biology***2**, 237-46 (2001).

6. Grigull, J., Mnaimneh, S., Pootoolal, J., Robinson, M.D. & Hughes, T.R. Genome-wide analysis of mRNA stability using transcription inhibitors and microarrays reveals posttranscriptional control of ribosome biogenesis factors. *Molecular and cellular biology***24**, 5534-47 (2004).

7. Yang, E. *et al*. Decay rates of human mRNAs: correlation with functional characteristics and sequence attributes. *Genome research***13**, 1863-72 (2003).

8. Bernstein, J.A., Khodursky, A.B., Lin, P.-H., Lin-Chao, S. & Cohen, S.N. Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. *Proceedings of the National Academy of Sciences of the United States of America***99**, 9697-702 (2002).

9. Dahan, O., Gingold, H. & Pilpel, Y. Regulatory mechanisms and networks couple the different phases of gene expression. *Trends in genetics : TIG***27**, 316-22 (2011).

10. Tucker, M. *et al*. The transcription factor associated Ccr4 and Caf1 proteins are components of the major cytoplasmic mRNA deadenylase in Saccharomyces cerevisiae. *Cell***104**, 377-86 (2001).

11. Liu, H.Y. *et al*. Characterization of CAF4 and CAF16 reveals a functional connection between the CCR4-NOT complex and a subset of SRB proteins of the RNA polymerase II holoenzyme. *The Journal of biological chemistry***276**, 7541-8 (2001).

12. Benson, J.D., Benson, M., Howley, P.M. & Struhl, K. Association of distinct yeast Not2 functional domains with components of Gcn5 histone acetylase and Ccr4 transcriptional regulatory complexes. *The EMBO journal***17**, 6714-22 (1998).

13. Denis, C.L. & Chen, J. The CCR4-NOT complex plays diverse roles in mRNA metabolism. *Progress in nucleic acid research and molecular biology***73**, 221-50 (2003).

14. Collart, M.A. Global control of gene expression in yeast by the Ccr4-Not complex. *Gene***313**, 1-16 (2003).

15. Choder, M. mRNA imprinting: Additional level in the regulation of gene expression. *Cellular logistics***1**, 37-40 (2011).

16. Goler-Baron, V. *et al*. Transcription in the nucleus and mRNA decay in the cytoplasm are coupled processes. *Genes & development***22**, 2022-7 (2008).

17. Lotan, R. *et al*. The RNA polymerase II subunit Rpb4p mediates decay of a specific class of mRNAs. *Genes & development***19**, 3004-16 (2005).

18. Lotan, R., Goler-Baron, V., Duek, L., Haimovich, G. & Choder, M. The Rpb7p subunit of yeast RNA polymerase II plays roles in the two major cytoplasmic mRNA decay mechanisms. *The Journal of cell biology***178**, 1133-43 (2007).

19. Runner, V.M., Podolny, V. & Buratowski, S. The Rpb4 subunit of RNA polymerase II contributes to cotranscriptional recruitment of 3' processing factors. *Molecular and cellular biology***28**, 1883-91 (2008).

20. Harel-Sharvit, L. *et al*. RNA polymerase II subunits link transcription and mRNA decay to translation. *Cell***143**, 552-63 (2010).

21. Choder, M. mRNA imprinting: Additional level in the regulation of gene expression. *Cellular logistics***1**, 37-40 (2011).

22. Shalem, O. *et al*. Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation. *Molecular systems biology***4**, 223 (2008).

23. Amorim, M.J., Cotobal, C., Duncan, C. & Mata, J. Global coordination of transcriptional control and mRNA decay during cellular differentiation. *Molecular systems biology***6**, 380 (2010).

24. Castells-Roca, L. *et al*. Heat shock response in yeast involves changes in both transcription rates and mRNA stabilities. *PloS one***6**, e17272 (2011).

25. Miller, C. *et al*. Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. *Molecular systems biology***7**, 458 (2011).

26. Molin, C., Jauhiainen, A., Warringer, J., Nerman, O. & Sunnerhagen, P. mRNA stability changes precede changes in steady-state mRNA amounts during hyperosmotic stress. *RNA (New York, N.Y.)***15**, 600-14 (2009).

27. Rabani, M. *et al*. Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nature biotechnology* (2011).doi:10.1038/nbt.1861

28. Lotan, R. *et al*. The RNA polymerase II subunit Rpb4p mediates decay of a specific class of mRNAs. *Genes & development***19**, 3004-16 (2005).

29. Lotan, R., Goler-Baron, V., Duek, L., Haimovich, G. & Choder, M. The Rpb7p subunit of yeast RNA polymerase II plays roles in the two major cytoplasmic mRNA decay mechanisms. *The Journal of cell biology***178**, 1133-43 (2007).

30. Yassour, M. *et al*. Ab initio construction of a eukaryotic transcriptome by massively parallel mRNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America***106**, 3264-9 (2009).

31. Spellman, P.T. *et al*. Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. *Molecular biology of the cell***9**, 3273-97 (1998).

32. Gasch, A.P. *et al*. Genomic expression programs in the response of yeast cells to environmental changes. *Molecular biology of the cell***11**, 4241-57 (2000).

33. Nonet, M., Scafe, C., Sexton, J. & Young, R. Eucaryotic RNA polymerase conditional mutant that rapidly ceases mRNA synthesis. *Molecular and cellular biology***7**, 1602-11 (1987).

34. Shalem, O., Groisman, B., Choder, M., Dahan, O. & Pilpel, Y. Transcriptome kinetics is governed by a genome-wide coupling of mRNA production and degradation: a role for RNA Pol II. *PLoS genetics***7**, e1002273 (2011).

35. Paulsson, J. & Ehrenberg, M. Random signal fluctuations can reduce random fluctuations in regulated components of chemical regulatory networks. *Physical review letters***84**, 5447-50 (2000).

36. Friedman, N., Cai, L. & Xie, X.S. Linking stochastic dynamics to population distribution: an analytical framework of gene expression. *Physical review letters***97**, 168302 (2006).

37. Castells-Roca, L. *et al*. Heat shock response in yeast involves changes in both transcription rates and mRNA stabilities. *PloS one***6**, e17272 (2011).

38. Molin, C., Jauhiainen, A., Warringer, J., Nerman, O. & Sunnerhagen, P. mRNA stability changes precede changes in steady-state mRNA amounts during hyperosmotic stress. *RNA (New York, N.Y.)***15**, 600-14 (2009).

39. Elkon, R., Zlotorynski, E., Zeller, K.I. & Agami, R. Major role for mRNA stability in shaping the kinetics of gene induction. *BMC genomics***11**, 259 (2010).

40. Amorim, M.J., Cotobal, C., Duncan, C. & Mata, J. Global coordination of transcriptional control and mRNA decay during cellular differentiation. *Molecular systems biology***6**, 380 (2010).

41. Rabani, M. *et al*. Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nature biotechnology***29**, 436-42 (2011).

42. Miller, C. *et al*. Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. *Molecular systems biology***7**, 458 (2011).

43. Zeisel, A. *et al*. Coupled pre-mRNA and mRNA dynamics unveil operational strategies underlying transcriptional responses to stimuli. *Molecular systems biology***7**, 529 (2011).

44. Shalgi, R., Lieber, D., Oren, M. & Pilpel, Y. Global and local architecture of the mammalian microRNA-transcription factor regulatory network. *PLoS computational biology***3**, e131 (2007).

45. Proshkin, S., Rahmouni, A.R., Mironov, A. & Nudler, E. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science (New York, N.Y.)***328**, 504-8 (2010).

46. Moore, M.J. & Proudfoot, N.J. Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell***136**, 688-700 (2009).

47. Lasko, P. & Sonenberg, N. Coordinated transcriptional and translational control in metabolic homeostasis in flies. *Genes & development***21**, 235-7 (2007).

48. Hu, W., Sweet, T.J., Chamnongpol, S., Baker, K.E. & Coller, J. Co-translational mRNA decay in Saccharomyces cerevisiae. *Nature***461**, 225-9 (2009).

49. Halbeisen, R.E. & Gerber, A.P. Stress-Dependent Coordination of Transcriptome and Translatome in Yeast. *PLoS biology***7**, e105 (2009).

50. Tan, Q., Prysak, M.H. & Woychik, N.A. Loss of the Rpb4/Rpb7 subcomplex in a mutant form of the Rpb6 subunit shared by RNA polymerases I, II, and III. *Molecular and cellular biology***23**, 3329-38 (2003).

51. Herrick, D., Parker, R. & Jacobson, A. Identification and comparison of stable and unstable mRNAs in Saccharomyces cerevisiae. *Molecular and cellular biology***10**, 2269-84 (1990).

52. Rabani, M. *et al*. Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nature biotechnology***29**, 436-42 (2011).

53. Zeisel, A. *et al*. Coupled pre-mRNA and mRNA dynamics unveil operational strategies underlying transcriptional responses to stimuli. *Molecular systems biology***7**, 529 (2011).

54. Kuehner, J.N., Pearson, E.L. & Moore, C. Unravelling the means to an end: RNA polymerase II transcription termination. *Nature reviews. Molecular cell biology***12**, 283-94 (2011).

55. West, S. & Proudfoot, N.J. Transcriptional termination enhances protein expression in human cells. *Molecular cell***33**, 354-64 (2009).

56. Mapendano, C.K., Lykke-Andersen, S., Kjems, J., Bertrand, E. & Jensen, T.H. Crosstalk between mRNA 3' end processing and transcription initiation. *Molecular cell***40**, 410-22 (2010).

57. Guo, Z. & Sherman, F. 3'-end-forming signals of yeast mRNA. *Trends in biochemical sciences***21**, 477-81 (1996).

58. Guo, Z. & Sherman, F. Signals sufficient for 3'-end formation of yeast mRNA. *Molecular and cellular biology***16**, 2772-6 (1996).

59. Graber, J.H., Cantor, C.R., Mohr, S.C. & Smith, T.F. In silico detection of control signals: mRNA 3'-end-processing sequences in diverse species. *Proceedings of the National Academy of Sciences of the United States of America***96**, 14055-60 (1999).

60. Komissarova, N., Becker, J., Solter, S., Kireeva, M. & Kashlev, M. Shortening of RNA:DNA hybrid in the elongation complex of RNA polymerase is a prerequisite for transcription termination. *Molecular cell***10**, 1151-62 (2002).

61. Kireeva, M.L., Komissarova, N., Waugh, D.S. & Kashlev, M. The 8-nucleotide-long RNA:DNA hybrid is a primary stability determinant of the RNA polymerase II elongation complex. *The Journal of biological chemistry***275**, 6530-6 (2000).

Declaration

This thesis summarizes my independent work under the supervision of Yitzhak Pilpel joint with Eran Segal. I have preformed all the experimental work and downstream analysis by myself excluding part of the experimental work related to the *PLoS Genetics* paper, which was preformed by Bella Groisman, a technician in the lab.

Acknowledgements

First and foremost I would like to thank Yitzhak Pilpel for his endless scientific enthusiasm, insightful advice, criticism when needed and for providing me with a great environment to grow and develop on my own. Eran Segal was my second advisor since day one. I have gained tremendously from his insight, encouragement and critical and direct point of view. Together, with their differences and similarities, they have shaped my scientific thinking and I am extremely grateful to have had both of them as my advisors.
A special thanks goes to Orna Dahan for making a computer science student into a proud biologist and for helping me with my first steps with a pipette.
Enumerating all my friends and colleges in the Weizmann Institute will probably take a whole page, and being myself I will probably forget a few very important people. So I will simply thank all the members of the Pilpel and Segal labs for the great scientific discussions, the friendship, and for creating an amazing learning and fun environment I looked forward to get to every morning.
Last and not least, I thank my family: my parents, my sister and my grandparents, for their love and support, without which I might have not been able take this path.

Appendix A – supplementary material for the paper "Transient transcriptional responses to stress are generated by opposing affect of mRNA production and degradation"

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

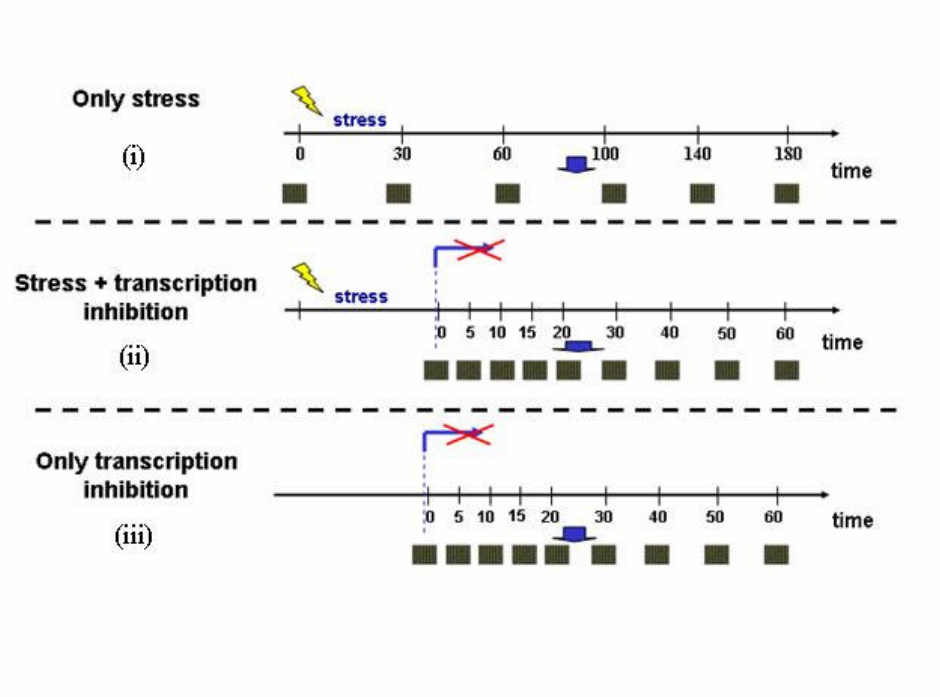


Figure S1: **A schematic illustration of the experimental procedure:** Three types of experiments were conducted: (i) A conventional microarray experiment where mRNA abundance was measured following the perturbation, (ii) a stress followed by transcription inhibition to measure condition specific decay 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.

