

SUPPLEMENTARY NOTES  
for  
Adaptation as a genome-wide autoregulatory  
principle in the stress response of yeast

G. De Palo<sup>†</sup>, F. Eduati<sup>‡</sup>, M. Zampieri<sup>†</sup>,  
B. Di Camillo<sup>‡</sup>, G. Toffolo<sup>‡</sup> and C. Altafini<sup>†</sup> \*

<sup>†</sup> SISSA Int. School for Advanced Studies  
via Beirut 2-4, 34014 Trieste, Italy

<sup>‡</sup> Department of Information Engineering,  
University of Padova  
via Gradenigo, 6/B, 35131 Padova, Italy

February 12, 2010

## Contents

<b>1</b>	<b>Materials</b>	<b>2</b>
<b>2</b>	<b>Open-loop models of transcription rates and rise time</b>	<b>3</b>
<b>3</b>	<b>Significance of known gene-gene interactions for the stress response time-series</b>	<b>5</b>
<b>4</b>	<b>Closed loop model with delay</b>	<b>6</b>

---

\*Corresponding author: [altafini@sissa.it](mailto:altafini@sissa.it)

<b>5</b>	<b>Closed loop model with integral feedback from the growth rate</b>	<b>6</b>
<b>6</b>	<b>Estimating the ribosome density rate constant</b>	<b>8</b>

## List of Tables

S1	Tests of significance for gene-gene interactions . . . . .	6
S2	Fitting for gene and protein time-series . . . . .	7

## List of Figures

S1	Response of a single mode with zero order synthesis term . . .	9
S2	Significance of the gene-gene interactions . . . . .	10
S3	Areas of the KEGG pathways . . . . .	11
S4	Areas of the KEGG macrocategories . . . . .	12
S5	Comparison with the “yeast metabolic cycle” . . . . .	13

## 1 Materials

Of the compendium of data in [1] we consider 5 time series: two heat shocks ( $25^\circ \rightarrow 37^\circ$  and  $29^\circ \rightarrow 33^\circ$ ), Hydrogen Peroxide, Sulfhydryl oxidizing agent, hyper-osmotic shock. The four time series from [2] represent exposure to high  $\text{Ca}^{2+}$  and  $\text{Na}^{2+}$  and combinations of these with the immunosuppressive drug FK506. The six series from [3] represent the following responses: heat shock ( $25^\circ \rightarrow 37^\circ$ ), acid (succinic acid, pH 4), alkali (Tris-HCl, pH 8.25), Hydrogen Peroxide, NaCl, Sorbitol. From [4] we consider 3 time series (heat shock  $30^\circ \rightarrow 37^\circ$ ,  $\text{H}_2\text{O}_2$  response, and transfer from glucose to glycerol as carbon sources). Finally the two series from [5] represent responses to pulses of glucose of different amplitude (for steady-state yeast on a chemostat). Other environmental stress response time series such as [6] are not considered because the sampling is too sporadic (qualitatively the results of [6] are very similar). In other time series, such as the DNA damage response [7] or the response to DTT [1, 8], the kinetic response triggered can be much slower than the one we consider here and therefore these time series are not included in this study.

The HL considered in this paper are obtained averaging the values contained in the three datasets [9, 10, 11]. The values of HL are often said to be different on different experimental conditions [12, 9]. However, HL specific to the stresses/stimuli considered here are largely missing (see [7] though).

The protein complexes were downloaded from the MPACT subsection (<http://mips.gsf.de/genre/proj/mpact/>) of the CYGD database at MIPS. Only complexes manually annotated from the literature are considered; those obtained from high throughput experiments are disregarded (according to the MIPS classification scheme these last are labeled “550”). The KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways are downloaded from <http://www.genome.jp/kegg>. Also the assembling into the 15 macrocategories discussed in the paper follows the KEGG hierarchy.

