Supplementary Material

Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast

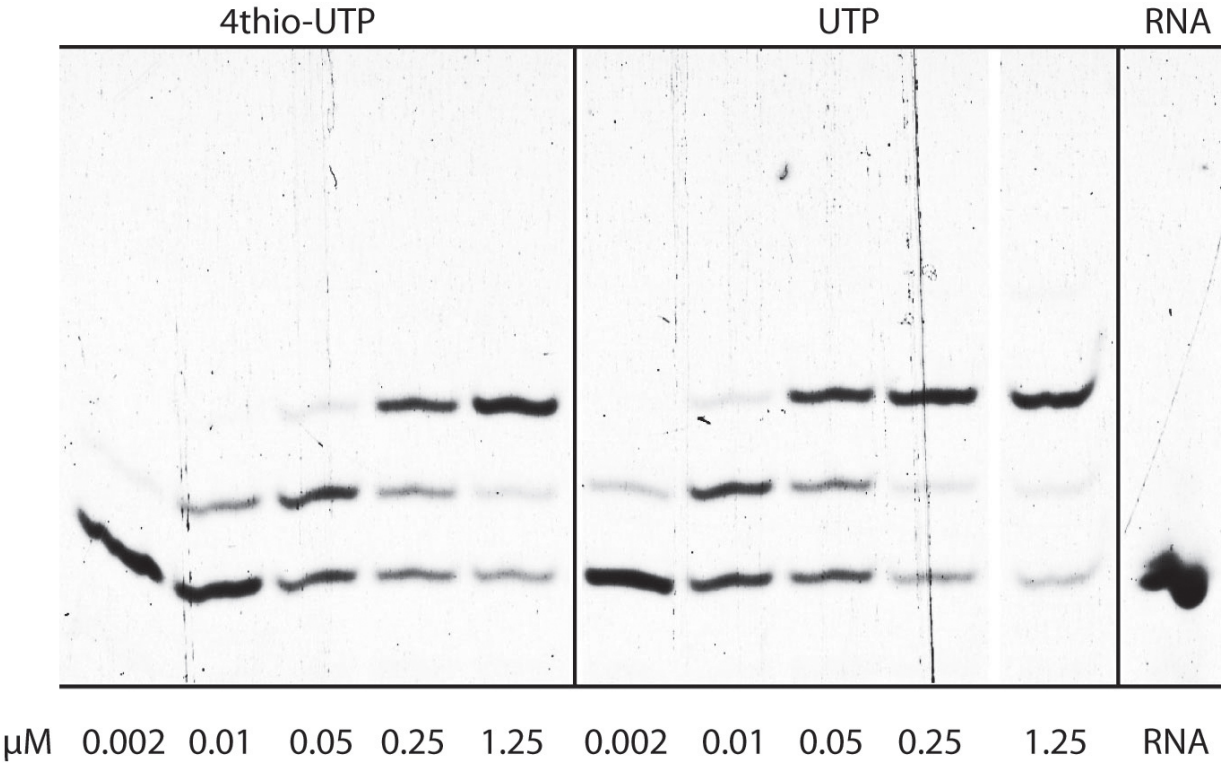
November 12, 2010

Christian Miller, Björn Schwalb, Kerstin Maier, Daniel Schulz, Sebastian Dümcke, Benedikt Zacher, Andreas Mayer, Jasmin Sydow, Lisa Marcinowski, Lars Dölken, Dietmar E. Martin, Achim Tresch, and Patrick Cramer.

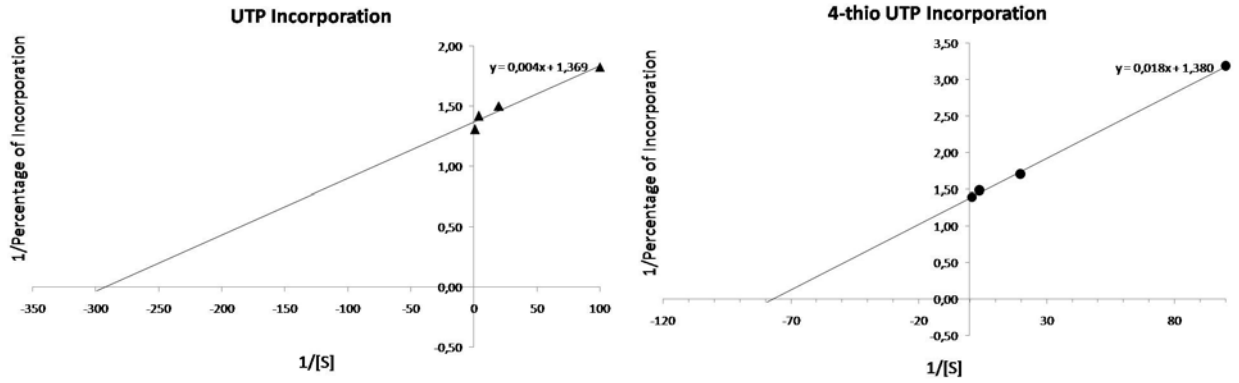
Contents

Supplementary figures

1 Supplementary Figure S1



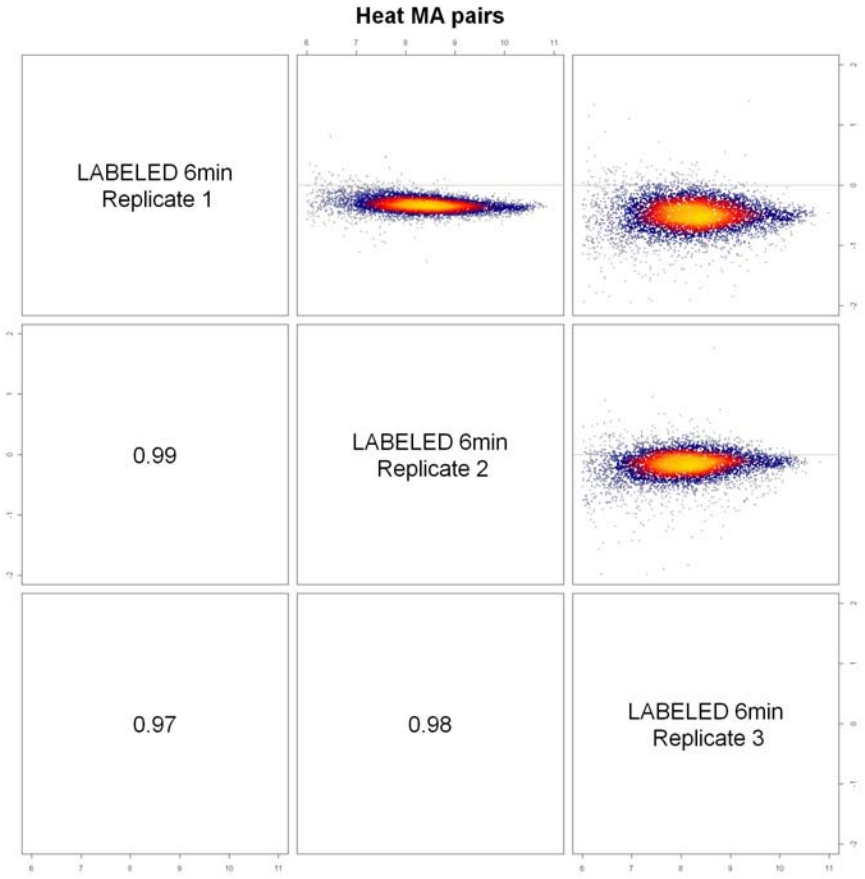
(a)



(b)

(a) Results of the in-vitro RNA extension assay to determine the nucleotide incorporation efficiency by RNA polymerase II. (b) Lineweaver-Burk diagrams (left: UTP, right: 4sUTP) showing similar kinetics. The maximal reaction rate v_{max} is virtually unchanged ($v_{max} \approx 0.72$), the Michaelis-Menten constant K_M increased from 3 nM for UTP to 13 nM for 4sUTP.