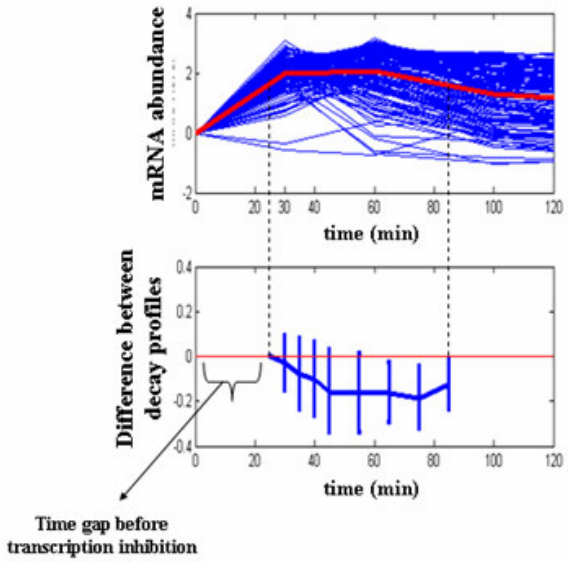


Figure S2: **Difference between decay profiles as a function of time:** For each gene, at each time point, the reference decay profile is subtracted from the oxidative stress decay profile, then at each time point the mean and standard deviation of the difference is calculated and plotted (lower panel) 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).

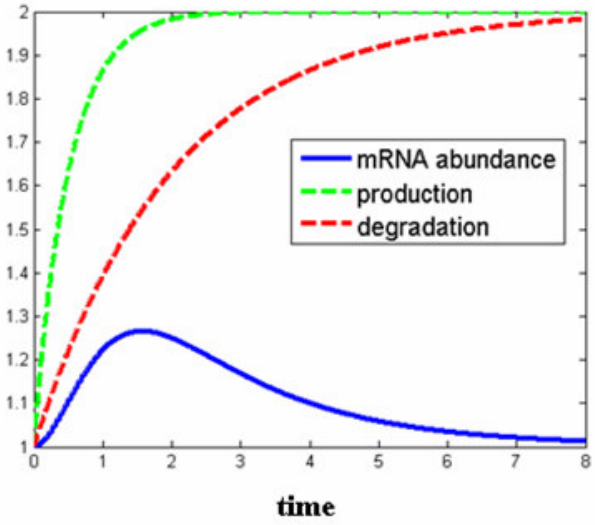


Figure S3: **A proposed model that accounts for the transient changes in mRNA abundance:** Increase in both production and degradation creates a transient increase in mRNA abundance if we assume that the increase in degradation is slower than the increase in production. This time delay will produce a 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.

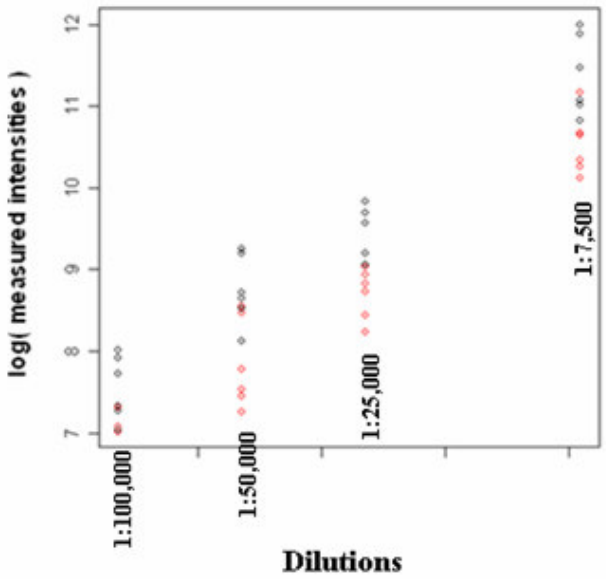


Figure S4: **linear relationship between the spikes known dilutions to measured intensities:** Plot of the known dilutions vs. the measured intensities of all probe sets representing the "spiked in" RNA genes. Each of the four genes is represented by several probe sets. The two colors represent two different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

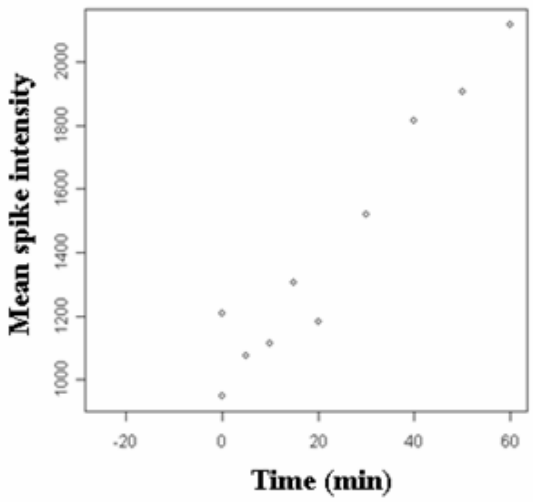


Figure S5: **Increase in the relative concentration of "spiked in" RNA:** The mean spike intensity for each sample is plotted against the time at which it was taken after transcription inhibition. A monotonic increase is apparent as expected.

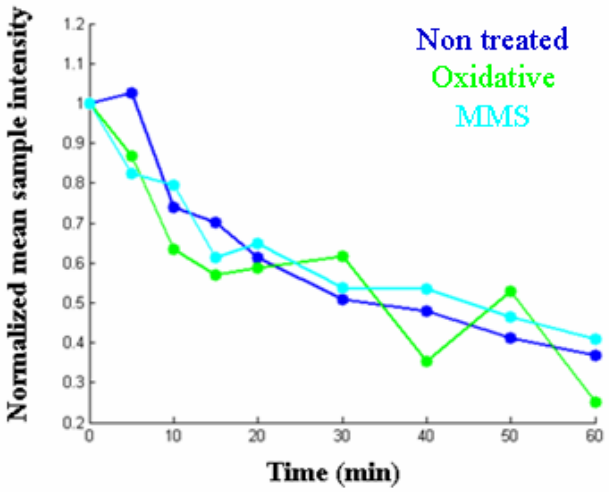


Figure S6: **Mean mRNA amount decreases with time:** The mean mRNA amount for each sample is plotted as a function of time following transcription inhibition for all three conditions. The values for each time course are normalized to the first time point in order to compare the decay kinetics of each 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.

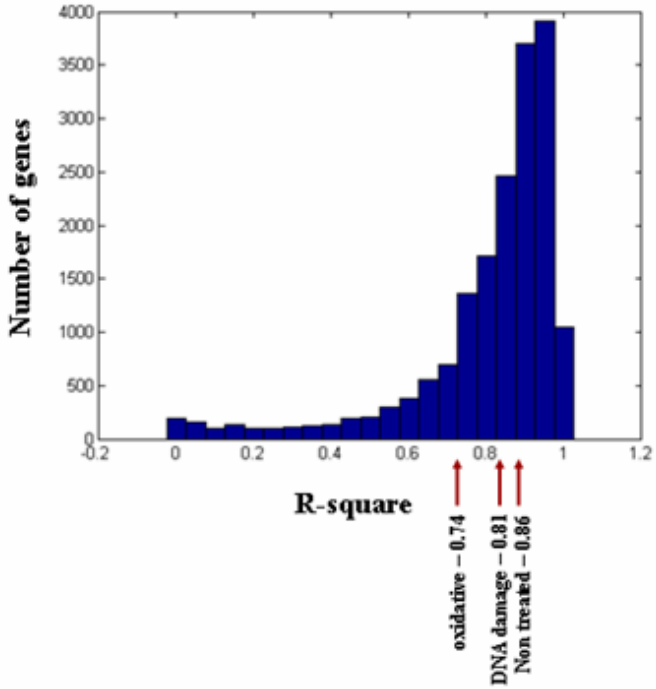


Figure S7: **Most genes decay in a constant rate during the measured time course:** Histogram of R-square ($1-\frac{\sum_i (y_i - f(x_i))^2}{\sum_i (y_i - \hat{y})^2}$) values for the fit to a constant rate decay model for all genes in all three conditions together. The values represent the goodness 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.

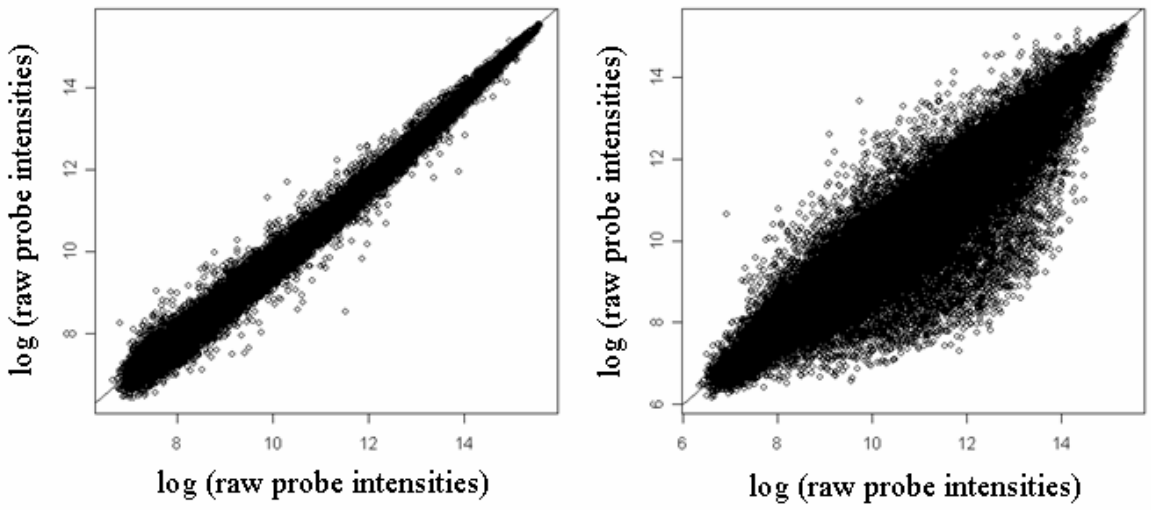


Figure S8: **Dot plot of replicate arrays and consecutive arrays:** Dot plot of raw probe intensities, left: correlation between replicate arrays, right: correlation between two different time points (time point zero and time point 20 minutes).

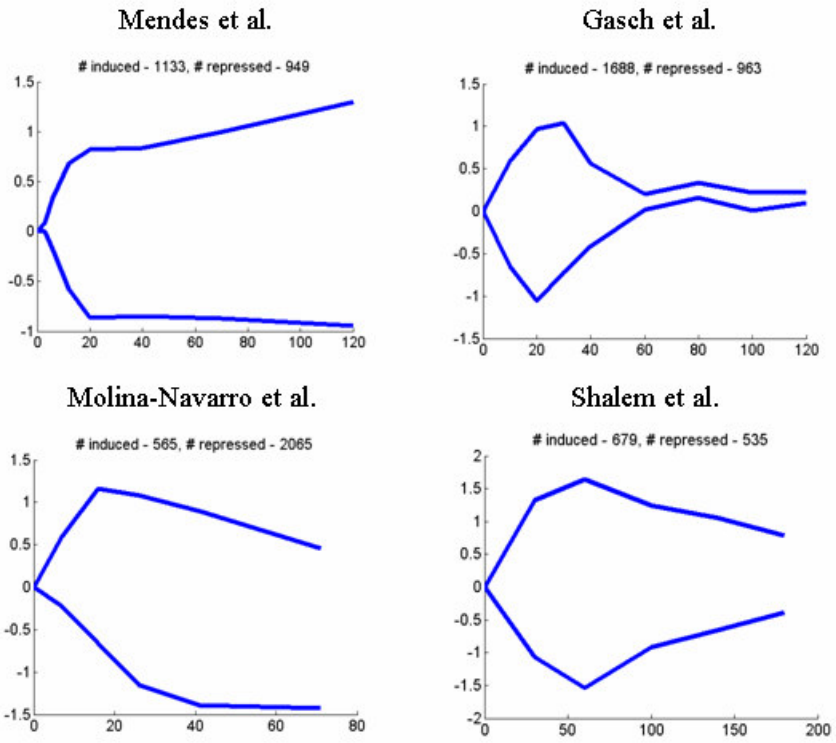


Figure S9: **Mean change in mRNA abundance following oxidative stress:** For each on the four works that measured changes in mRNA abundance following oxidative stress the mean profile of all responsive genes (that respond at least two fold relative to the reference time point) is plotted as a function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