## 2 Open-loop models of transcription rates and rise time

**Models for transcription rates based on transcription factors** There is a wide literature on the use of kinetic ODE models to describe transcription processes [13, 14, 15, 16, 5, 17, 18]. The simplest such model is of the form (notation is the same as the main paper)

$$\frac{dm_i}{dt} = -\delta_i m_i + f_i \quad (\text{S1})$$

where in most cases it is assumed that the synthesis rate  $f_i$  is a function of the transcription factors  $w_i$  promoting/inhibiting  $m_i$  and it is of zero order in the mRNA abundance  $m_i$  [15, 14]. A few elementary choices for  $f_i$  are:

- $f_i$  is of zero order kinetics also in  $w_i$

$$f_i = \alpha_i \quad (\text{S2})$$

- $f_i$  is of first order in  $w_i$

$$f_i = \alpha_i \pm \beta_i w_i \quad (\text{S3})$$

- $f_i$  is of Michaelis-Menten form [16]

$$f_i = \frac{\alpha_i}{\beta_i + w_i} \quad \text{for a repressor} \quad (\text{S4})$$

$$f_i = \frac{\alpha_i w_i}{\beta_i + w_i} \quad \text{for an activator} \quad (\text{S5})$$

$$(\text{S6})$$

- $f_i$  is of Hill form [14]

$$f_i = \frac{\alpha_i}{\beta_i + w_i^n} \quad \text{for a repressor} \quad (\text{S7})$$

$$f_i = \frac{\alpha_i w_i^n}{\beta_i + w_i^n} \quad \text{for an activator.} \quad (\text{S8})$$

Given a stimulus  $u(t)$ , for example a step:

$$u(t) = \begin{cases} 0 & \text{for } t < 0; \\ 1 & \text{for } t \geq 0, \end{cases}$$

in correspondence of (S3)-(S8) one needs to specify the function  $w_i(u)$ , i.e., how an external stimulus promotes/inhibits the activity of the transcription factor(s)  $w_i$ . Very little is known in general about the functional form of  $w_i(u)$  (which may describe events such as crossing of the nuclear envelope, phosphorylations, dimerizations, etc.). Alternatively,  $f_i(\cdot)$  can be expressed as  $f_i = f_i(w_i, u) = \tilde{f}_i(w_i)u$  with  $\tilde{f}_i(w_i)$  any of (S2)-(S8).

**Open-loop response and rise time** Assuming that  $f_i$  is activated by a stimulus at time  $t = 0$ , the synthesis rate term will jump from 0 (or a basal level) to a value  $f_i(w_i)$  in response to the stimulus. As long as  $f_i$  is chosen of zero order in  $m_i$ , regardless of the functional form of  $f_i(w_i)$ , the time constant of the solution  $x(t)$  is independent of the amplitude of the input step and limited by the degradation constant  $\delta_i$ . Under these assumptions, in order to have a fast transient one must assume that  $\delta_i$  in the input response is time-varying and uncorrelated with the known experimental HL, as  $f_i$  alone, being of zero-order in  $m_i$ , cannot change the time constant of the response but only its amplitude and sign, see Fig. S1. Therefore, in order to explain the fast rising – slow decay trend observed in the transient, the  $\delta_i$  must change considerably *during* the transient [15]. Furthermore, in order to account for adaptation, models such as (S1) must assume that the synthesis term  $f_i(w_i)$  returns to its basal level in spite of the persistent stimulus. While this is easily achieved e.g. through a vanishing kernel for  $f_i(\cdot)$  [18], a clear biological justification is still missing. Moreover any such “open-loop” model must entail anyway a form of memory of the “optimal” working level for each gene, in order to achieve exact adaptation.