2 Supplementary Figure S2



Comparison of the labeled mRNA fractions in all three replicate measurements for the wild type samples after 6 min. labeling time (Wild type experiment (see section 12.1)). Sample cultures of replicate 1 and 2 were grown on the same day. Upper triangle: MA plots of the log-intensities (M: intensity ratio, versus A: average intensity). Lower triangle: respective pairwise Spearman correlations. Plots were produced with the R package LSD [Schwalb *et al.*, 2010].

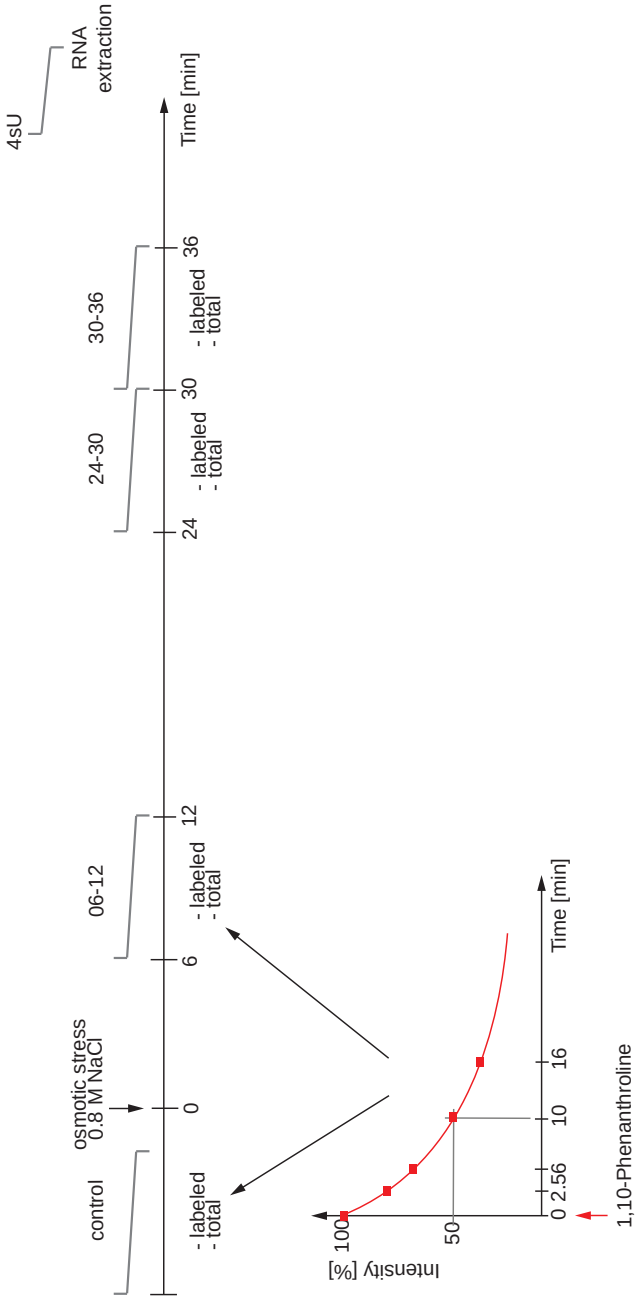
	L3 R1	L3 R2	L6 R1	L6 R3	L6 R3	L12 R1	L12 R3	L12 R3	L24 R1	L24 R3	L24 R3
L3 R1	1	0.99	0.98	0.98	0.96	0.92	0.92	0.92	0.89	0.89	0.89
L3 R2	0.99	1	0.97	0.97	0.94	0.9	0.91	0.91	0.87	0.87	0.87
L6 R1	0.98	0.97	1	0.99	0.97	0.95	0.96	0.95	0.94	0.93	0.93
L6 R3	0.98	0.97	0.99	1	0.98	0.96	0.96	0.96	0.94	0.94	0.94
L6 R3	0.96	0.94	0.97	0.98	1	0.98	0.98	0.98	0.95	0.95	0.96
L12 R1	0.92	0.9	0.95	0.96	0.98	1	0.99	0.99	0.99	0.99	0.99
L12 R3	0.92	0.91	0.96	0.96	0.98	0.99	1	1	0.98	0.99	0.99
L12 R3	0.92	0.91	0.95	0.96	0.98	0.99	1	1	0.98	0.98	0.99
L24 R1	0.89	0.87	0.94	0.94	0.95	0.99	0.98	0.98	1	1	1
L24 R3	0.89	0.87	0.93	0.94	0.95	0.99	0.99	0.98	1	1	1
L24 R3	0.89	0.87	0.93	0.94	0.96	0.99	0.99	0.99	1	1	1

Assessment of reproducibility for all measurements of labeled mRNA in the Wild type experiment (see section 12.1). The table shows the pairwise Spearman correlations between unnormalized mRNA abundances. (Abbreviation: mRNA fraction followed by labeling time and replicate number.)

	T0 R1	T0 R2	T0 R3	T0 R4	T3 R1	T3 R2	T6 R1	T6 R3	T6 R3	T12 R1	T12 R3	T12 R3	T24 R1	T24 R3	T24 R3
T0 R1	1	0.99	0.99	0.99	1	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.98	0.98
T0 R2	0.99	1	0.99	0.99	0.99	1	0.99	0.99	0.98	0.99	0.98	0.98	0.97	0.97	0.97
T0 R3	0.99	0.99	1	1	0.99	0.98	0.99	0.98	1	1	0.99	0.99	0.98	0.98	0.98
T0 R4	0.99	0.99	1	1	0.99	0.98	0.99	0.98	1	0.99	0.99	0.99	0.98	0.98	0.98
T3 R1	1	0.99	0.99	0.99	1	0.99	1	0.99	0.99	0.99	0.99	0.99	0.98	0.98	0.98
T3 R2	0.99	1	0.98	0.98	0.99	1	0.99	1	0.98	0.98	0.99	0.98	0.97	0.97	0.97
T6 R1	0.99	0.99	0.99	0.99	1	0.99	1	0.99	0.99	0.99	0.99	0.99	0.98	0.98	0.98
T6 R3	0.99	0.99	0.98	0.98	0.99	1	0.99	1	0.98	0.98	0.98	0.98	0.97	0.97	0.97
T6 R3	0.99	0.98	1	1	0.99	0.98	0.99	0.98	1	1	0.99	0.99	0.99	0.99	0.99
T12 R1	0.99	0.99	1	0.99	0.99	0.98	0.99	0.98	1	1	1	1	0.99	0.99	0.99
T12 R3	0.99	0.98	0.99	0.99	0.99	0.99	0.99	0.98	0.99	1	1	0.99	0.99	0.99	0.99
T12 R3	0.99	0.98	0.99	0.99	0.99	0.98	0.99	0.98	0.99	1	0.99	1	0.98	0.99	0.99
T24 R1	0.98	0.97	0.98	0.98	0.98	0.97	0.98	0.97	0.99	0.99	0.98	1	1	0.99	1
T24 R3	0.98	0.97	0.98	0.98	0.98	0.97	0.98	0.97	0.99	0.99	0.99	0.99	0.99	1	0.99
T24 R3	0.98	0.97	0.98	0.98	0.98	0.97	0.98	0.97	0.99	0.99	0.99	1	0.99	0.99	1