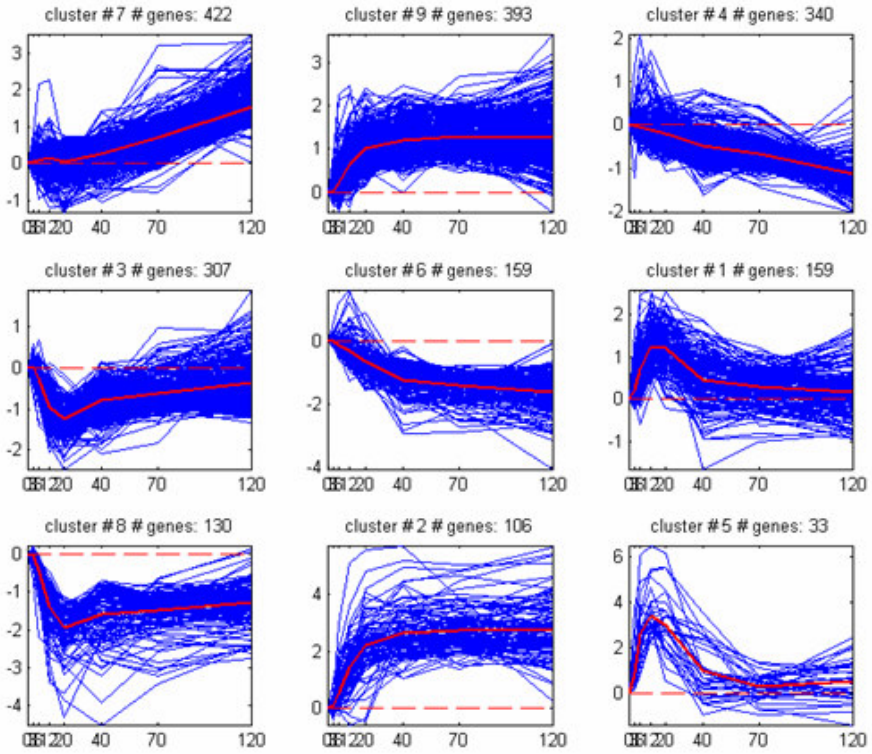


Figure S10: **Clustering of mRNA abundance changes following oxidative stress (data of Mendes et al.):** Responsive gene profiles of change in mRNA abundance following oxidative stress in the data of Mendes et al. are clustered to show two types of behaviors. Clustering was performed using k-means with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

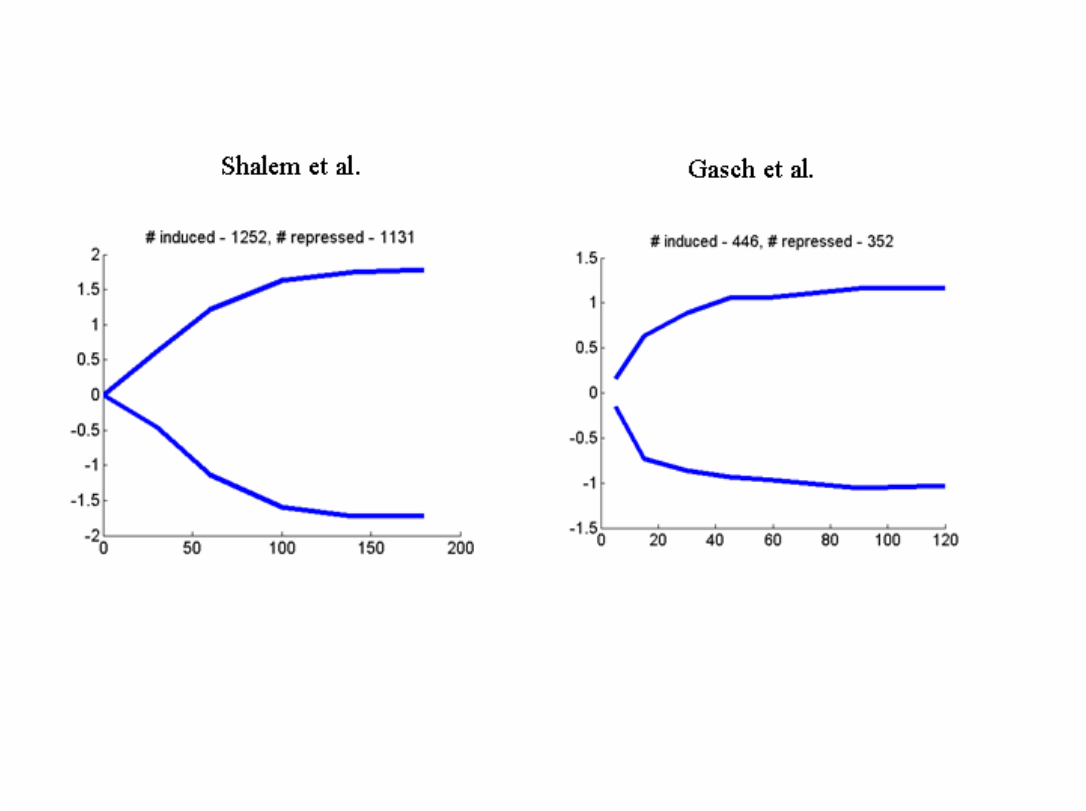


Figure S11: **Mean change in mRNA abundance following DNA damage:** The mean profile of all responsive genes (that respond at least two fold relative to the reference time point) is plotted as a function of time, separately for induced and repressed genes. Our data is compared to the data of Gasch et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

Figure S12 **(A)**:

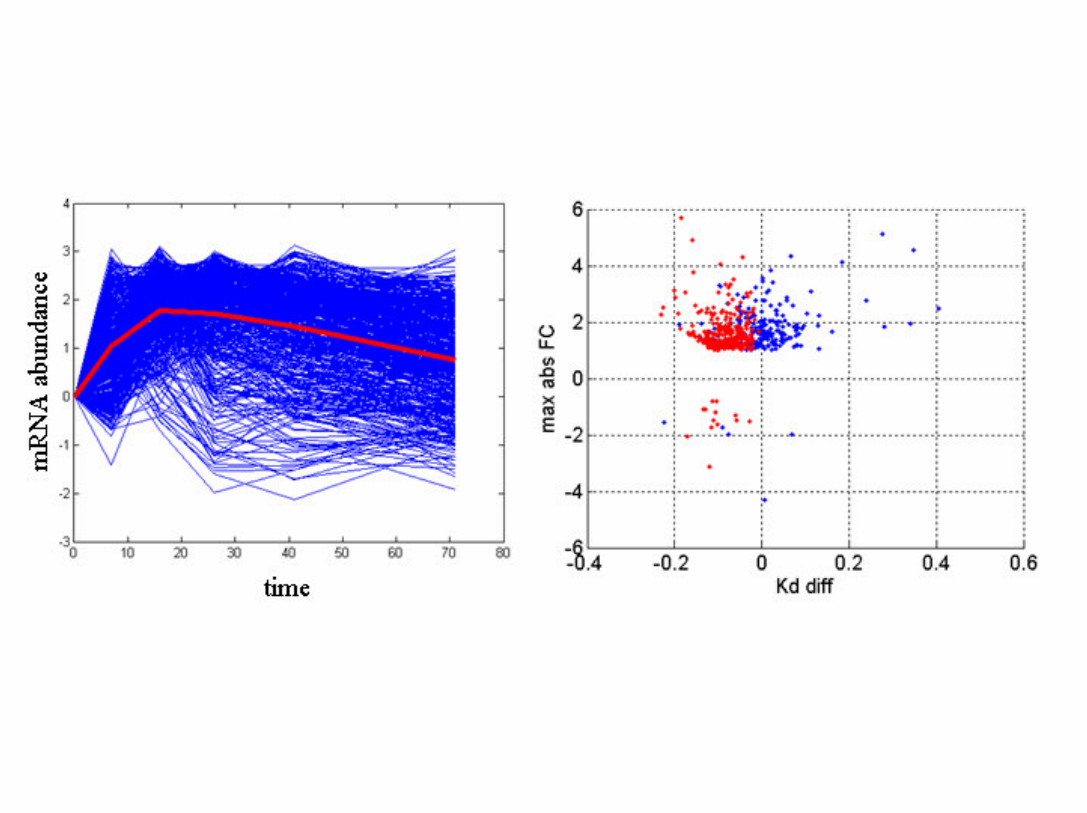


Figure S12 **(B)**:

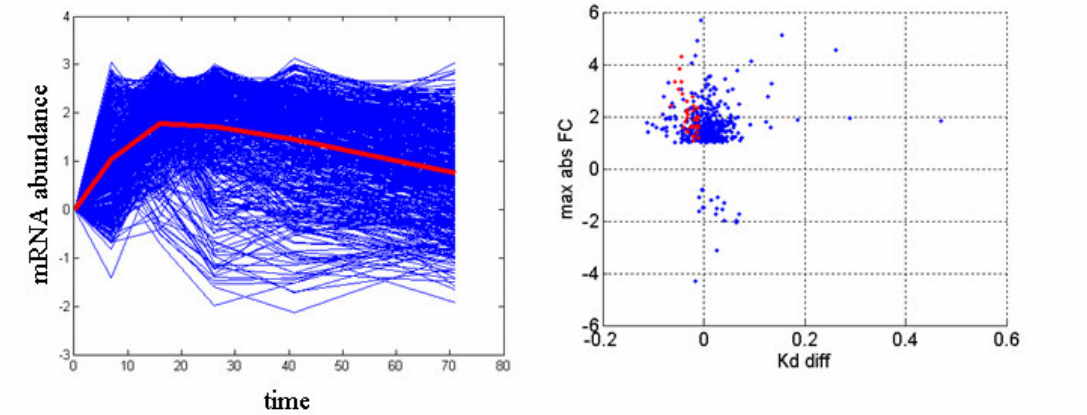


Figure S12 **(C)**:

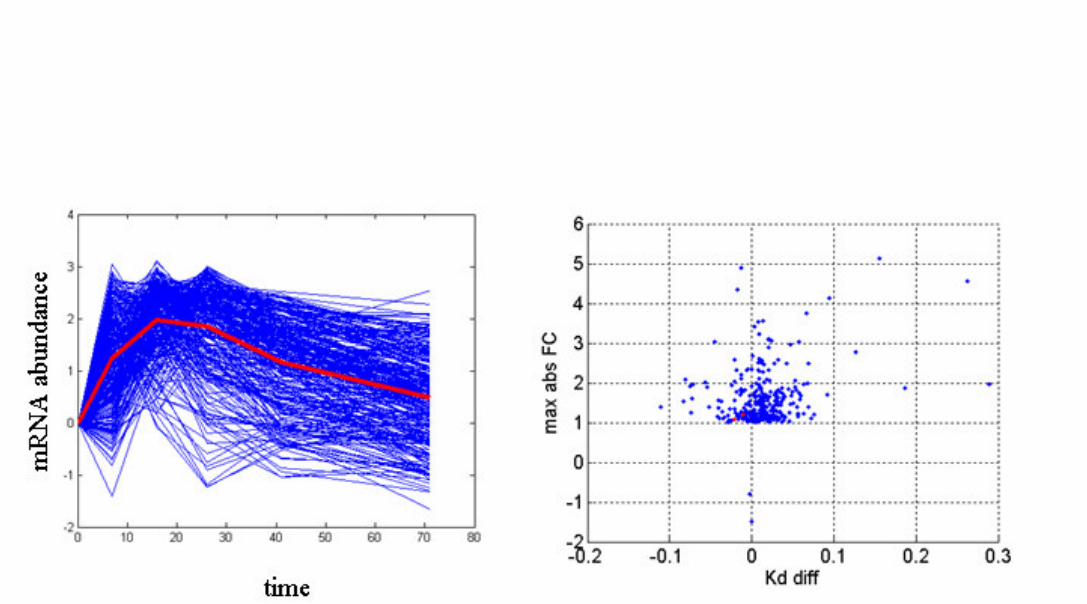


Figure S12 **(D)**:

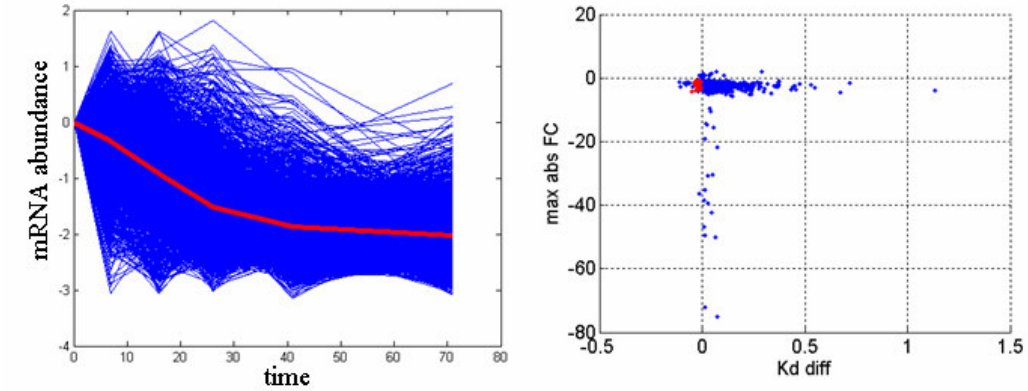


Figure S12: **Change in mRNA abundance Vs. change in Kd values (Molinha-Navarro et al. data): (A)** mRNA abundance profiles (standard deviation and mean normalized) of all induced genes are plotted alongside with the relationship between the maximal fold change to the difference in Kd in time point 7 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.