### 3 Significance of known gene-gene interactions for the stress response time-series

To test if the interactions between the transcription factors (TF) and the target genes (TG) are significant or negligible, for the TF-TG interactions available from [19], we computed the mean of the correlation (in absolute value) in all our time-series and we compared it with a null model created by 1000 random samples (obtained choosing a pool of random TG of equal cardinality as the real interactions for each TF). This test was repeated for the five datasets we are considering and it is possible to see in Fig. S2 (left column) that the mean of the correlations of real TF-TG is always comparable to that of the null model, i.e., that the TF-TG interactions are not significant for the datasets (p-values are shown in Table S1). Even when testing (for the Gasch database) every single transcription factor by itself, with its own null model, we found that out of 153 TFs, 17 were considered significantly different from the random model and, among these, only 9 were more correlated than the null model when signs were added to the correlations.

As mentioned in the paper, the activity of the TF is probably modulated post-translationally in a decisive manner. The external stimulation activates (or inactivates), through suitable signaling mechanisms, the TFs, which then trigger the transcriptional response. Even if for us the activity of TF is an hidden variable, one could expect its effect to be indirectly observable on the TG which are co-regulated by the same TF. We therefore performed a second test of significance of the correlation of co-transcribed TG. For the 5 datasets, this is shown in the central column of Fig. S2 and of Table S1. It can be noticed that the correlation is similar or slightly higher than in the previous test and that the null model still unanimously rejects the hypothesis of significance of such values on each of the 5 datasets. When instead we repeat the test replacing co-transcription by the same TF with co-participation to the same PC, the results are totally different, as can be seen on the histograms on the right in Fig. S2 and on Table S1. In this case the correlation is always much more significant than in a random choice of partners of equal cardinality.

Table S1: Significance tests for the correlation for classes of known gene-gene interactions. The first column shows the mean of the correlations between TF and corresponding TG for the five datasets. The p-value is computed by means of a null model obtained choosing an equal number of random TG for each TF. The second and third columns show the results of a similar analysis (correlation of true interactions vs null model generated with 1000 samples) performed between co-regulated genes (i.e. genes transcribed by the same TF), and between genes whose products belong to the same protein complex.

Dataset	mean(TF-TG)	p-val	mean(co-TG)	p-val	mean(PC)	p-val
Causton	0.2782	0.3750	0.286	0.9777	0.491	$10^{-71}$
Gasch	0.2442	0.4941	0.274	0.7343	0.473	$10^{-86}$
Ronen	0.2822	0.7611	0.291	0.4793	0.484	$10^{-63}$
Tirosh	0.3253	0.7829	0.326	0.0789	0.532	$10^{-52}$
Yoshi	0.3396	0.7597	0.365	0.8650	0.547	$10^{-45}$

## 4 Closed loop model with delay

Adding a delay  $\tau$  on the protein synthesis, the model (1) of the paper becomes

$$\begin{aligned} \frac{dm_i(t)}{dt} &= -\delta_i(m_i(t) - 1) - a_i(p_i(t) - 1) + b_i u \\ \frac{dp_i(t)}{dt} &= -\lambda_i(p_i(t) - 1) + r(m_i(t - \tau) - 1), \end{aligned} \quad (\text{S9})$$

and similarly for the model (2) of the paper. The fitting to the gene/protein data of [8] is shown in Table S2. In this case the ratio  $\delta_i/\lambda_i$  is even higher than in the model without delay.

## 5 Closed loop model with integral feedback from the growth rate

In [6], it is shown that in response to a heat shock the growth rate decreases with a kinetic time constant which is slower than that of the transcriptional transient. In line with the main idea developed in the paper, one could

Table S2: Gene and protein parameters estimated by fitting model (S9) to the time series of [8].