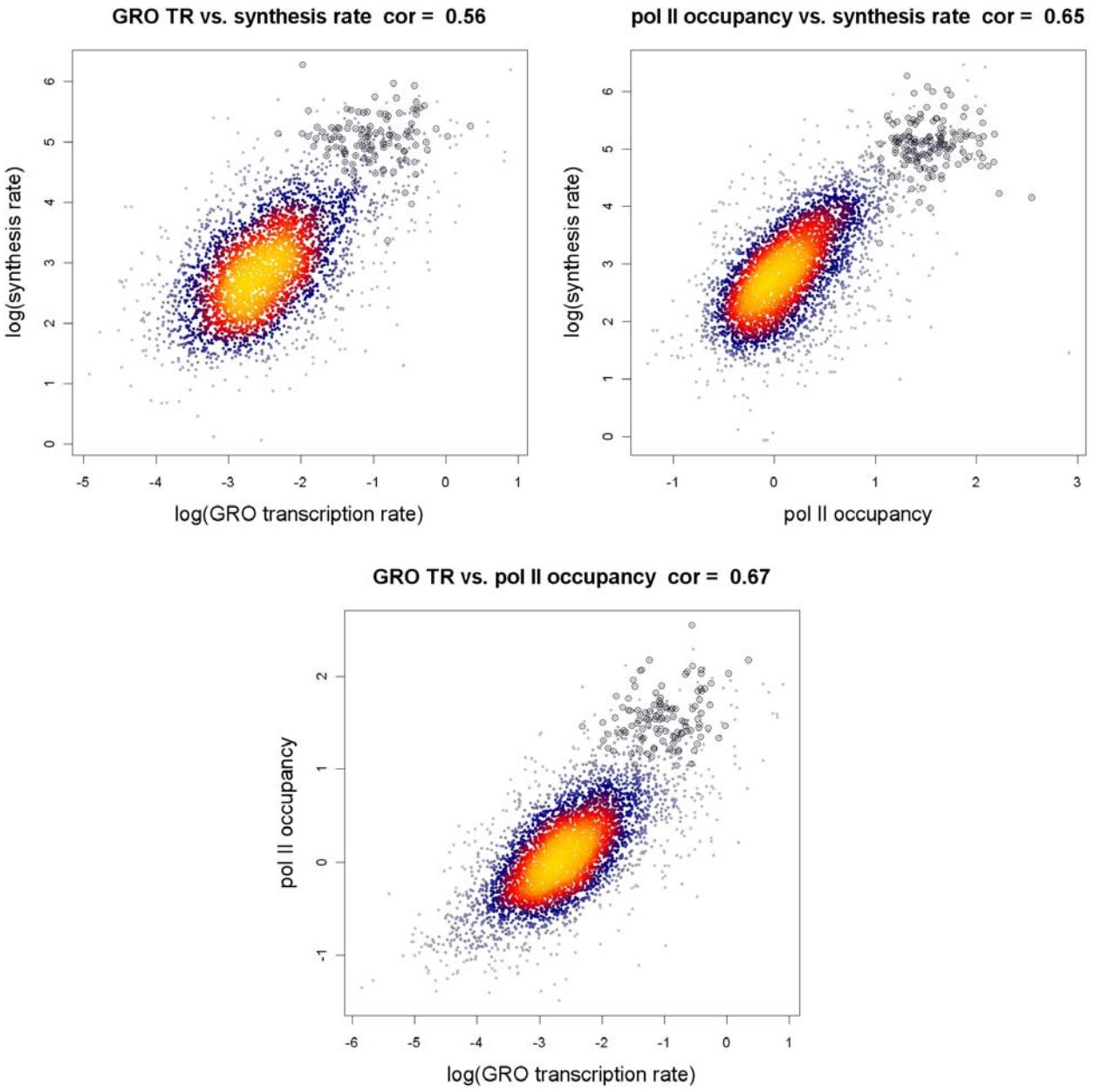
Assessment of reproducibility for all measurements of total mRNA in the Wild type experiment (see section 12.1). The table shows the pairwise Spearman correlations between unnormalized mRNA abundances. (Abbreviation: mRNA fraction followed by labeling time and replicate number.)

3 Supplementary Figure S3



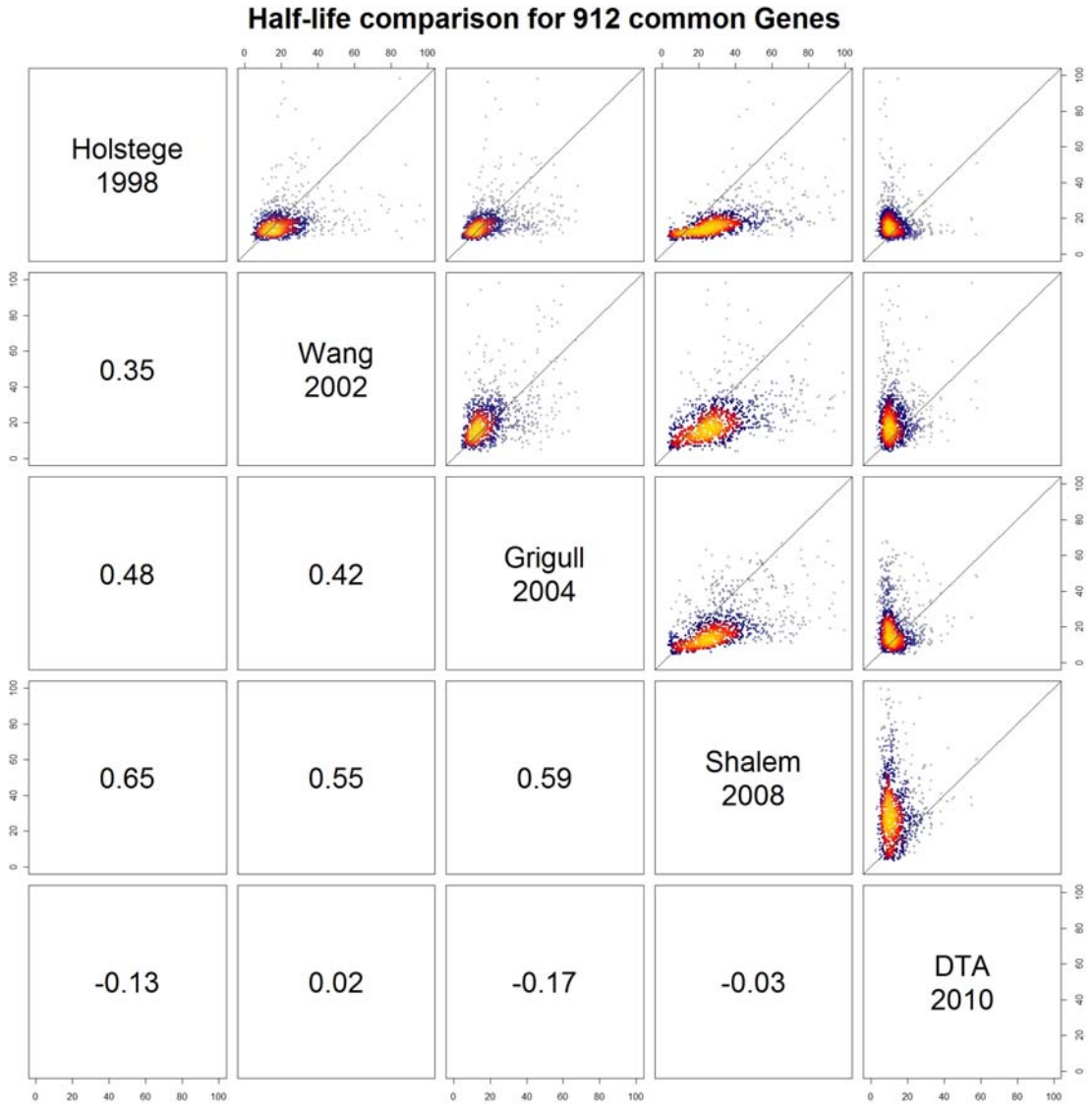
Design of the RTqPCR experiments. Extracts of total and labeled mRNA were taken (after a labeling period of 6 min) of wild type samples and during osmotic stress at t = 12, 30 and 36 min. RTqPCR was performed for a set of selected genes. Moreover, a transcriptional shutoff experiment was performed adding 1,10-Phenanthroline for the wild type samples and after 12 min of osmotic stress. The mRNA decay rates of selected genes were determined with RTqPCR by an mRNA decay time series taken at t = 0, 2.5, 6, 10, 16 min after transcriptional shut off (see Supplementary methods, Section 15).