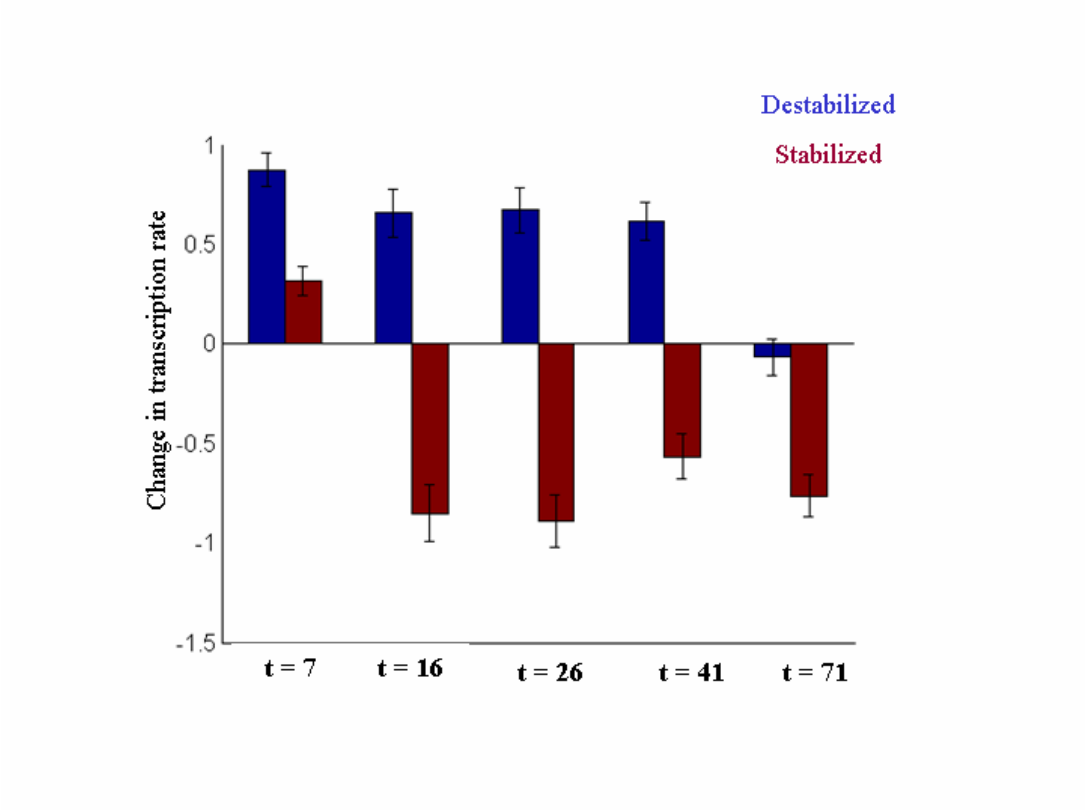


Figure S13: **Changes in transcription rates for stabilized and destabilized genes:** Bar plot of changes in transcription rates at each time point for stabilized and destabilized genes (log2 of half life ratio of above 0.8 or below -0.8), mRNA abundance and stability values are taken from our data while 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)**

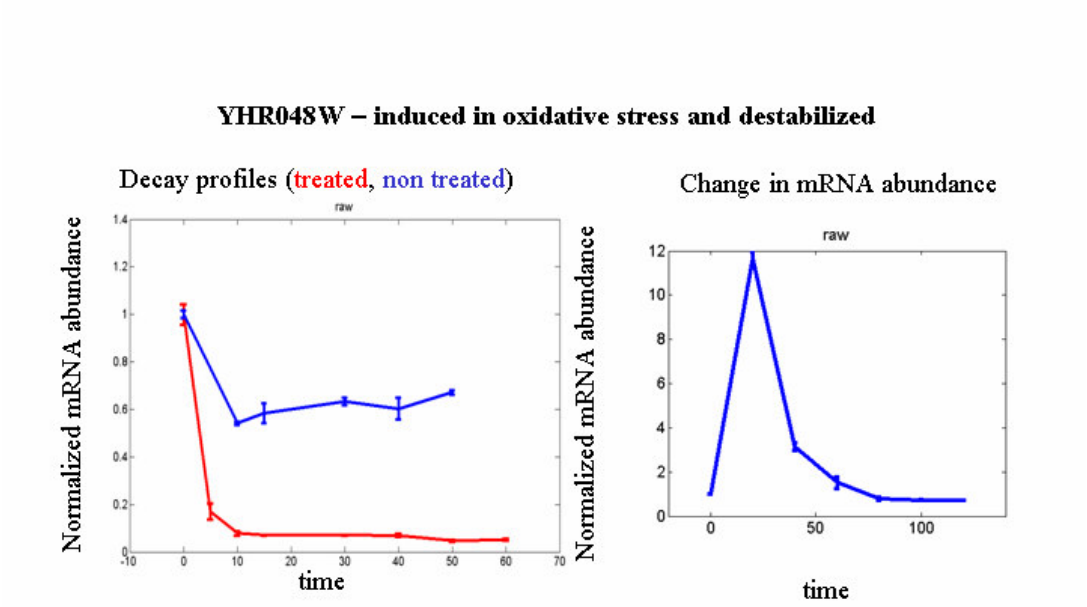


Figure S14 **(B)**

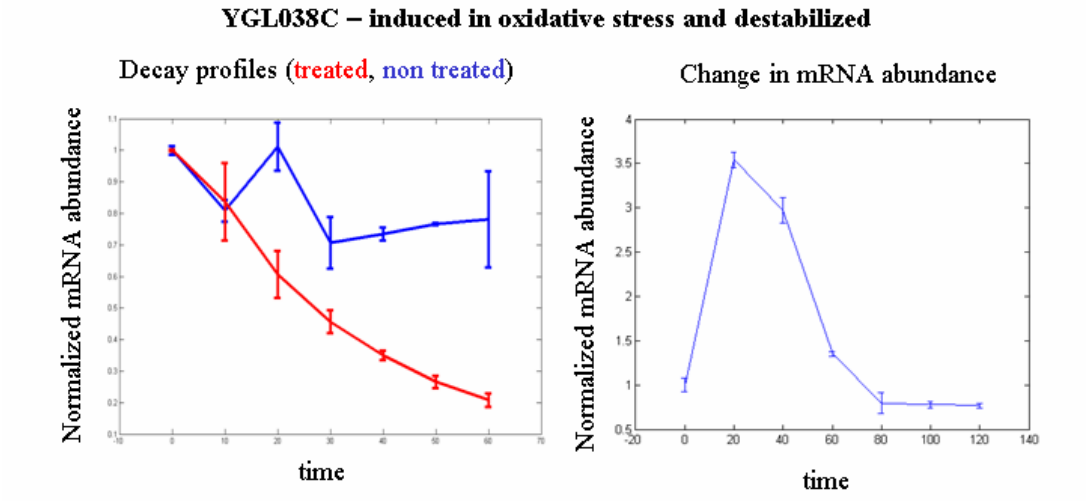


Figure S14 **(C)**

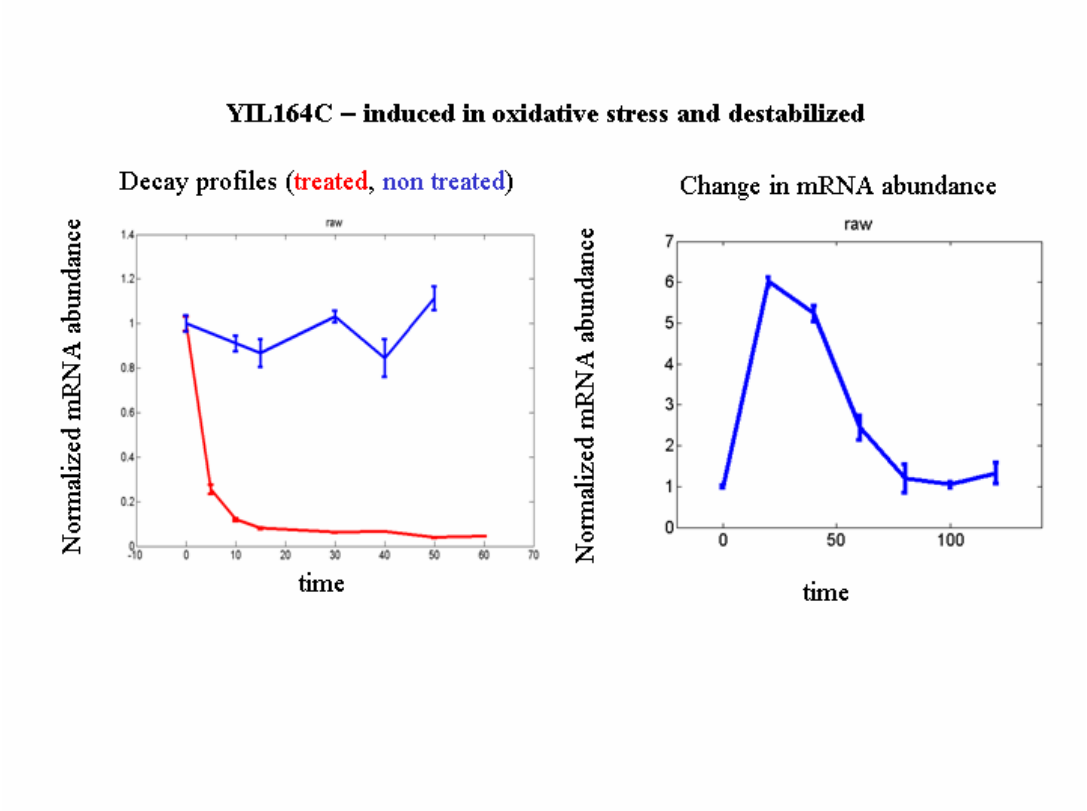


Figure S14 **(D)**

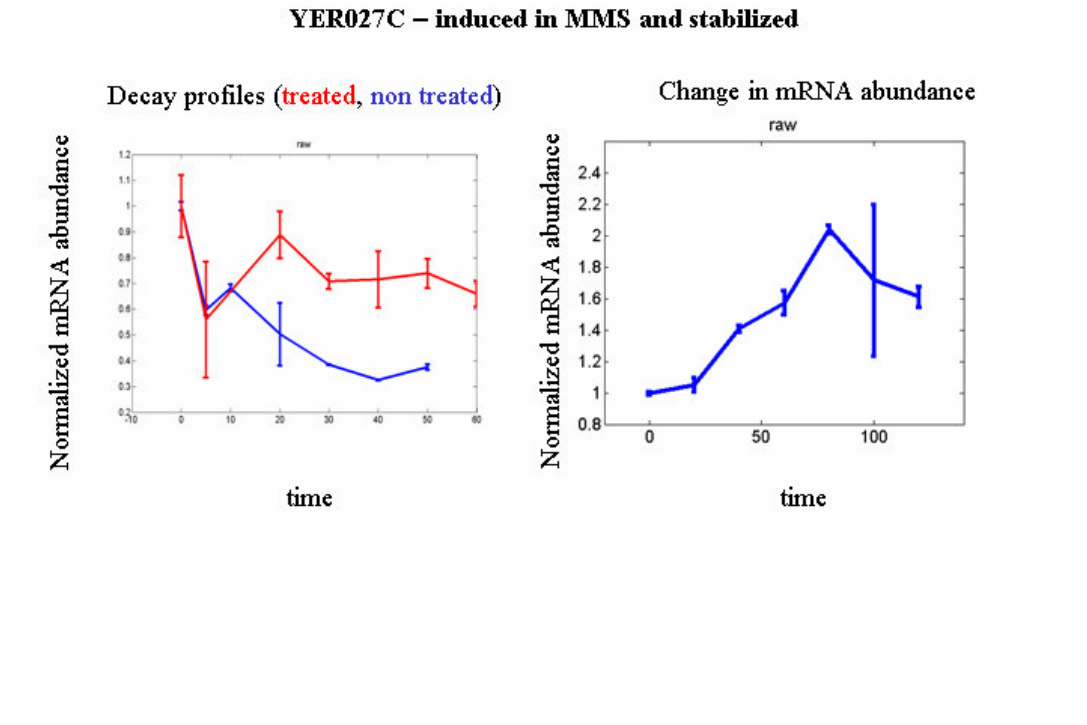


Figure S14: **Validation of microarray results using Real Time PCR for a few specific genes:** Real time PCR was used to validate microarray results, both mRNA abundance and decay profiles, for a few candidate genes (see Materials and Methods).

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.

Appendix B – supplementary data for the paper "Transcriptome kinetics is governed by a genome-wide coupling of mRNA production and degradation: A role for RNA Pol II"

(Figure and table legends are provided in the paper itself)