ORF	$a_i$	$b_i$	$\delta_i$	$\lambda_i$	$r$
YBR001C	0.003	0.010	0.035	0.004	0.027
YBR196C	0.029	0.047	0.078	0.002	0.014
YDR074W	0.008	0.007	0.046	0.001	0.044
YDR261C	0.005	0.014	0.053	0.001	0.022
YER003C	0.021	0.019	0.015	0.001	0.009
YGL253W	0.010	0.014	0.023	0.001	0.011
YHR163W	0.058	0.127	0.408	0.001	0.018
YKL127W	0.010	0.009	0.011	0.001	0.010
YNL241C	0.005	0.007	0.008	0.010	0.024
mean	0.017	0.028	0.075	0.002	0.020

think of using the growth rate as the variable used for the negative integral feedback regulation of the mRNA levels, instead of the  $p_i$ . Call  $g(t)$  the relative change with respect to the nominal growth rate value of the non-stressed yeast culture, in a log scale. If the yeast is in the exponential growth phase, then  $g(t)$  is a constant before the stimulation, and decreases after it with a slow transient. By properly choosing the feedback gains (including their sign), the system (1) or (2) of the paper can be replaced by

$$\frac{dm_i}{dt} = -\delta_i(m_i - 1) - a_i(g - 1) + b_i u$$

The correction given by  $-a_i(g - 1)$  can in principle be enough to explain the rapid transient excursion. In order to “close the loop”, a suitable differential equation for  $g(t)$  should be added. However, too little is known on how the growth rate is dynamically modified. In [20] it is affirmed that the current experimental techniques do not have a sufficiently high temporal resolution to draw experimental conclusions on how the growth rate changes with the mRNA concentrations. In between mRNA expression and growth rate changes, all considerations about protein synthesis made in the paper remain valid and should be included.

## 6 Estimating the ribosome density rate constant

In (5), an upper bound on the rate constant  $r$  (representing ribosomal density) can be obtained when consistency conditions (i.e.,  $p_i$  is a concentration hence  $p_i(t) \geq 0$  for all  $t$ ) are imposed on the data. Consider the time series of gene expression and compute (by numerical integration) the area under each expression profile using a formula like (3). For a gene experiencing no perturbation  $m_i(t) = \bar{m}_i = 1$ , hence  $p_i(t) = \bar{p}_i = 1$  for all  $t$ . For a downregulated gene  $m_i(t) < \bar{m}_i$ , implying  $p_i(t) - \bar{p}_i < 0$ . As  $p_i(t)$  is a relative concentration, we require  $p_i(t) \geq 0$  for all  $t$ . This condition gives an upper bound for the value of  $r$ . Call  $\mu = \max \left| \int_0^t (m_i(\tau) - 1) d\tau \right|$  for all genes for which  $\int_0^t (m_i(\tau) - 1) d\tau < 0$  (downregulated). Then  $p_i(t) \geq 0$  is satisfied provided  $1 - r\mu \geq 0$  i.e.,  $r < 1/\mu$ . For all time series considered, a choice of  $r = 0.01$  (motivated by the experimental data rather than by the dynamical model chosen) is sufficient to have biologically consistent values of  $\tilde{p}_i$  for the range of  $a_i, b_i$  required by the fitting procedure. Notice that the fitting of Table 1 of the paper (in which  $r$  is treated as a gene-specific parameter) confirms the order of magnitude chosen here.

## References

- [1] Audrey P Gasch, Paul T Spellman, Camilla M Kao, Orna Carmel-Harel, Michael B Eisen, Gisela Storz, David Botstein, and Patrick O Brown. Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell*, 11(12):4241–4257, 2000.
- [2] H Yoshimoto, K Saltsman, A P Gasch, H X Li, N Ogawa, D Botstein, P O Brown, and M S Cyert. Genome-wide analysis of gene expression regulated by the calcineurin/crz1p signaling pathway in *saccharomyces cerevisiae*. *J Biol Chem*, 277(34):31079–31088, 2002.
- [3] H C Causton, B Ren, S S Koh, C T Harbison, E Kanin, E G Jennings, T I Lee, H L True, E S Lander, and R A Young. Remodeling of yeast genome expression in response to environmental changes. *Mol Biol Cell*, 12(2):323–337, 2001.
- [4] I Tirosh, A Weinberger, M Carmi, and N Barkai. A genetic signature of interspecies variations in gene expression. *Nat. Genet.*, 38(7):830–834, 2006.