4 Supplementary Figure S4



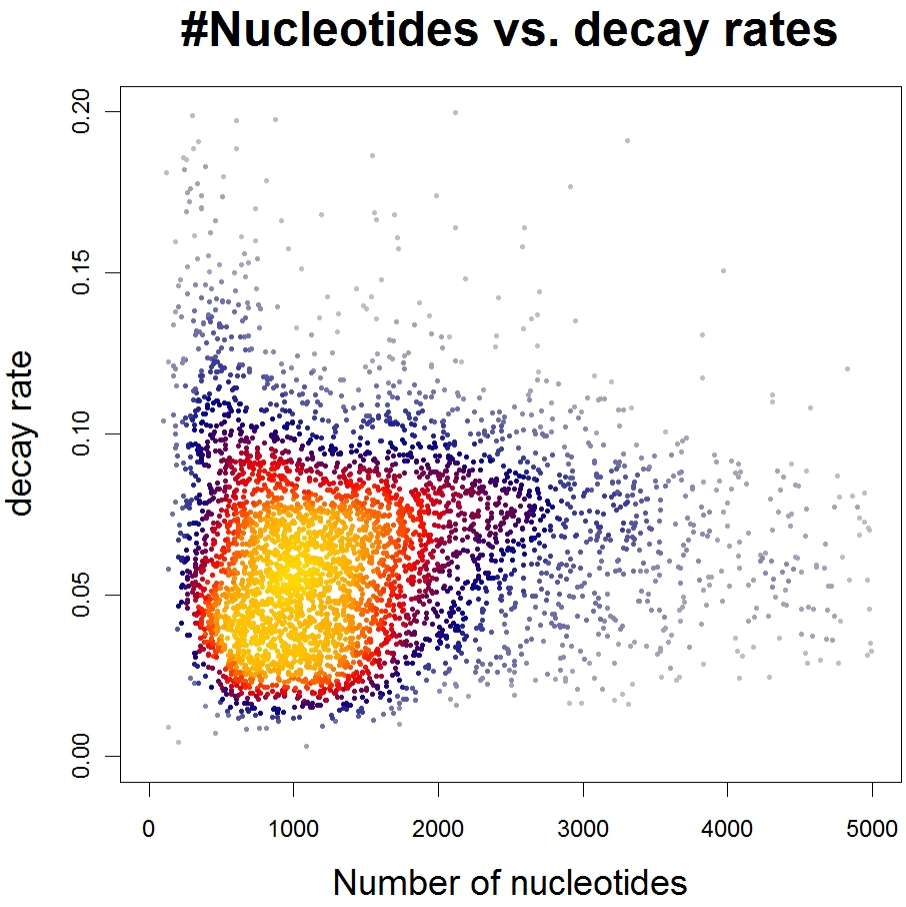
Comparison of synthesis rates determined by DTA (Dynamic Transcriptome Analysis, see Supplementary methods, part I) with two alternative measures of transcriptional activity: Pol II occupancy (obtained by ChIP-chip experiments) [Mayer *et al.*, 2010] and transcription rates as measured with the genomic run-on (GRO) method [Pelechano & Pérez-Ortín, 2010]. Top left: Scatterplot of GRO transcription rates vs. DTA synthesis rates. Top right: Scatterplot of Pol II occupancy vs. DTA synthesis rates. Bottom: GRO rates vs. Pol II occupancy. Highlighted are the ribosomal protein genes. The Pol II occupancy shown is the genewise mean from start to stop codon, processed by means of the Bioconductor Starr package [Zacher *et al.*, 2010]. Plots were produced with the R package LSD [Schwalb *et al.*, 2010].

5 Supplementary Figure S5



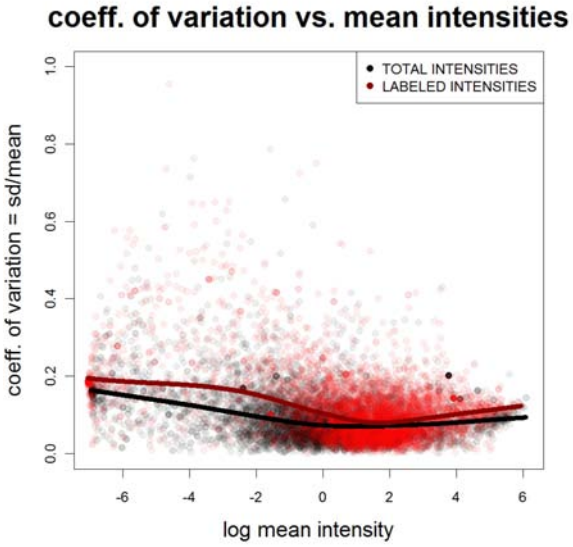
Scatterplots comparing DTA (Dynamic Transcriptome Analysis, see Supplementary methods, part I) half-life estimates with literature results obtained by experiments using transcriptional arrest [Holstege *et al.*, 1998, Wang *et al.*, 2002, Grigull *et al.*, 2004, Shalem *et al.*, 2008]. All four data sets were obtained with a Yeast strain containing the RNA polymerase II temperature sensitive mutant rpb1-1. Decay rates can be measured after blocking transcription, but this requires a perturbing heat shock (cells have to be shifted to the non-permissive temperature of 37°C). The intensity of each mRNA species relative to that observed in a wild type cell gives their measure for mRNA decay, applying the usual first-order exponential decay model. Three experiments were conducted using the same Yeast strain. Their median/mean mRNA half-lives are reported as 16/19 min [Holstege *et al.*, 1998], 20/23 min [Wang *et al.*, 2002], 18/22 min [Grigull *et al.*, 2004], and 30/34 min [Shalem *et al.*, 2008]. The lower panel shows the respective Spearman correlations. Generally, the estimates show poor agreement. DTA does not correlate with any of the invasive methods, though all show a typical right-skewed distribution (see main text Figure 2B).

6 Supplementary Figure S6

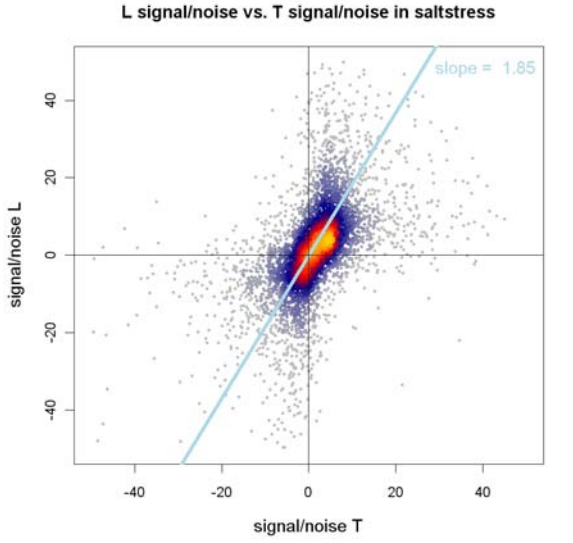


The comparison of transcript length and DTA (Dynamic Transcriptome Analysis, see Supplementary methods, part I) decay rates (estimated with DTA) shows that degradation speed (= decay rate = log(2)/half-life) is uncorrelated with transcript length. The spearman correlation coefficient is 0.06. It is noteworthy that a correlation coefficient of 0.64 is obtained, if discrepancies that are due to 4sU/Biotin labeling are ignored (Supplementary methods 13.2). Without bias removal, the half-lives of 72% of the mRNAs are artificially elongated by a factor of at least 2, so that the overall ranking of the half-lives is strongly altered.

7 Supplementary Figure S7

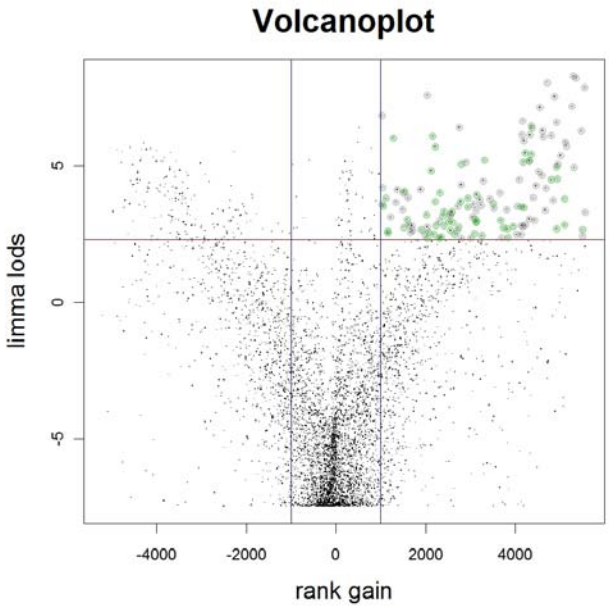


The coefficient of variation of the total resp. labeled measurements is plotted as a function of the resp. mean value. Three replicate measurements of the total (black) and the labeled intensities (red) of wild type yeast after a 6 min labeling period were used to calculate a standard deviation (sd), a mean intensity, and a coefficient of variation (cv=sd/mean) for each gene. The solid lines are smoothed estimates of the cv intensities obtained by loess regression.