Figure S1

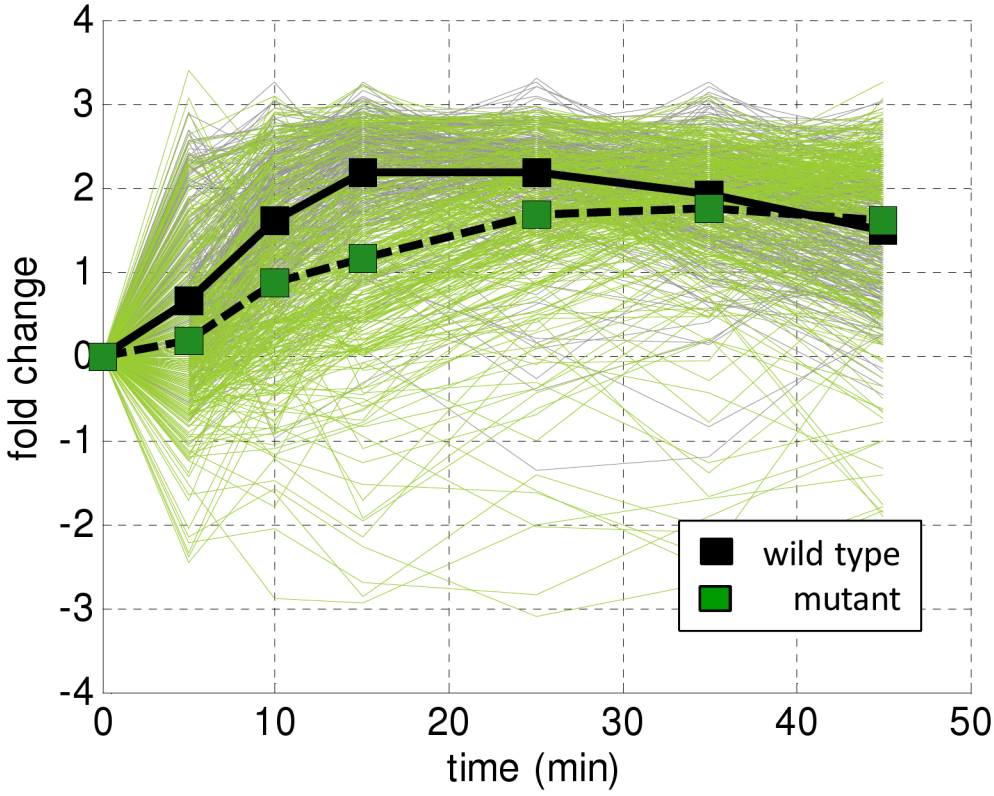

4
3
2
1
0
-1
-2
-3
-4
fold change
0
10
20
30
40
50
time (min)
wild type
mutant

Figure S2

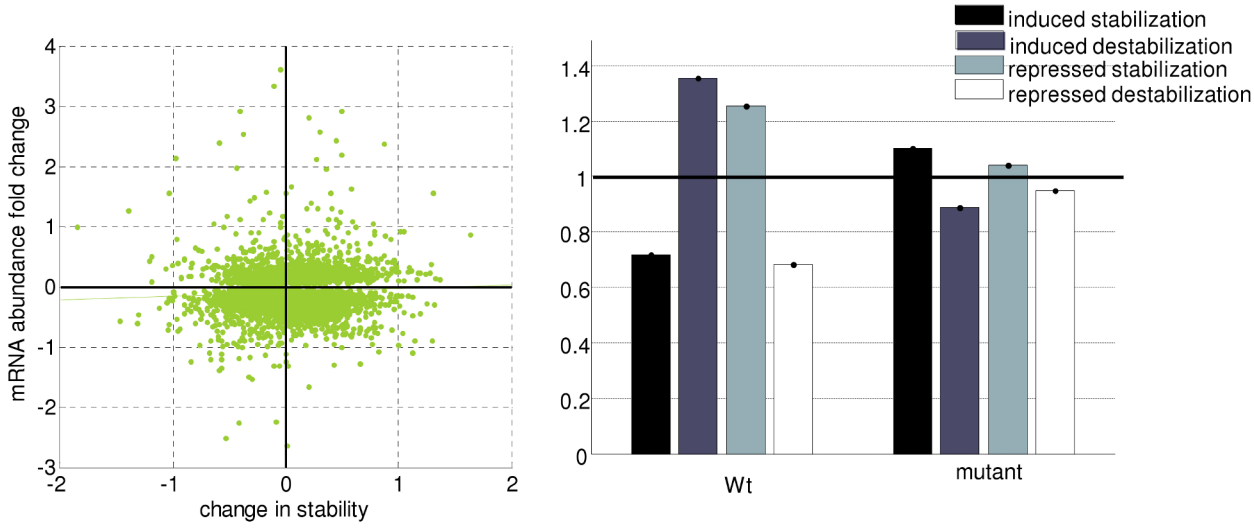

mRNA abundance fold change
change in stability
4
3
2
1
0
-1
-2
-3
-2
-1
0
1
2
induced stabilization
induced destabilization
repressed stabilization
repressed destabilization
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Wt
mutant

Figure S3

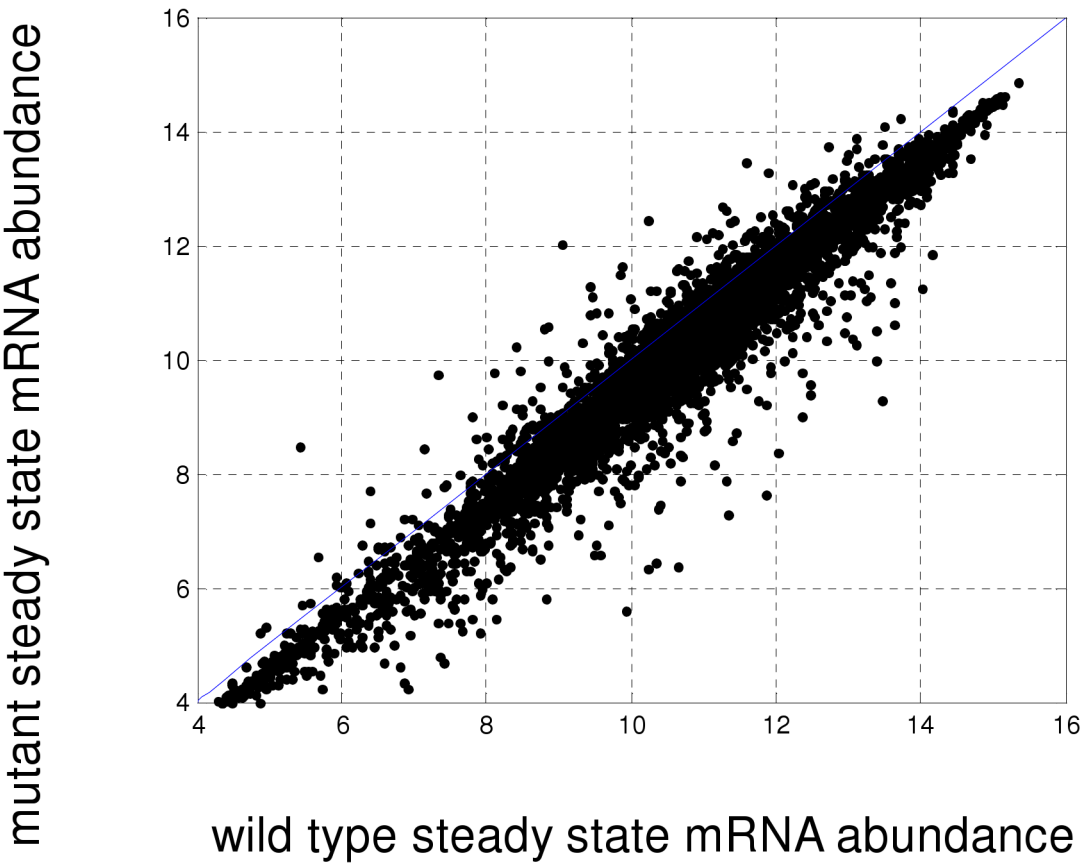

mutant steady state mRNA abundance
wild type steady state mRNA abundance
4
6
8
10
12
14
16

Figure S4

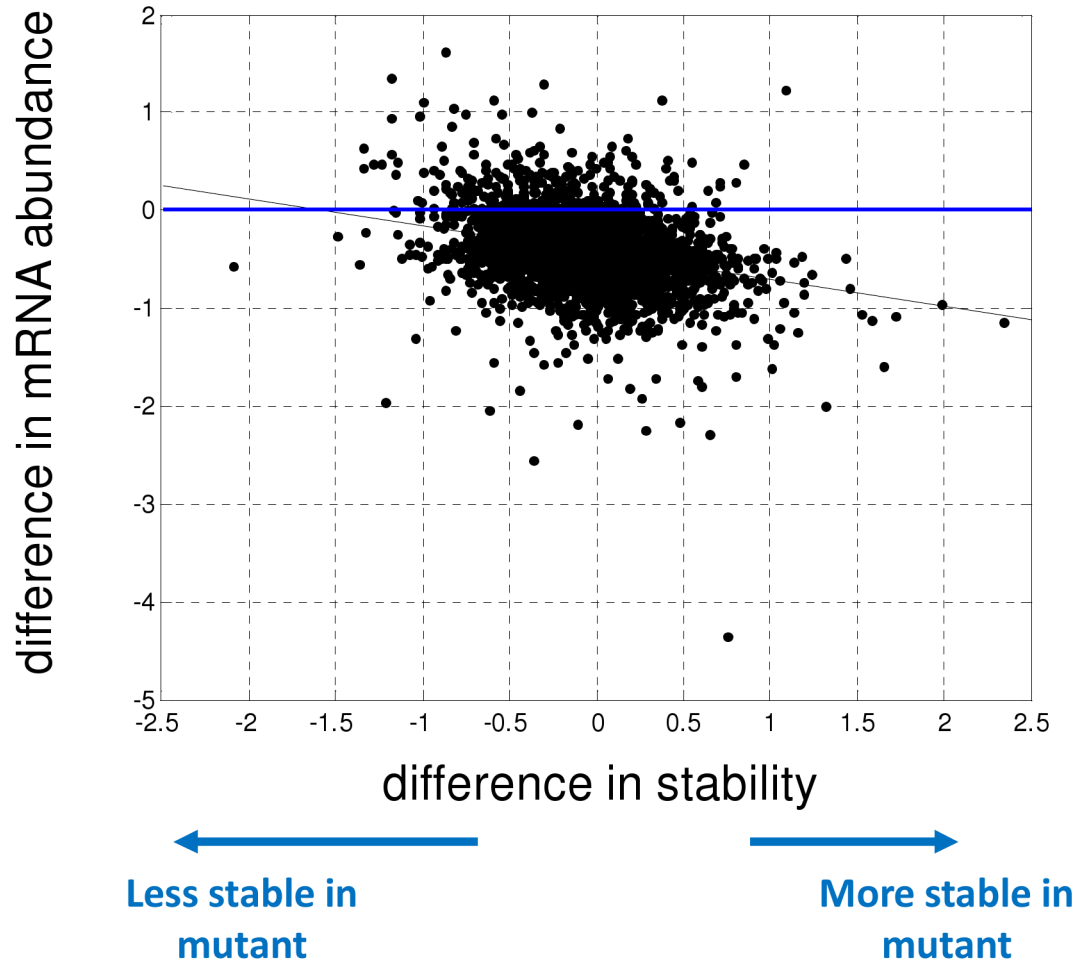
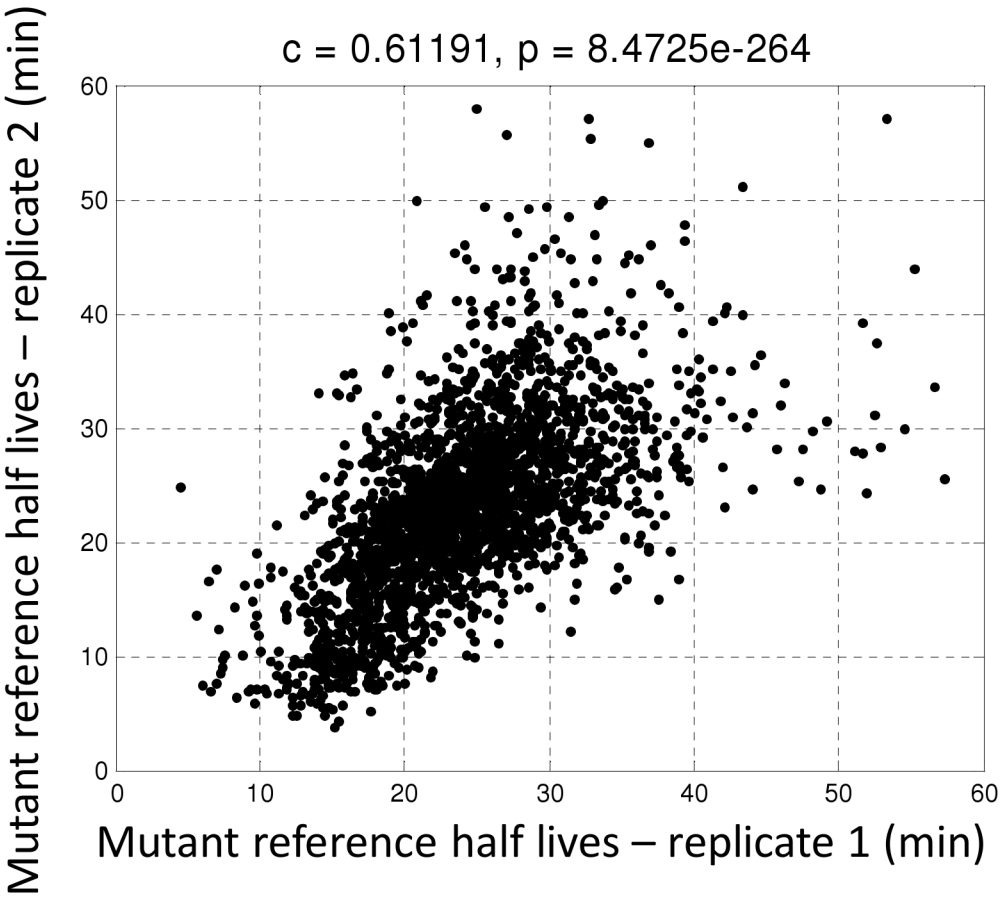

difference in mRNA abundance
2
1
0
-1
-2
-3
-4
-5
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
difference in stability
Less stable in mutant
More stable in mutant

Figure S5


c = 0.61191, p = 8.4725e-264
Mutant reference half lives – replicate 2 (min)
Mutant reference half lives – replicate 1 (min)
0
10
20
30
40
50
60