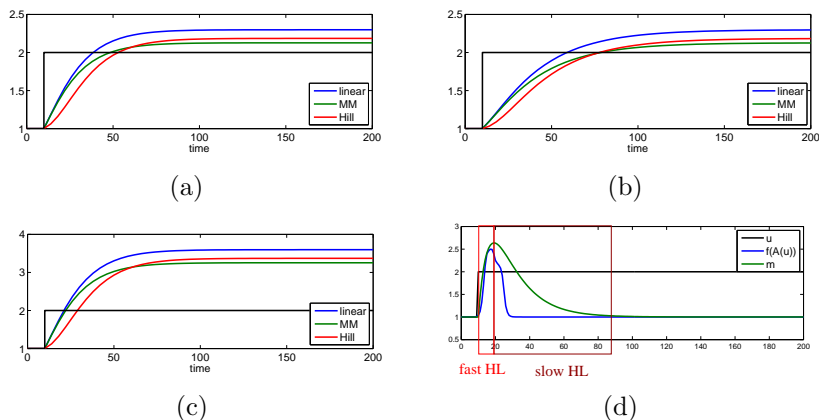


Figure S1: **Response of a single mode with zero order synthesis term.** Assuming a model like (S1) with as synthesis term any of (S2), (S3), (S5) and (S8), activated at  $t = 0$ , then a typical response is shown in (a) for a degradation constant equal to 10 min (i.e.,  $\delta_i = \ln(2)/10$ ) and in (b) for  $\delta_i = \ln(2)/30$ . In these plots, if the stimulus is the black step, the 3 colored lines representing the responses do not return to the basal level of 1. If adaptation can be achieved modulating the functional form of  $f_i$  (e.g. choosing  $f_i$  that vanishes a certain time interval after the application of the stimulus), what cannot be modified in a model like (S1) is the slope of the transient rising front, which depends on  $\delta_i$ . For example for  $\delta_i = \ln(2)/10$  (fast turnover gene) the rising front exhausts at  $t \sim 40 - 50$  min, see (a), which is in disagreement with the time series considered. Increasing the magnitude of the synthesis rate only yields an amplified response of equal slope, see (c). The situation is of course worse for  $\delta_i = \ln(2)/30$  see (b). Hence the need to resort to a time-varying  $\delta_i$  discussed e.g. in [15] requiring however an extremely fast turnover during the rising front of the transient, in conjunction with a vanishing functional form for  $f(w(u))$ , as shown for example in (d).