Sensitivity comparison of the standard transcriptomics measurements (total mRNA, T) to DTA (labeled mRNA, L). For a precise explanation of the mRNA fractions, see Supplementary methods, section 12. To compare transcriptional changes in both methods, we calculated "signal-to-noise" ratios after 18 min of osmotic stress relative to the wild type. One for each gene and each mRNA fraction. The signal-to-noise ratios were defined as the quotient of the fold of the intensity after 18 min of stress over wild type (as estimated by standard transcriptomics resp. DTA) and the standard deviation obtained by the analysis above (see upper figure). The blue line indicates that on average, DTA is 1.85 times more sensitive than the classical method.

8 Supplementary Figure S8



Volcanoplot for the comparison of the synthesis rates 36 min after osmotic stress induction against wild type synthesis rates. The x-axis shows the difference of the ranks of a gene in the 36 min synthesis rates distribution and the wild type synthesis rates distribution. The y-axis shows the significance of a change in synthesis rates, as measured with limma [Smyth, 2004]. It is given as the log odds (synthesis rate is different/synthesis rate is unchanged) for each gene. Grey dots: Hog1 and/or Msn2/4 dependent osmotic stress genes identified by [Capaldi *et al.*, 2008]. The 58 dots in green are novel genes also clearly involved in the transcriptional response to osmotic stress.

YPR098C	DAN3	THO1	PAU2	YMR034C
YLR031W	TGL2	SRL3	YNR068C	YLR108C
SGF11	YNL040W	ECI1	PRM8	UBC5
ARR2	AFR1	YIL055C	YIL046W-A	FMP23
GIP1	ECM12	YLR285C-A	YDL085C-A	YBR056W-A
ATH1	YNL211C	STF1	CPS1	YKL133C
SPL2	PET10	YET1	GSP2	FSH1
DIA1	PCL1	SPG5	YPR172W	SCS22
YER185W	RNR3	YJL185C	HMX1	REC102
ICT1	YGL010W	MAG1	STB2	PEP12
UGX2	YOL024W	MST27	LEE1	
PFK26	GSM1	YNL130C-A	BOP2	

List of the 58 genes depicted in green in the figure above, which are involved in the transcriptional response to osmotic stress.

9 Supplementary Table T1

	profile name	fraction	time window	time point	osmotic stress	bio rep	day	exp series
1	C 00-00 - WT 1 1 WT.CEL	C	00-00	0	-	1	1	WT
2	C 00-00 - WT 2 1 WT.CEL	C	00-00	0	-	2	1	WT
3	L 00-03 - WT 1 1 WT.CEL	L	00-03	3	-	1	1	WT
4	L 00-03 - WT 2 1 WT.CEL	L	00-03	3	-	2	1	WT
5	L 00-06 - WT 1 1 WT.CEL	L	00-06	6	-	1	1	WT
6	L 00-06 - WT 2 1 WT.CEL	L	00-06	6	-	2	1	WT
7	L 00-06 - WT 3 2 WT.CEL	L	00-06	6	-	3	2	WT
8	L 00-12 - WT 1 2 WT.CEL	L	00-12	12	-	1	2	WT
9	L 00-12 - WT 2 2 WT.CEL	L	00-12	12	-	2	2	WT
10	L 00-12 - WT 3 2 WT.CEL	L	00-12	12	-	3	2	WT
11	L 00-24 - WT 1 2 WT.CEL	L	00-24	24	-	1	2	WT
12	L 00-24 - WT 2 2 WT.CEL	L	00-24	24	-	2	2	WT
13	L 00-24 - WT 3 2 WT.CEL	L	00-24	24	-	3	2	WT
14	T 00-00 - WT 1 1 WT.CEL	T	00-00	0	-	1	1	WT
15	T 00-00 - WT 2 1 WT.CEL	T	00-00	0	-	2	1	WT
16	T 00-00 - WT 3 1 WT.CEL	T	00-00	0	-	3	1	WT
17	T 00-00 - WT 4 1 WT.CEL	T	00-00	0	-	4	1	WT
18	T 00-03 - WT 1 1 WT.CEL	T	00-03	3	-	1	1	WT
19	T 00-03 - WT 2 1 WT.CEL	T	00-03	3	-	2	1	WT
20	T 00-06 - WT 1 1 WT.CEL	T	00-06	6	-	1	1	WT
21	T 00-06 - WT 2 1 WT.CEL	T	00-06	6	-	2	1	WT
22	T 00-06 - WT 3 2 WT.CEL	T	00-06	6	-	3	2	WT
23	T 00-12 - WT 1 2 WT.CEL	T	00-12	12	-	1	2	WT
24	T 00-12 - WT 2 2 WT.CEL	T	00-12	12	-	2	2	WT
25	T 00-12 - WT 3 2 WT.CEL	T	00-12	12	-	3	2	WT
26	T 00-24 - WT 1 2 WT.CEL	T	00-24	24	-	1	2	WT
27	T 00-24 - WT 2 2 WT.CEL	T	00-24	24	-	2	2	WT
28	T 00-24 - WT 3 2 WT.CEL	T	00-24	24	-	3	2	WT
29	U 00-03 - WT 1 1 WT.CEL	U	00-03	3	-	1	1	WT
30	U 00-03 - WT 2 1 WT.CEL	U	00-03	3	-	2	1	WT
31	U 00-06 - WT 1 1 WT.CEL	U	00-06	6	-	1	1	WT
32	U 00-06 - WT 2 1 WT.CEL	U	00-06	6	-	2	1	WT
33	U 00-06 - WT 3 2 WT.CEL	U	00-06	6	-	3	2	WT
34	U 00-12 - WT 1 2 WT.CEL	U	00-12	12	-	1	2	WT
35	U 00-12 - WT 2 2 WT.CEL	U	00-12	12	-	2	2	WT
36	U 00-12 - WT 3 2 WT.CEL	U	00-12	12	-	3	2	WT
37	U 00-24 - WT 1 2 WT.CEL	U	00-24	24	-	1	2	WT
38	U 00-24 - WT 2 2 WT.CEL	U	00-24	24	-	2	2	WT
39	U 00-24 - WT 3 2 WT.CEL	U	00-24	24	-	3	2	WT

Table of all produced gene expression arrays of the wild type experiment (see supplementary methods, section 12.1). Profile name column: File name as uploaded to Array Express (accession number E-MTAB-439). Fraction column: (C) cells lacking the transporter plasmid (control mRNA samples), (L) labeled mRNA, (T) total mRNA, (U) unlabeled mRNA. Time window column: start and end of the labeling period (4sU). Timepoint column: extraction timepoint of mRNA. Osmotic stress: (+) addition of sodium chloride to the sample culture to a concentration of 0.8 M, (-) no addition of sodium chloride. Biorep column: number of the biological replicate. Day column: Indicator of the experiment date. Sample cultures were grown on two different days (day 1, day 2). Expseries column: abbreviation for the experiment series (WT: Wild type experiment (see supplementary methods, section 12.1)).