GO analysis for clustering in figure 3, all categories passed FDR with q value of 0.05

	categories	# group	# gene set	# intersecti	p-value
Cluster 1	energy reserve metabolism	296	36	14	8.32E-10
	generation of precursor metabolites and	296	230	35	3.31E-09
	carbohydrate metabolism	296	200	31	1.80E-08
	oxidoreductase activity	296	254	35	4.62E-08
	biological process unknown	296	1449	112	3.52E-07
	trehalose metabolism	296	7	5	6.48E-06
	glycogen metabolism	296	29	9	8.48E-06
	glucan metabolism	296	40	10	2.24E-05
	carbohydrate biosynthesis	296	67	13	2.59E-05
	carbohydrate catabolism	296	52	11	4.72E-05
	glucosyltransferase activity	296	15	6	5.71E-05
	disaccharide metabolism	296	10	5	6.85E-05
	sulfur utilization	296	10	5	6.85E-05
	peroxidase activity	296	17	6	0.000129
	UDP-glycosyltransferase activity	296	24	7	0.000138
	pyridoxine metabolism	296	8	4	0.000396
	siderophore transporter activity	296	4	3	0.000508
	protease inhibitor activity	296	4	3	0.000508
	trehalose biosynthesis	296	4	3	0.000508
Cluster 2	response to chemical stimulus	168	232	25	8.48E-09
	response to abiotic stimulus	168	305	26	5.08E-07
	response to oxidative stress	168	53	9	1.73E-05
	response to inorganic substance	168	21	6	2.05E-05
	oxidoreductase activity	168	254	20	3.84E-05
	response to drug	168	38	7	8.97E-05
	iron ion homeostasis	168	30	6	0.000181
	aldehyde metabolism	168	20	5	0.00021
cluster 3	ribosome biogenesis	359	210	107	2.52E-77
	nucleolus	359	223	106	1.83E-72
	rRNA metabolism	359	172	82	2.25E-55
	nuclear lumen	359	517	118	8.93E-41
	RNA processing	359	332	81	2.51E-29
	organelle lumen	359	719	123	3.10E-29
	RNA metabolism	359	461	96	4.33E-29
	35S primary transcript processing	359	74	35	1.70E-23
	nuclear part	359	897	126	1.11E-21
	processing of 20S pre-rRNA	359	43	26	2.50E-21
	nucleolar part	359	98	36	1.30E-19
	non-membrane-bound organelle	359	854	117	5.06E-19
	snoRNA binding	359	26	19	3.04E-18
	small nucleolar ribonucleoprotein comp	359	57	25	4.13E-16
	ribosomal large subunit biogenesis	359	21	14	7.49E-13
	ribosome assembly	359	62	22	4.92E-12
	ribosomal large subunit assembly and r	359	40	16	5.41E-10
	RNA binding	359	226	38	9.05E-09
	RNA helicase activity	359	41	14	7.23E-08
	rRNA binding	359	11	7	8.78E-07
	RNA-dependent ATPase activity	359	26	10	1.59E-06
	ribosome export from nucleus	359	21	9	1.84E-06
	protein complex assembly	359	147	25	2.87E-06
	RNA methyltransferase activity	359	23	9	4.57E-06
	RNA modification	359	56	14	5.06E-06
	methyltransferase activity	359	82	17	7.85E-06

	tRNA methyltransferase activity	359	15	7	1.38E-05
	S-adenosylmethionine-dependent meth	359	62	14	1.80E-05
	biopolymer metabolism	359	1457	124	2.53E-05
	ATP-dependent helicase activity	359	42	11	3.31E-05
	RNA methylation	359	17	7	3.73E-05
	ribosomal large subunit export from nuc	359	8	5	4.25E-05
	cyclin-dependent protein kinase regulat	359	23	8	4.31E-05
	small subunit processome	359	5	4	6.89E-05
	90S preribosome	359	5	4	6.89E-05
	helicase activity	359	90	16	0.000106
	regulation of cyclin-dependent protein k	359	14	6	0.000106
	snoRNA metabolism	359	15	6	0.000168
	preribosome	359	10	5	0.000172
	processing of 27S pre-rRNA	359	11	5	0.0003
	rRNA modification	359	17	6	0.000373
	tRNA (guanine) methyltransferase activ	359	7	4	0.000436
	nucleolar preribosome	359	8	4	0.00083
	spermine transporter activity	359	4	3	0.0009
	DNA-directed RNA polymerase I compl	359	14	5	0.001114
	nucleic acid binding	359	474	46	0.001232

Appendix C – supplementary figures for the paper "The effect of 3' UTR sequences on protein expression in yeast"

(Figure legends are provided in the paper itself)

Figure S1

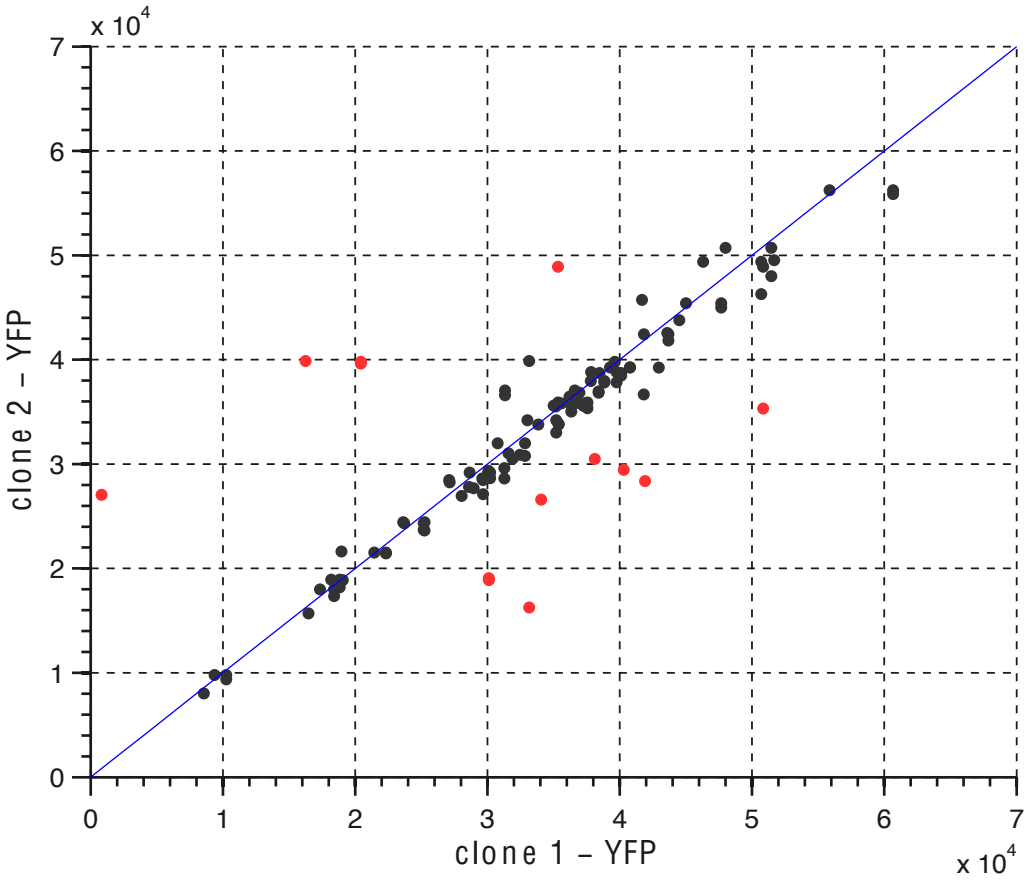
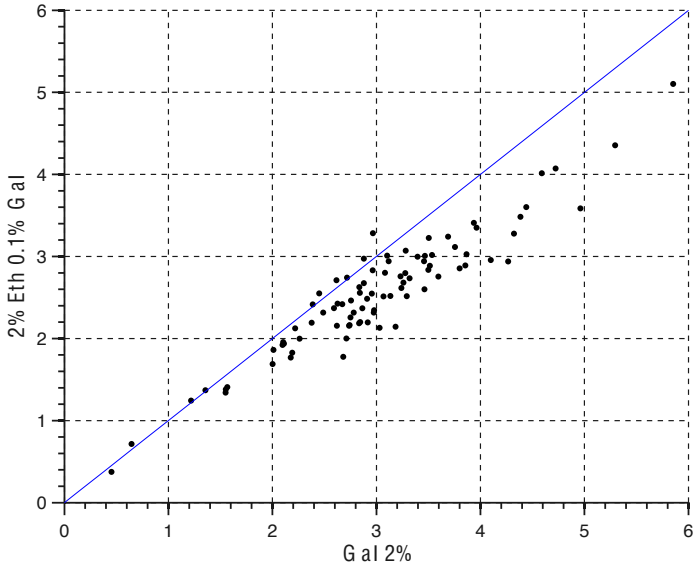
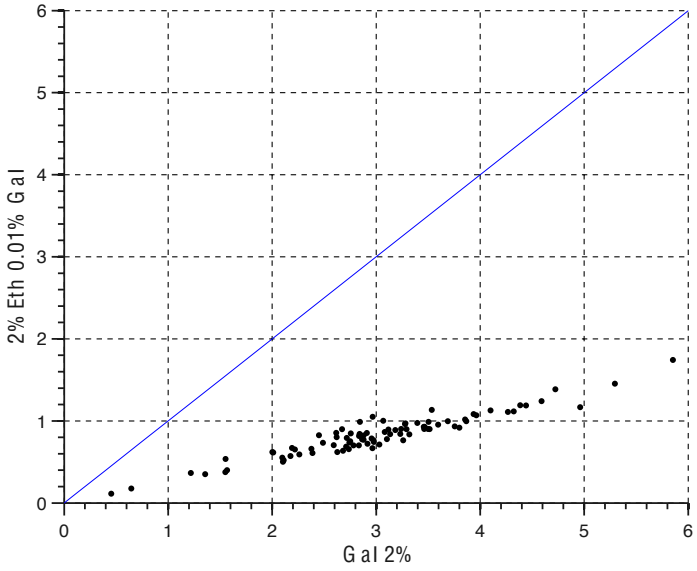
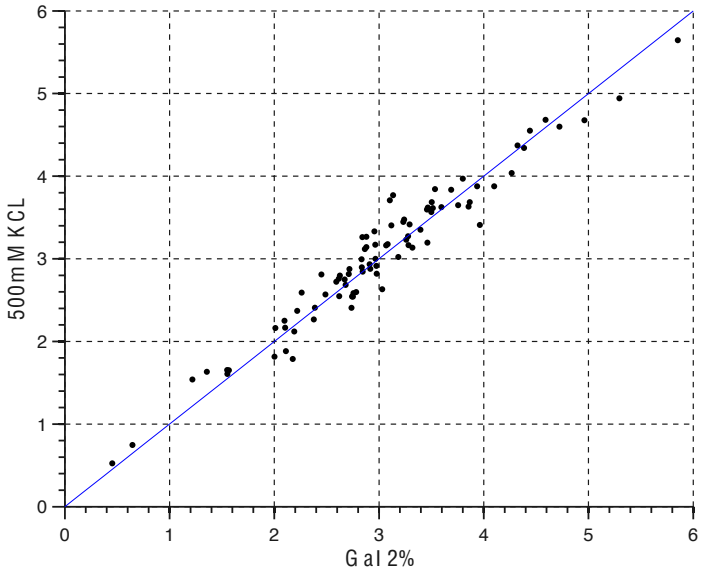
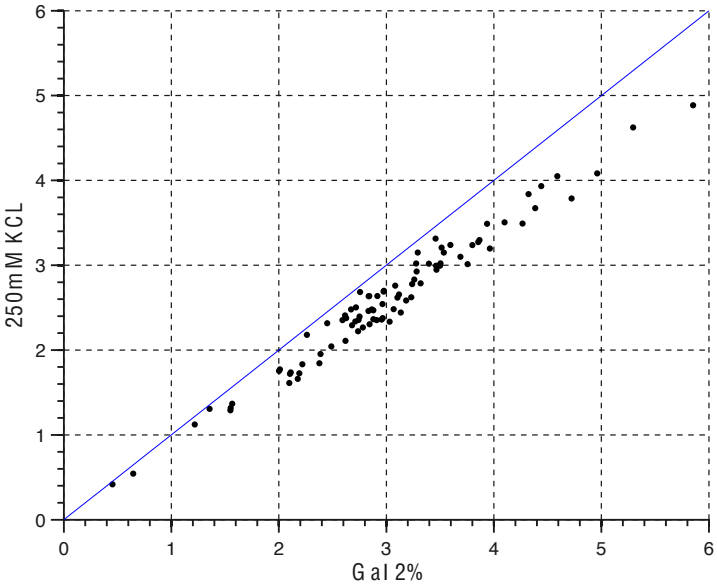