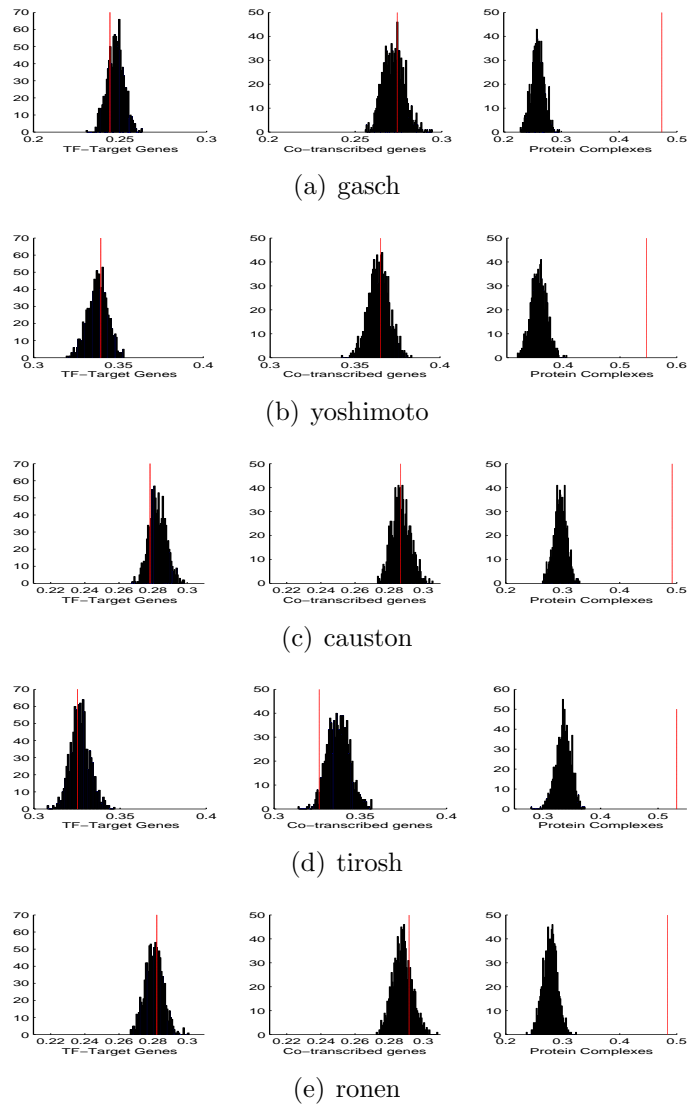


Figure S2: **Significance of gene-gene interactions.** Left column: for the 5 datasets the mean over all TF of the correlations (taken in absolute value) between each TF and its TG (red vertical line) is compared with the corresponding values obtained with 1000 random choices of TG for each TF (shown in the black histograms). The same test is repeated for the TG co-transcribed by the same TF and for the subunits of a PC. Only the PC are significantly more correlated than expected by the null model, while gene-gene interactions such as TF-TG and co-transcribed TG are not significant.

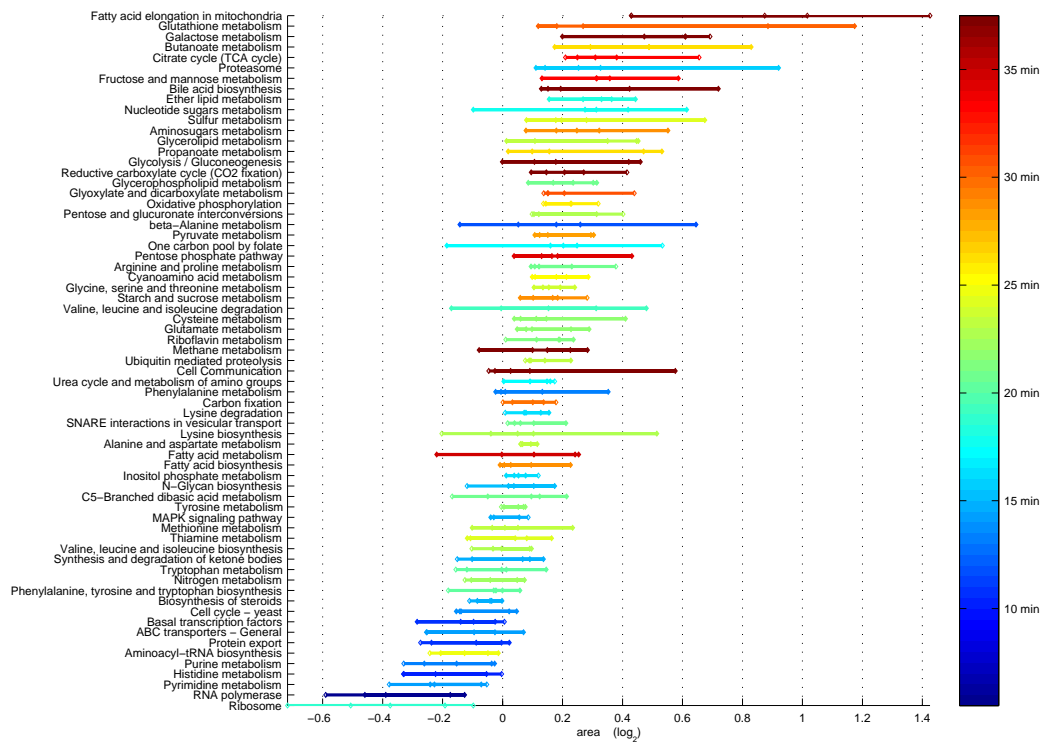


Figure S3: **Areas of the KEGG pathways.** For the 5 time series of [1], the average areas of the KEGG pathways are shown. Up/down regulated functional categories are very similar to Fig. 4 of the paper.