10 Supplementary Table T2

	profile name	fraction	time window	time point	osmotic stress	bio rep	day	exp series
1	L 00-06 + WT 1 1 SS.CEL	L	00-06	6	+	1	1	SS
2	L 06-12 - WT 1 1 SS.CEL	L	06-12	6	-	1	1	SS
3	L 06-12 + WT 1 1 SS.CEL	L	06-12	6	+	1	1	SS
4	L 12-18 - WT 1 1 SS.CEL	L	12-18	6	-	1	1	SS
5	L 12-18 + WT 1 1 SS.CEL	L	12-18	6	+	1	1	SS
6	L 18-24 + WT 1 1 SS.CEL	L	18-24	6	+	1	1	SS
7	L 24-30 + WT 1 1 SS.CEL	L	24-30	6	+	1	1	SS
8	L 30-36 - WT 1 1 SS.CEL	L	30-36	6	-	1	1	SS
9	L 30-36 + WT 1 1 SS.CEL	L	30-36	6	+	1	1	SS
10	T 00-06 + WT 1 1 SS.CEL	T	00-06	6	+	1	1	SS
11	T 06-12 - WT 1 1 SS.CEL	T	06-12	6	-	1	1	SS
12	T 06-12 + WT 1 1 SS.CEL	T	06-12	6	+	1	1	SS
13	T 12-18 - WT 1 1 SS.CEL	T	12-18	6	-	1	1	SS
14	T 12-18 + WT 1 1 SS.CEL	T	12-18	6	+	1	1	SS
15	T 18-24 + WT 1 1 SS.CEL	T	18-24	6	+	1	1	SS
16	T 24-30 + WT 1 1 SS.CEL	T	24-30	6	+	1	1	SS
17	T 30-36 - WT 1 1 SS.CEL	T	30-36	6	-	1	1	SS
18	T 30-36 + WT 1 1 SS.CEL	T	30-36	6	+	1	1	SS
19	U 06-12 - WT 1 1 SS.CEL	U	06-12	6	-	1	1	SS
20	U 06-12 + WT 1 1 SS.CEL	U	06-12	6	+	1	1	SS
21	U 12-18 - WT 1 1 SS.CEL	U	12-18	6	-	1	1	SS
22	U 12-18 + WT 1 1 SS.CEL	U	12-18	6	+	1	1	SS
23	U 30-36 - WT 1 1 SS.CEL	U	30-36	6	-	1	1	SS
24	U 30-36 + WT 1 1 SS.CEL	U	30-36	6	+	1	1	SS

Table of all produced gene expression arrays of the wild type experiment (see supplementary methods, section 12.1). Profile name column: File name as uploaded to Array Express (accession number E-MTAB-439). Fraction column: (C) cells lacking the transporter plasmid (control mRNA samples), (L) labeled mRNA, (T) total mRNA, (U) unlabeled mRNA. Time window column: start and end of the labeling period (4sU). Timepoint column: extraction timepoint of mRNA. Osmotic stress: (+) addition of sodium chloride to the sample culture to a concentration of 0.8 M, (-) no addition of sodium chloride. Biorep column: number of the biological replicate. Day column: Indicator of the experiment date. Sample cultures were grown on two different days (day 1, day 2). Expseries column: abbreviation for the experiment series (SS: Osmotic stress experiment (see supplementary methods, section 12.2)).

11 Supplementary Table T3

	profile name	extracts	time point	osmotic stress	bio rep	day	exp series
1	IP 00 - 1 W.CEL	IP	0	-	1	1	WT
2	IP 00 - 2 W.CEL	IP	0	-	2	1	WT
3	Input 00 - 1 W.CEL	Input	0	-	1	1	WT
4	IP 12 + 1 S.CEL	IP	12	+	1	2	SS
5	IP 12 + 2 S.CEL	IP	12	+	2	2	SS
6	Input 12 + 1 S.CEL	Input	12	+	1	2	SS
7	IP 24 + 1 S.CEL	IP	24	+	1	2	SS
8	IP 24 + 2 S.CEL	IP	24	+	2	2	SS
9	Input 24 + 1 S.CEL	Input	24	+	1	2	SS

Table of all produced tiling arrays of the Pol II ChIP-chip experiment. Profile name column: File name as uploaded to Array Express (accession number E-MTAB-439). Extracts column: (IP) Chromatin immunoprecipitation (Input) Genomic Input. Timepoint column: timepoint of ChIP-chip after salt addition. Osmotic stress: (+) addition of sodium chloride to the sample culture to a concentration of 0.8 M. (-) no addition of sodium chloride. Biorep column: number of the biological replicate. Day column: Indicator of the experiment date. Sample cultures were grown on two different days (day 1, day 2). Expseries column: abbreviation for the experiment series (WT: Occupancy profiles for the Pol II subunit Rpb3 by ChIP-chip analysis for the wild type, see [Mayer *et al.*, 2010], SS: Occupancy profiles for the Pol II subunit Rpb3 by ChIP-chip analysis under osmotic stress conditions (see supplementary methods, section 16.1)).

Supplementary methods

Part I
Dynamic Transcriptome Analysis (DTA)

12 Preprocessing and Quality Control

12.1 Wild type experiment

We used S. cerevisiae strain BY4741 (MATa, his2Δ1, leu2Δ0, met15Δ0, ura3Δ0). The strain was transformed with plasmid YEpEBI311 carrying the human equilibrative nucleoside transporter hENT1. Samples for establishing DTA were grown in SD medium overnight. 4sU (Sigma) was added to the media to a final concentration of 500 µM at timepoint $t = 0$, and cells were harvested after different labeling times $t_r \in \{3, 6, 12, 24 \text{ min}\}$, where r denotes biological samples $r \in R = \{1, ..., 11\}$. The mRNA of each sample was split into three fractions (extracts): Total mRNA T_r (total cellular mRNA), labeled mRNA L_r (thiol-labeled newly transcribed mRNA), and unlabeled mRNA U_r (pre-existing mRNA) (see main text (Materials and methods)). All mRNA extracts were hybridized to GeneChip Yeast Genome 2.0 microarrays (Affymetrix). Let G be the set of genes that were measured on the array. The measured gene expression of gene $g \in G$ in a sample r in the fraction $F_r \in \mathcal{F}^{wildtype} = \{T_r, L_r, U_r \mid r \in R\}$ is denoted by F_{gr}. This notation emphasizes the fraction from which the mRNA was obtained. Additionally, we created quadruplicates of total mRNA T for timepoint $t = 0$. Duplicates of cells lacking the nucleoside transporter (hENT1) are termed control mRNA samples C. A listing of arrays can be found in (Supplementary Table T1).

12.2 Osmotic stress experiment

Sample cultures were obtained as above and divided in aliquots. 4sU was added at 0, 6, 12, 18, 24, and 30 min after the addition of sodium chloride to a concentration of 0.8 M. Cells were harvested after the labeling time of $t_L = 6$ min. Total mRNA T_i^{salt} and labeled mRNA L_i^{salt} was purified and analyzed to yield expression profiles for each time window $i \in I = \{\text{0-6, 6-12, 12-18, 18-24, 24-30, 30-36 min}\}$(see main text figure 3A). Unlabeled mRNA U_i^{salt} indeed, was only analyzed for $i \in \{\text{6-12, 12-18, 30-36 min}\}$. Control cultures were also produced to gain biological triplicate profiles of the wild type T_i^{wt},U_i^{wt} and L_i^{wt}, $i \in \{\text{6-12, 12-18, 30-36 min}\}$. By analogy to the notation in the wild type experiment (see Section 12.1), the measured gene expression of a set of genes G in a sample i in the fraction $F_i \in \mathcal{F}^{osmotic} = \{T_i, L_i, U_i \mid i \in I\}$ is denoted by F_{gi}, $g \in G$. All mRNA extracts were hybridized to GeneChip Yeast Genome 2.0 microarrays (Affymetrix). A listing of arrays can be found in (Supplementary Table T2).

12.3 Visual inspection, Quality metrics

Probe signal intensities were captured and processed with GeneChip Operating Software (Affymetrix), and the resulting CEL files were preprocessed using the GC Robust Multi-array Average (GCRMA) method [Wu *et al.*, 2004]. However we did not apply quantile normalization, which by default is applied as the last preprocessing step in GCRMA. This would have been inappropriate, because the expression distributions in the three fractions are expected to be substantially different.

We needed to develop a novel normalization strategy, that takes into account the particular way in which the sample fractions F_r were obtained (Section 13.1). To find a set of reliable genes, for which the parameter estimation is performed as described later (Section 13.2), the following heuristic was used: We excluded all genes, which are annotated as dubious or silenced by the Stanford Public Database (SGD™: Saccharomyces Genome Database) [Cherry *et al.*, 1998], and only kept those annotated as verified or uncharacterized. This leaves 5743 of initially 5976 genes. In a second step we excluded all 137 ribosomal protein genes G^{rpg} [Nakao *et al.*, 2004]. We noticed that these genes are generally expressed at a level which is considerably higher than that of the other genes. Therefore, ribosomal protein genes probably do not lie in the linear measurement range of the scanner and are likely to introduce a bias to our estimation procedure.