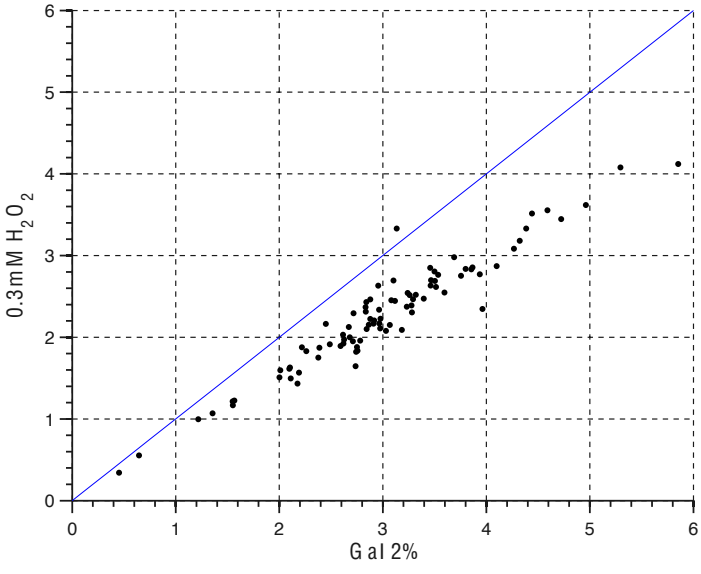
Figure S2


2% Eth 0.1% Gal
Gal 2%


2% Eth 0.01% Gal
Gal 2%


500mM KCL
Gal 2%


250mM KCL
Gal 2%


0.3mM H2O2
Gal 2%

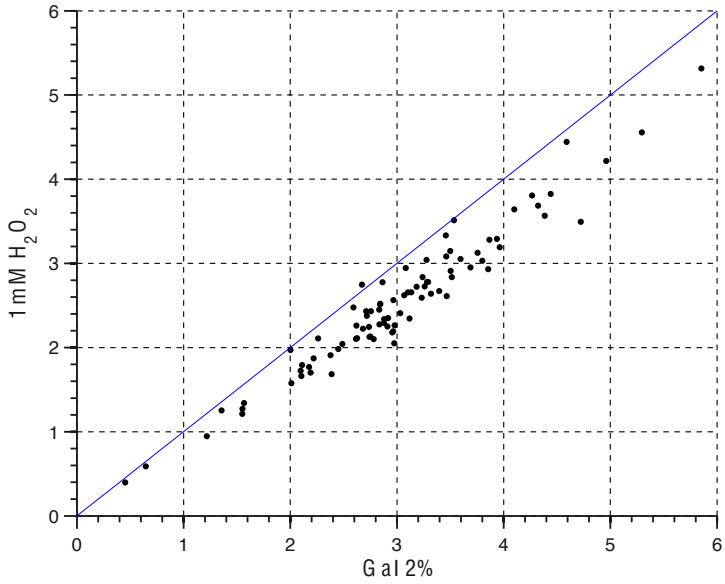

1mM H2O2
Gal 2%

Figure S3

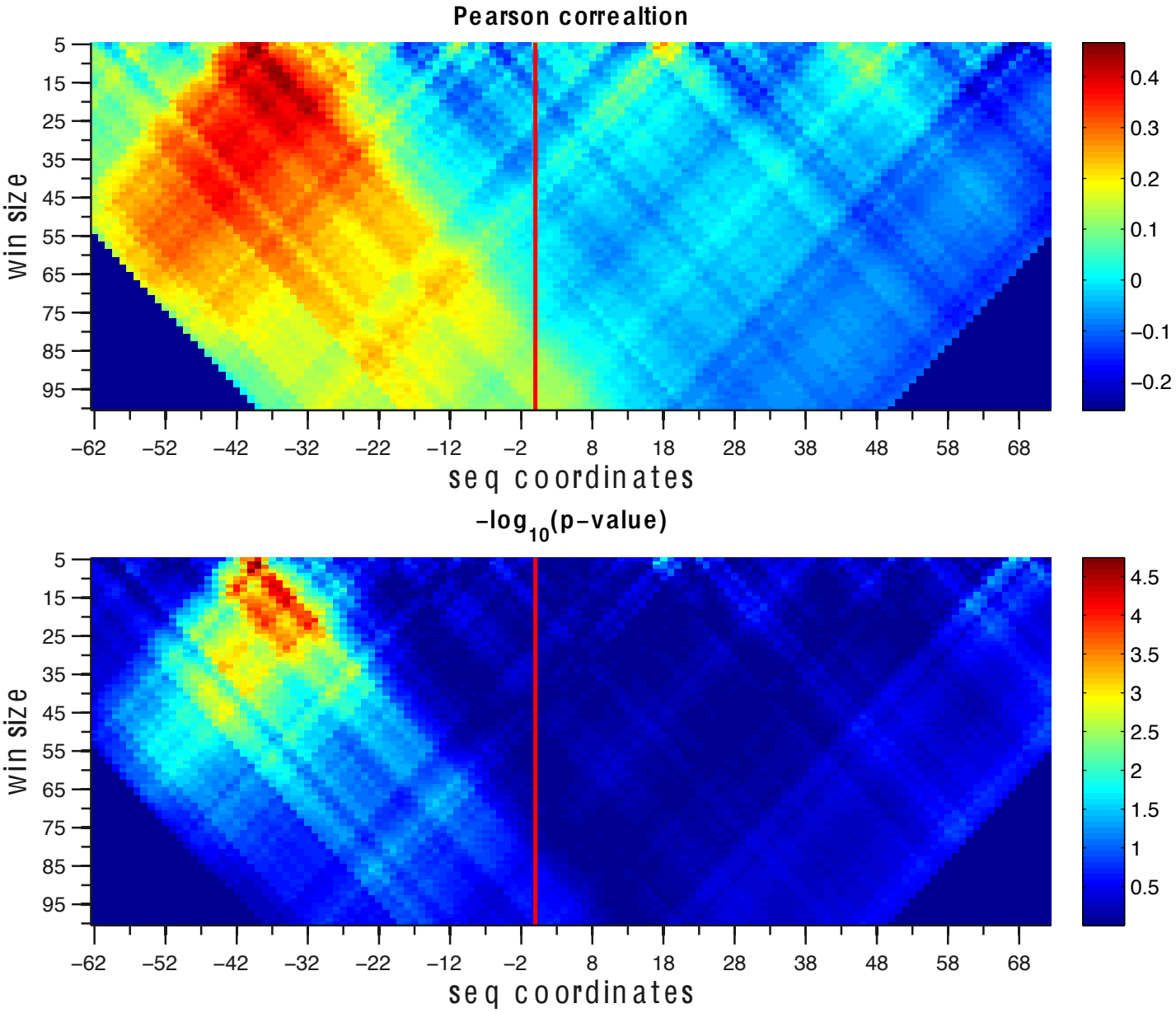

Pearson correaltion
win size
5
15
25
35
45
55
65
75
85
95
−62
−52
−42
−32
−22
−12
−2
8
18
28
38
48
58
68
seq coordinates
0.4
0.3
0.2
0.1
0
−0.1
−0.2
−log10(p−value)
win size
seq coordinates
4.5
4
3.5
3
2.5
2
1.5
1
0.5

Figure S4

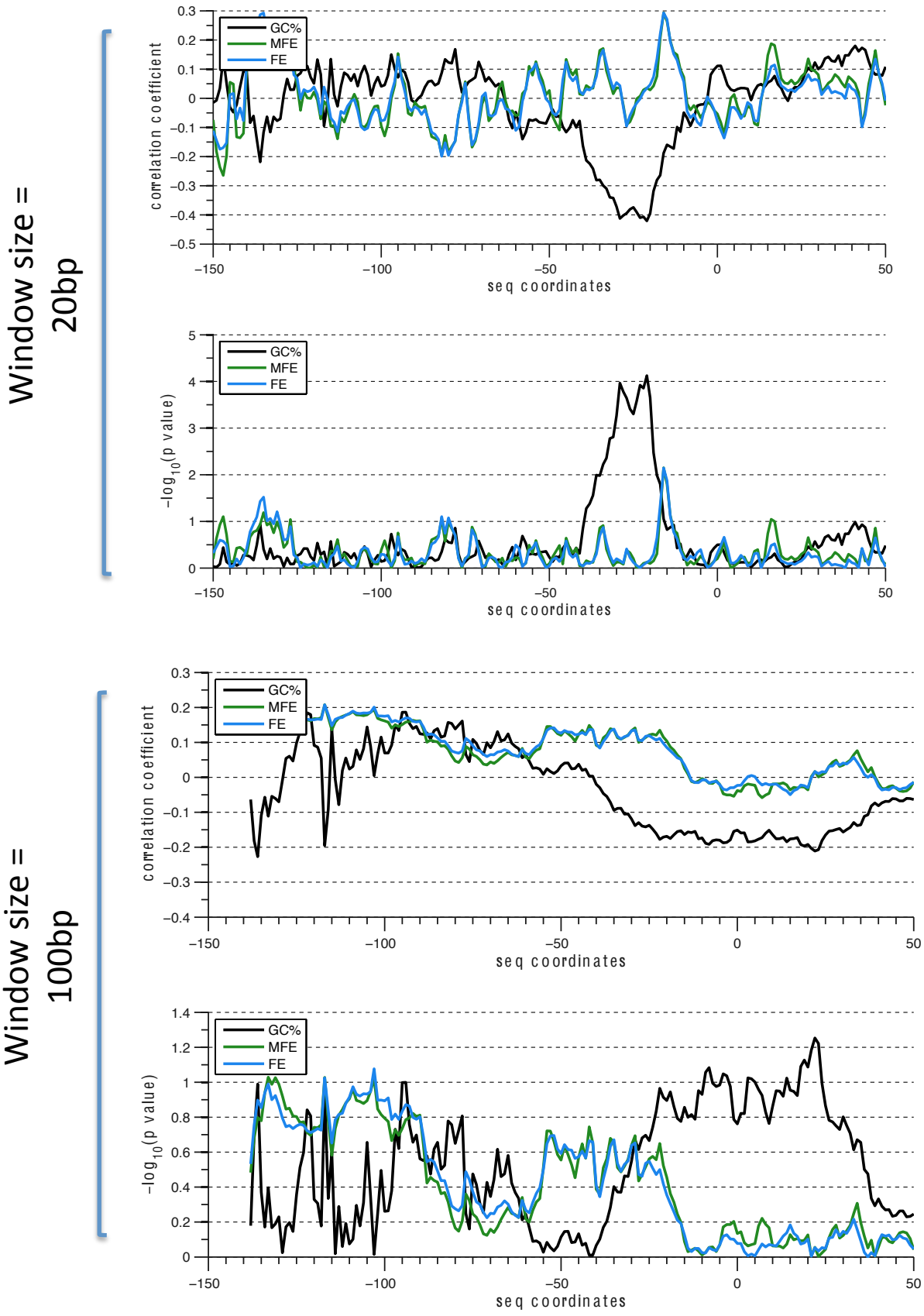

Window size = 20bp
GC%
MFE
FE
correlation coefficient
seq coordinates
−log10(p value)
Window size = 100bp

Figure S5

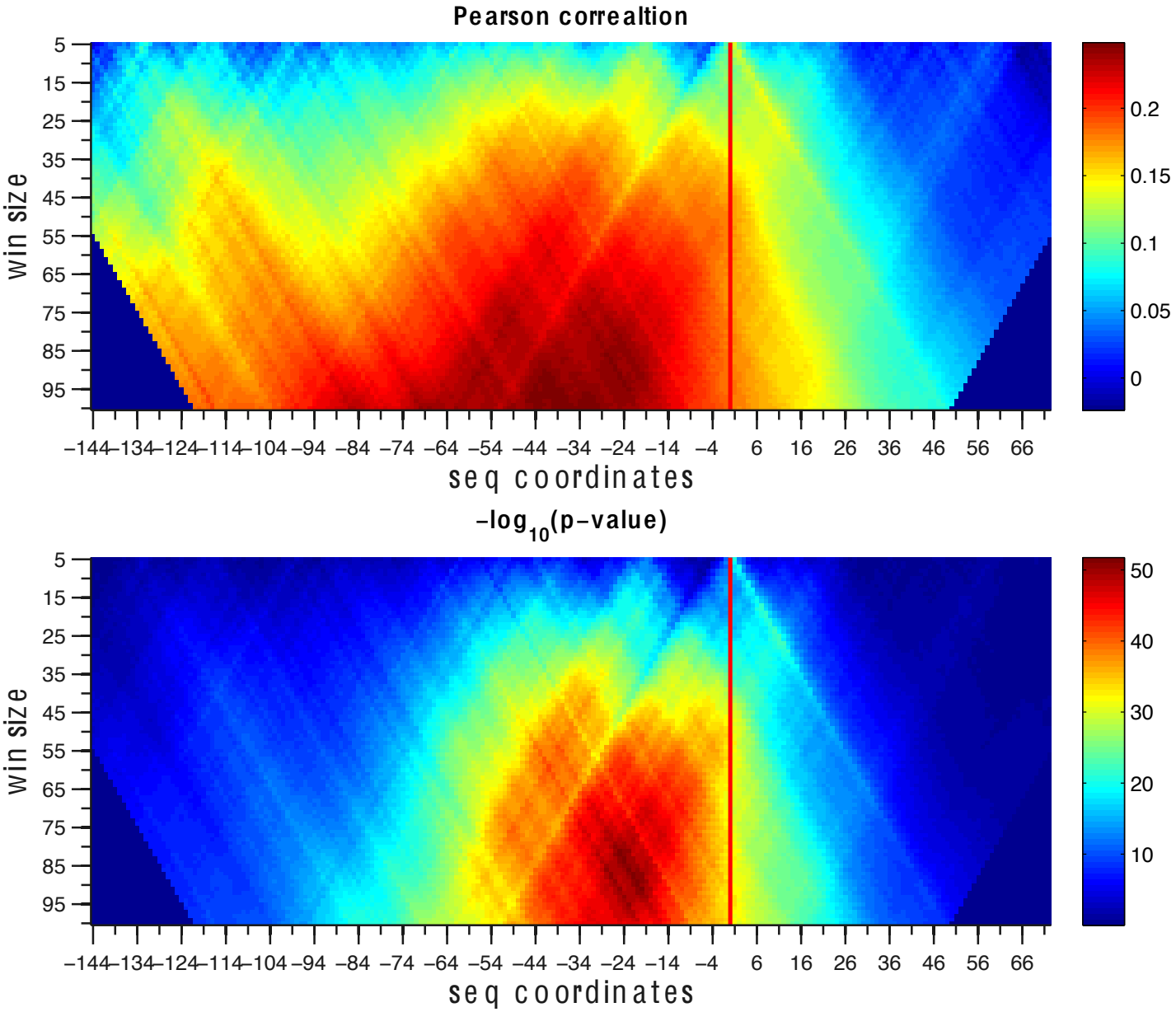
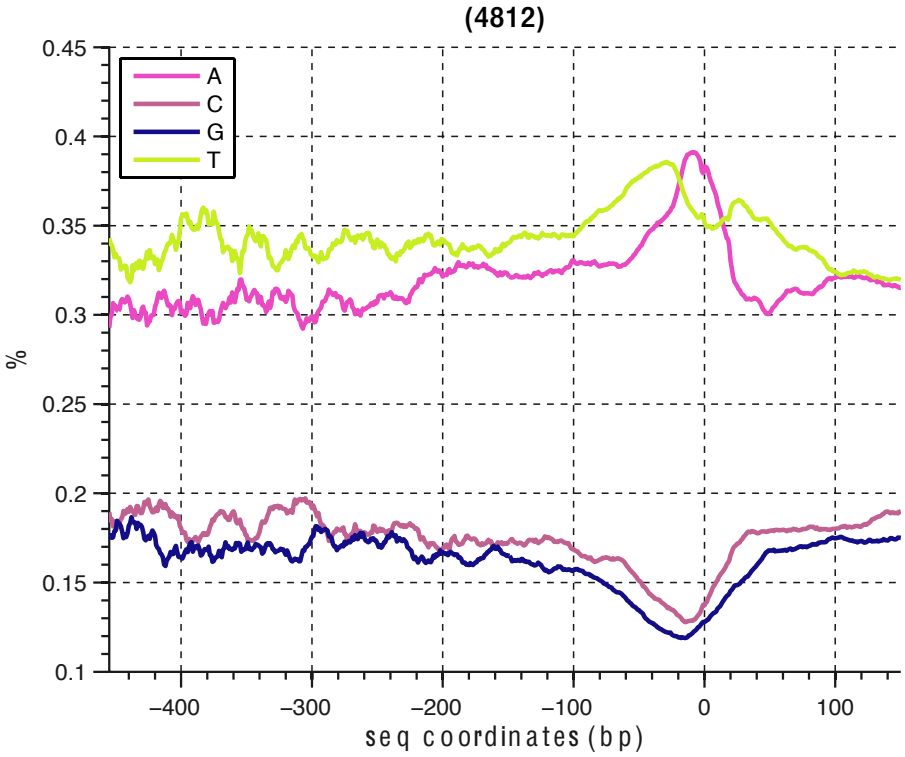