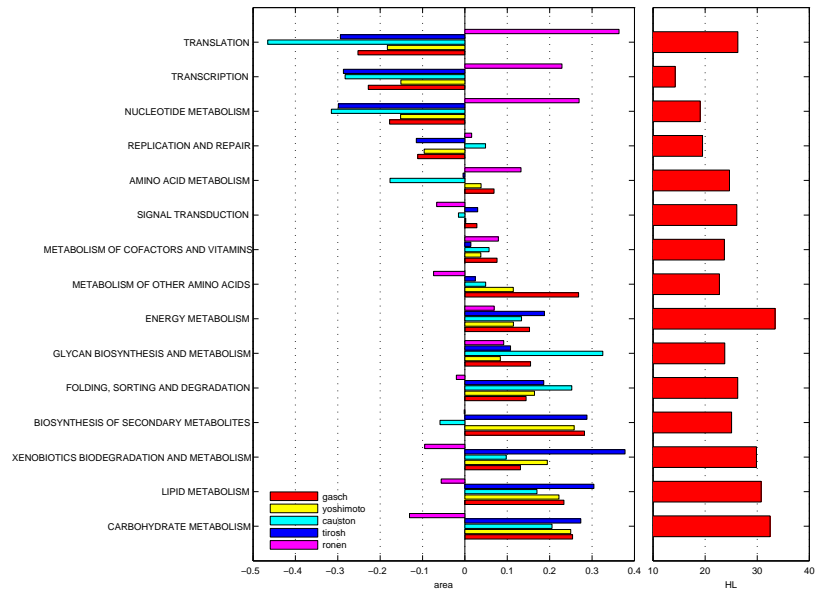


Figure S4: **Areas of the KEGG macrocategories.** For the 5 sets of time series mentioned in Fig. 6, the average areas of the KEGG pathways of Fig. S3 are lumped into 15 macrocategories, following KEGG hierarchy. Neatly downregulated categories in correspondence of a stress response (“transcription”, “translation” and “nucleotide metabolism”) are neatly up-regulated in presence of reciprocal stimuli such as “ronen” [5]. On the right the corresponding HL are gradually growing (notice the marked outlier in correspondence of “translation”).

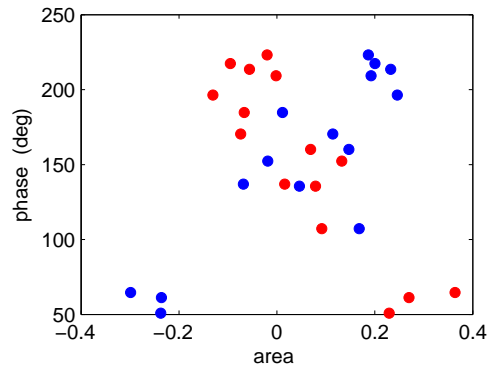


Figure S5: **Comparison with the “yeast metabolic cycle”** of [21]. In this time series,  $\sim 2000$  genes show a cyclic pattern with constant periodicity and different phase delays. See [22] for details. The area of the 15 macro-categories of Fig. S4 is compared with the phase of the same genes along this cycle. There is a clear correlation between area and phase for the stress responses (in blue) and a clear anticorrelation between area and phase for the glucose stimulations of [5] (in red). In [22] this is interpreted as the unfolding of a common gene expression program.