Visual inspection of pairwise scatterplots (Supplementary Figure S2) showed no systematic differences other than variation in noise regarding the day on which sample cultures were grown, and slight variations in the

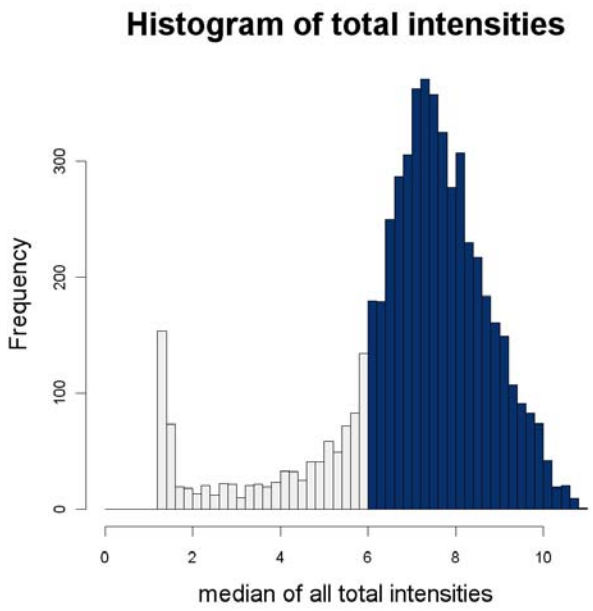


Figure 9: Genes that were below a log-intensity of 6 in at least 4 of the 15 total mRNA measurements are discarded.

global spot intensity level on the array. The latter requires normalization which will be discussed in subsection 12.4. Heatscatterplots for visual inspection were produced with the R package LSD [Schwalb *et al.*, 2010]. These colored scatterplots are based on a 2d density to highlight the distribution of the underlying data. This is essential to recognize systematic biases (see Figure 16 for the assessment of the labeling bias).
The density plots of the (log-)expression distributions have pronounced low intensity tails (Figure 9). For the same reasons as for ribosomal protein genes, we decided to cut off genes that had an expression value below a log-intensity of 6 (natural log basis) in at least 4 out of 15 total mRNA measurements. This cutoff is arbitrary, but it can be chosen even stricter without qualitatively effecting the results (data not shown). Finally a set $G^{reliable}$ of 4490 genes remained.
The pairwise correlation plot showed a high correlation of the respective mRNA fractions within replicates (Supplementary Figure S2, Figure 11), but also a surprisingly high correlation between different mRNA fractions (Figure 10).

12.4 Proportional rescaling of expression profiles

Variations in RNA extraction efficiencies, amplification steps in the biochemical protocol and scanner calibration of the flourescence readouts introduce slight differences in the global spot intensity levels on the arrays. This problem is often solved by centering the medians the respective expression profiles to a common value. This approach however, is only reasonable, if the assumption can be met, that the global level of expression has not changed. We have done so in Section 12.5. In our DTA procedure, this is not necessary. Proportional rescaling of expression profiles is completely compensated by the total least squares regression, which is a feature of our estimation procedure (Section 13.2).

12.5 Detection of differentially expressed genes

We mainly aim to identify genes that behave differently in the comparison of two groups of genome-wide measurements, be it total mRNA levels, labeled mRNA levels, or synthesis rates. The problem of identifying differentially expressed genes in microarray experiments with arbitrary numbers of groups and mRNA samples was considered by Gordon K. Smyth [Smyth, 2004]. His model is formulated as a linear regression problem. The estimators proposed show robust behavior even for small numbers of arrays. The approach is implemented in the R Bioconductor package "Limma" [Smyth, 2004], which was used in Section 12.6 after appropriate normalization of the used expression profiles (see Section 12.4). Multiple testing correction was done by converting p-values into local false discovery rates [Benjamini & Hochberg, 1995]. We consider a

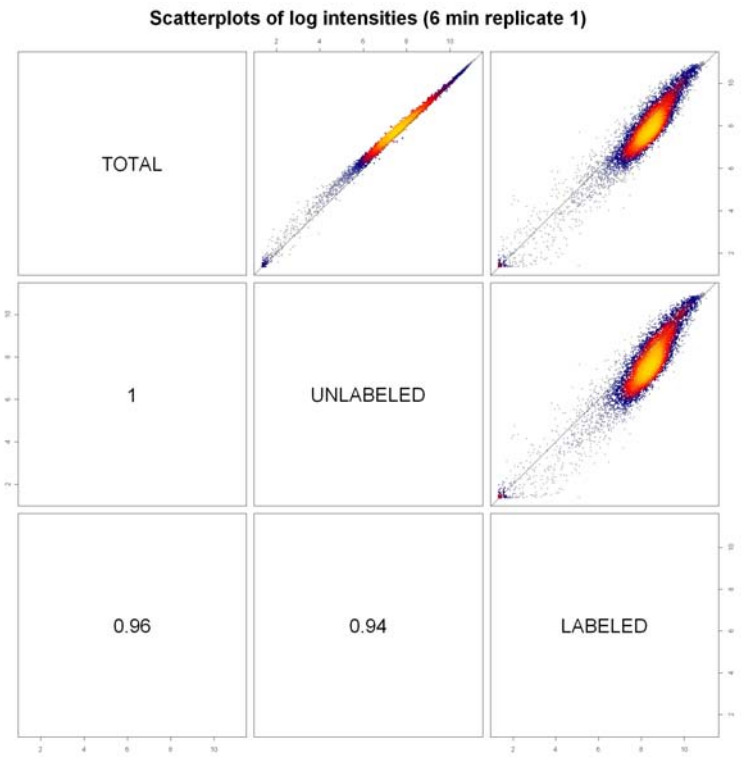


Figure 10: Upper triangle: scatterplots of the log-intensities $T_{gr,}U_{gr}$ and L_{gr} depicted for the 1^{st} replicate of the 6 min measurement. Lower triangle: respective spearman correlations.

gene *significantly differentially expressed* if it achieves a local false discovery rate smaller than 5% in the respective two-group comparison. The effect of induction/repression is called *relevant* if the mean expressions of the two groups differ by a factor of at least 2. A gene is called *induced/repressed* if it is significantly and relevantly up-/downregulated relative to the reference group. (Figure 12).

12.6 Side Effects of 4sU Labeling

At first glance one can only recognize a slight effect after a labeling period of 24 minutes (Figure 11).
This is consistent with the analysis for detecting differential expression: In the comparison of wild type samples with and without the human nucleoside transporter hENT (T00 versus C00) only one gene is detected to be *induced/repressed* (see Section 12.5, and Figure 12). In the comparison of the samples with 24 min of 4sU labeling time (T24) versus wild type control (C00), 117 genes are detected to be *induced/repressed* (see Section 12.5, and Figure 12). This is of course not a strong effect, but enough to decide against the 24 min labeling period.

repressed \ induced	C00	T00	T03	T06	T12	T24
C00	-	1	0	2	30	54
T00	0	-	0	5	10	42
T03	0	0	-	0	10	44
T06	1	2	0	-	4	28
T12	26	3	1	0	-	2
T24	63	22	47	24	0	-

Table 4: Counts of induced/repressed genes (upper triangle/lower triangle) among $G^{reliable} \cup G^{rpg}$(reliable and ribosomal protein genes). Selection criteria were a multiple testing corrected significance level (local false discovery rate) of 0.05 and an expression fold of at least 2 between groups. (Abbreviation: mRNA fraction followed by labeling time.)