Pearson correaltion
win size
5
15
25
35
45
55
65
75
85
95
−144 −134 −124 −114 −104 −94 −84 −74 −64 −54 −44 −34 −24 −14 −4 6 16 26 36 46 56 66
seq coordinates
0.2
0.15
0.1
0.05
0
−log10(p−value)
win size
5
15
25
35
45
55
65
75
85
95
−144 −134 −124 −114 −104 −94 −84 −74 −64 −54 −44 −34 −24 −14 −4 6 16 26 36 46 56 66
seq coordinates
50
40
30
20
10

Figure S6


(4812)
A
C
G
T
%
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
−400
−300
−200
−100
0
100
seq coordinates (bp)

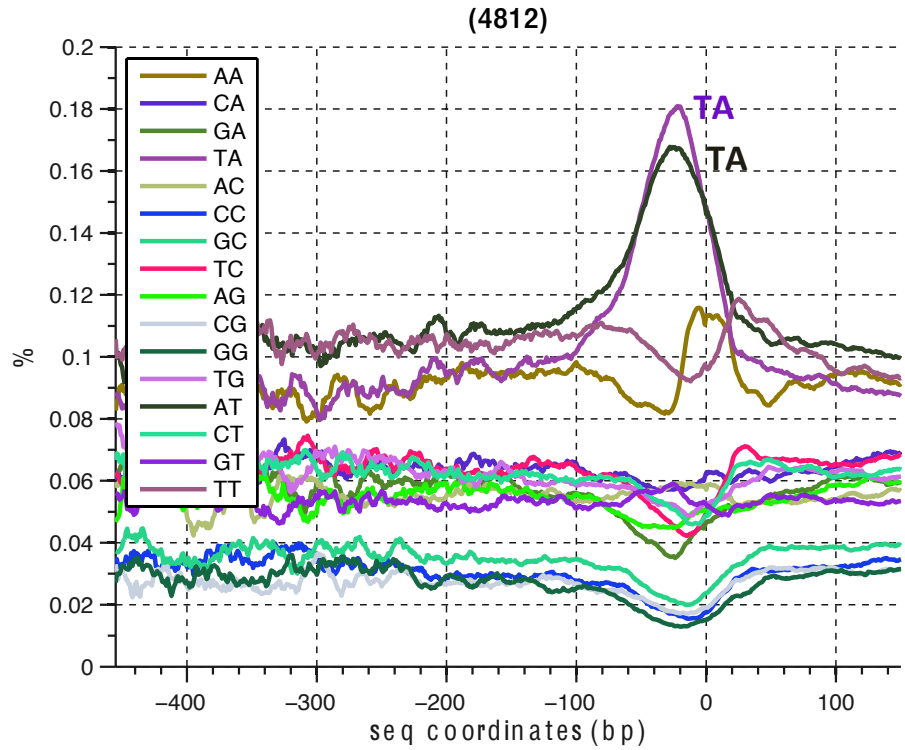

(4812)
AA
CA
GA
TA
AC
CC
GC
TC
AG
CG
GG
TG
AT
CT
GT
TT
TA
TA
%
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
−400
−300
−200
−100
0
100
seq coordinates (bp)

Figure S7

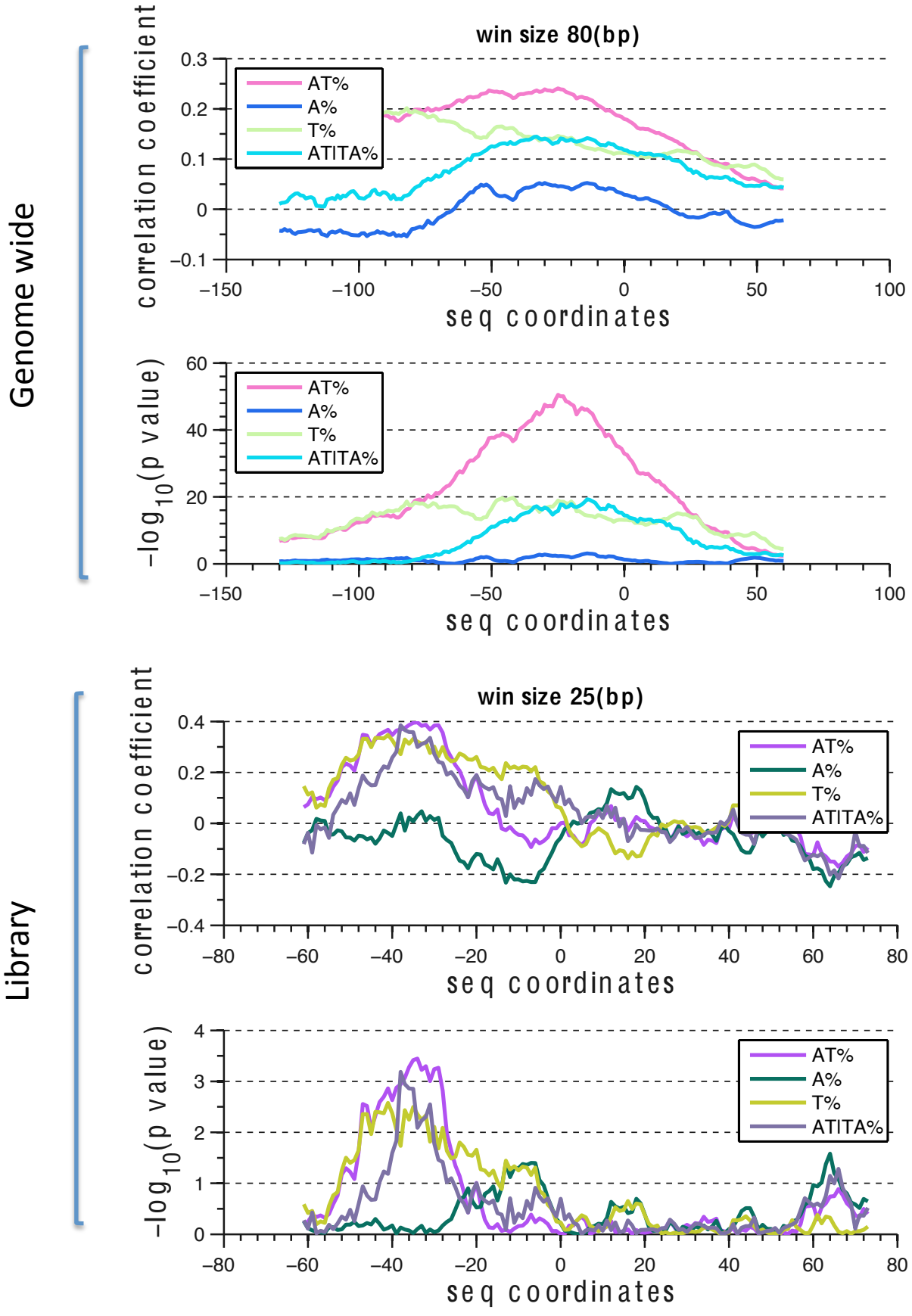

Genome wide
win size 80(bp)
correlation coefficient
AT%
A%
T%
ATITA%
seq coordinates
−log10(p value)
Library
win size 25(bp)

Figure S8

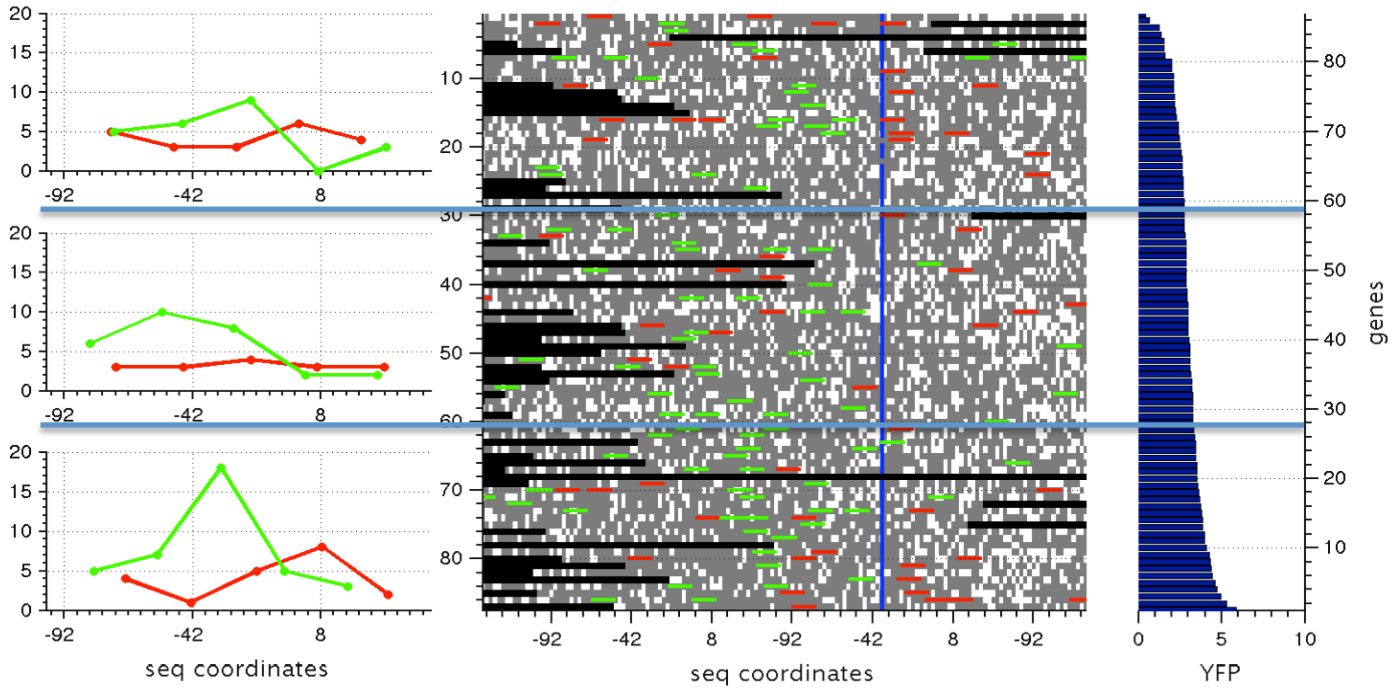
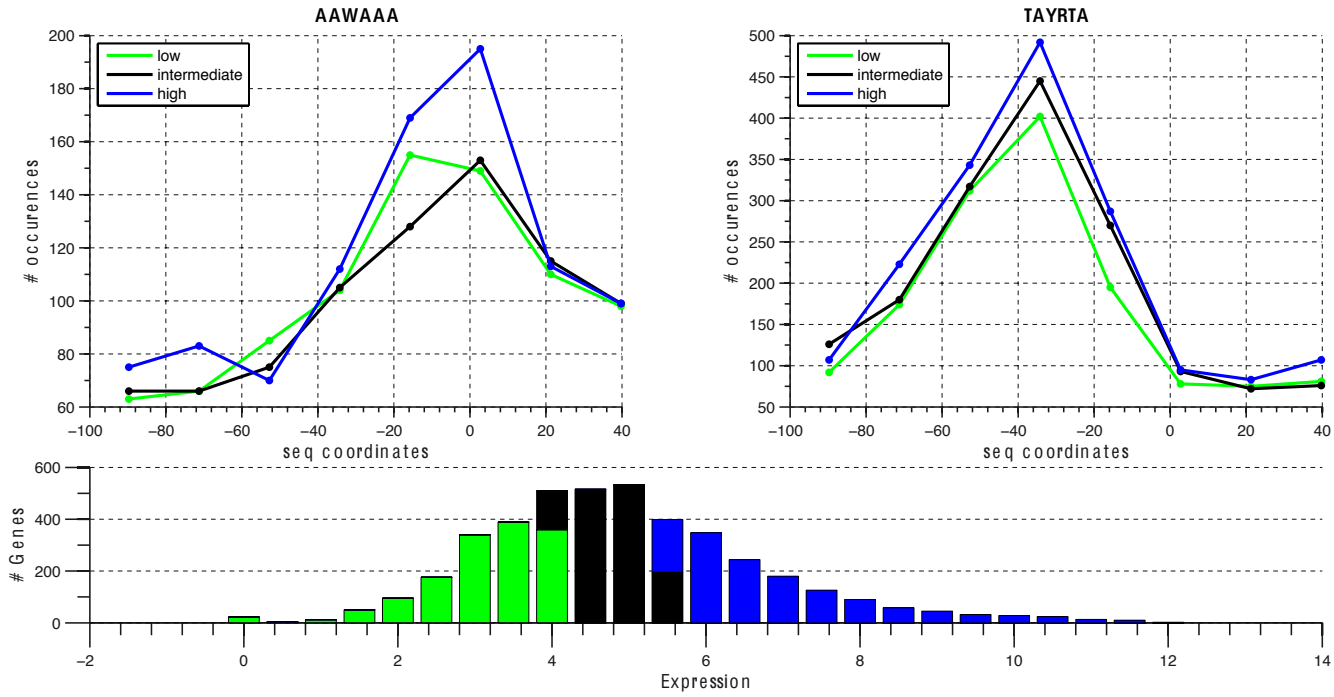


Figure S9


AAWAAA
low
intermediate
high
occurences
seq coordinates
TAYRTA
low
intermediate
high
occurences
seq coordinates
Genes
Expression