- [5] Michal Ronen and David Botstein. Transcriptional response of steady-state yeast cultures to transient perturbations in carbon source. *Proc. Natl. Acad. Sci. U.S.A.*, 103(2):389–394, 2006.
- [6] S Levy, J Ihmels, M Carmi, A Weinberger, G Friedlander, and N Barkai. Strategy of transcription regulation in the budding yeast. *PLoS ONE*, 2(2), 2007.
- [7] O Shalem, O Dahan, M Levo, M R Martinez, I Furman, E Segal, and Y Pilpel. Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation. *Mol Syst Biol*, 4:223–223, 2008.
- [8] Gal Chechik, Eugene Oh, Oliver Rando, Jonathan Weissman, Aviv Regev, and Daphne Koller. Activity motifs reveal principles of timing in transcriptional control of the yeast metabolic network. *Nat. Biotechnol.*, 26(11):1251–1259, 2008.
- [9] Jörg Grigull, Sanie Mnaimneh, Jeffrey Pootoolal, Mark D. Robinson, and Timothy R. Hughes. Genome-wide analysis of mRNA stability using transcription inhibitors and microarrays reveals posttranscriptional control of ribosome biogenesis factors. *Mol. Cell. Biol.*, 24(12):5534–5547, 2004.

- [10] Letian Kuai, Biswadip Das, and Fred Sherman. A nuclear degradation pathway controls the abundance of normal mRNAs in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A.*, 102(39):13962–13967, 2005.
- [11] Yulei Wang, Chih Long Liu, John D. Storey, Robert J. Tibshirani, Daniel Herschlag, and Patrick O. Brown. Precision and functional specificity in mRNA decay. *Proc. Natl. Acad. Sci. U.S.A.*, 99(9):5860–5865, 2002.
- [12] Juan Mata, Samuel Marguerat, and Jürg Bähler. Post-transcriptional control of gene expression: a genome-wide perspective. *Trends Biochem. Sci.*, 30(9):506–514, 2005.
- [13] J I Hargrove and F H Schmidt. The role of mRNA and protein stability in gene expression. *FASEB J.*, 3(12):2360–2370, 1989.
- [14] U. Alon. *An Introduction to Systems Biology - Design Principles of Biological Circuits*. Chapman & Hall/CRC;, 2006.
- [15] Paula M. Alepuz José E. Pérez-Ortín and Joaquín Moreno. Genomics and gene transcription kinetics in yeast. *Trends in Genetics*, 23(5):250–257, 2007.
- [16] M Ronen, R Rosenberg, B I Shraiman, and U Alon. Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics. *Proc Natl Acad Sci U S A*, 99(16):10555–10560, 2002.
- [17] Barrett C. Foat, S. Sean Houshmandi, Wendy M. Olivas, and Harmen J. Bussemaker. Profiling condition-specific, genome-wide regulation of mRNA stability in yeast. *Proc. Natl. Acad. Sci. U.S.A.*, 102(49):17675–17680, 2005.
- [18] G. Chechik and D. Koller. Timing of gene expression responses to environmental changes. *Journal of Computational Biology*, 16(2):279–290, February 2009.
- [19] S Balaji, M. Madan Babu, Lakshminarayan M Iyer, Nicholas M Luscombe, and L Aravind. Comprehensive analysis of combinatorial regulation using the transcriptional regulatory network of yeast. *J. Mol. Biol.*, 360(1):213–227, 2006.
- [20] E M Airoldi, C Huttenhower, D Gresham, C Lu, A A Caudy, M J Dunham, J R Broach, D Botstein, and O G Troyanskaya. Predicting cellular growth from gene expression signatures. *PLoS Comput Biol*, 5(1), Jan 2009.
- [21] Benjamin P Tu, Andrzej Kudlicki, Maga Rowicka, and Steven L McKnight. Logic of the yeast metabolic cycle: Temporal compartmentalization of cellular processes. *Science*, 310(5751):1152–1158, 2005.

- [22] Nicola Soranzo, Mattia Zampieri, Lorenzo Farina, and Claudio Altafini. mRNA stability and the unfolding of gene expression in the long-period yeast metabolic cycle. *BMC Systems Biology*, 3:18, 2009.