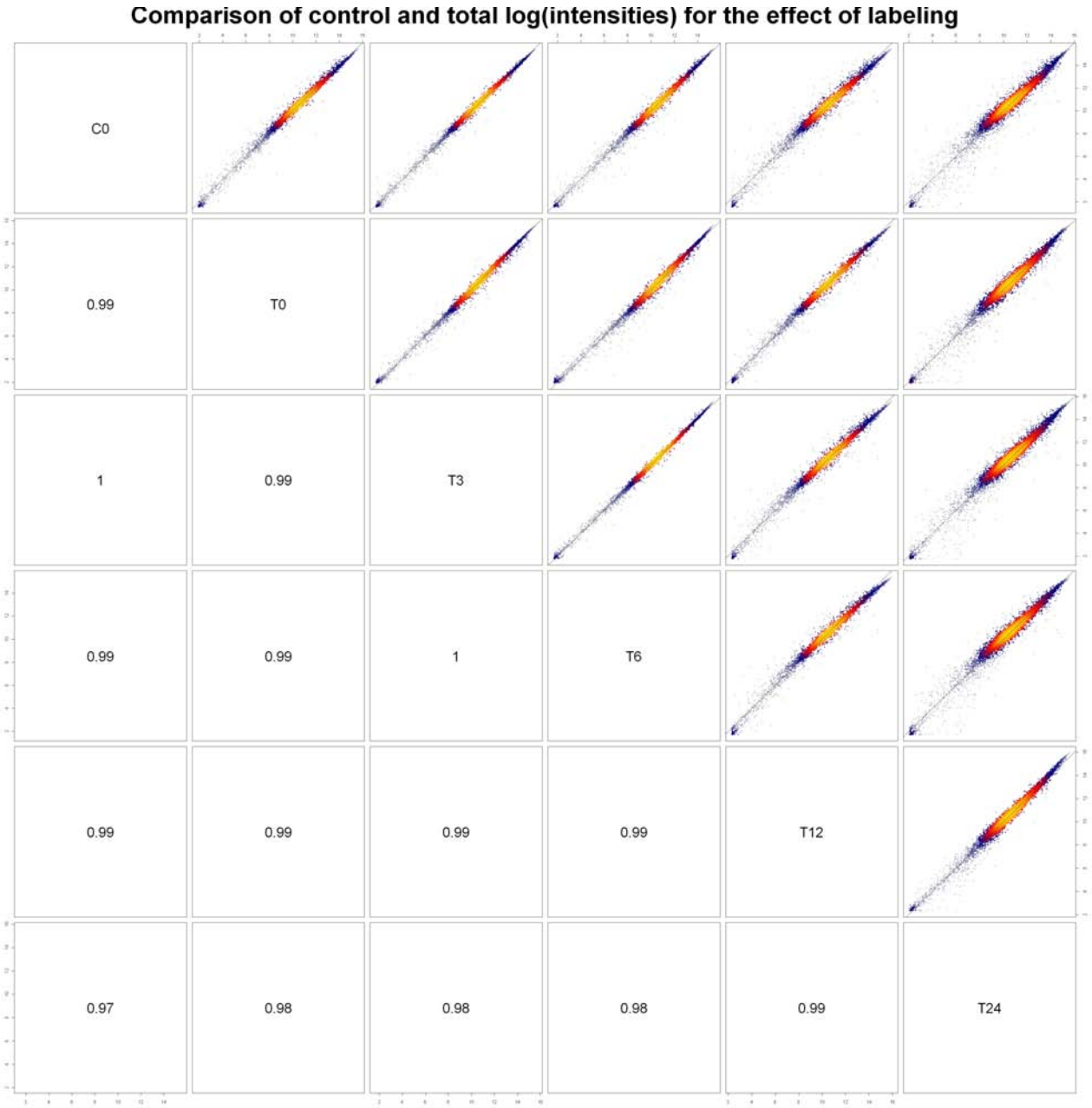


Figure 11: Pairwise scatterplots of log-intensities. The lower panel shows the respective Spearman correlations. The diagonal contains the abbreviation of the mRNA fraction followed by the length of labeling period in minutes. Compared fractions are obtained by taking the gene-wise median over all intensities of replicate measurements. C denotes control samples lacking the nucleoside transporter (hENT1). All other samples (T) carry the nucleoside transporter.

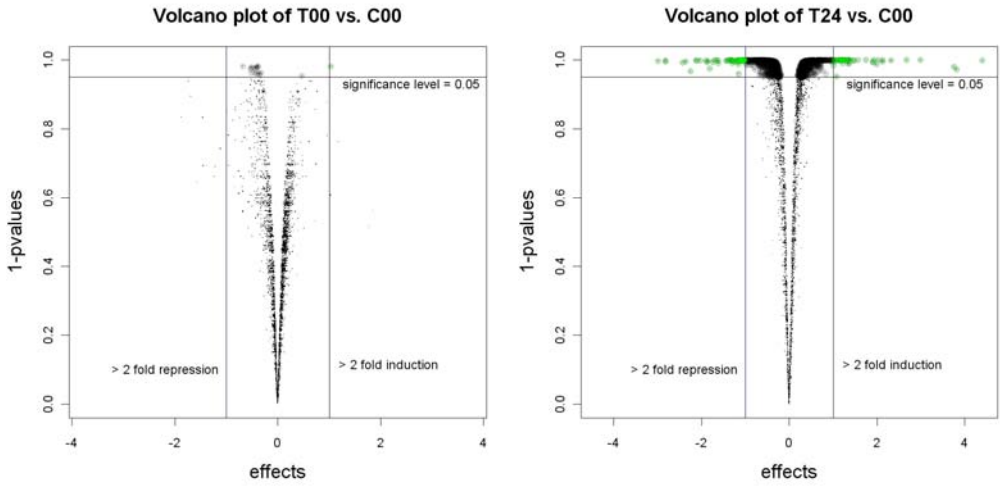


Figure 12: Volcano plot of all genes in the yeast transporter strain without labeling (T00) versus wild type control (C00) (left figure) and of all genes of the yeast transporter strain with 24 min of 4sU labeling (T24) versus wild type control (C00) (right figure). The x-axis shows the estimated folds in the limma analysis. The y-axis displays the adjusted p-values, 1 minus the adjusted p-values to be precise. The horizontal line marks the significance threshold of 0.05. The two vertical lines indicate induction/repression of 2 fold. The genes that are called *induced/repressed* (Section 12.5) are depicted in green.

The statistical comparison of the samples shows that the measurements are conform with the assumption (null hypothesis) that no genes are *induced/repressed*. Although these tests by their nature cannot produce an affirmative result, i.e. this comparison cannot prove the equality of the expression distributions of the wild type and $t = 6$ min 4sU labeling. It ensures that the requirements for the application of our subsequent procedures are met.

13 mRNA Synthesis and Decay in steady-state conditions

For sections 13.1 and 13.2, we assume that the cells exhibit constant growth under constant environmental conditions. In particular, this implies that the amount of each mRNA population is constant over time, being the result of a dynamic equilibrium of a constant mRNA synthesis and decay.

13.1 The steady-state Model

Let $r \in R$ be a sample. At time $t = 0$, we start the mRNA labeling. At the timepoint t_r, when the cells are harvested, the total mRNA amount $C_{gr}(t_r)$ of gene g in the sample r is composed of the amount $B_{gr}(t_r)$ of (pre-existing) mRNA that has been synthesized *before* $t = 0$ and the amount $A_{gr}(t_r)$ of mRNA that has been newly synthesized *after* $t = 0$,

$$\underbrace{C_{gr}(t_r)}_{\text{total RNA}} = \underbrace{A_{gr}(t_r)}_{\text{newly synthesized mRNA}} + \underbrace{B_{gr}(t_r)}_{\text{pre-existing mRNA}}$$

Let $N_r(t_r)$ denote the number of cells in the sample r at time t_r. The cells are grown and harvested during mid-log phase, i.e. the cell number follows an exponential law with growth rate $\alpha \geq 0$,

$$N_r(t_r) = N_r(0)e^{\alpha t_r} \tag{1}$$

We assume genes to have a (time averaged) constant cellular expression level m_g (transcripts of gene g per cell) during 4sU labeling. The total mRNA amount of a gene is therefore proportional to the cell number,

$$C_{gr}(t_r) = m_g N_r(t_r) = m_g N_r(0)e^{\alpha t_r} = C_{gr}(0)e^{\alpha t_r} \tag{2}$$

We assume that the mRNA population of a gene g decays at a constant rate λ_g if no other processes interfere. This means for the pre-existing mRNA fraction that

$$B_{gr}(t_r) = B_{gr}(0)e^{-\lambda_g t_r} = C_{gr}(0)e^{-\lambda_g t_r} \tag{3}$$

consequently, the newly synthesized mRNA fraction is

$$A_{gr}(t_r) = C_{gr}(t_r) - B_{gr}(t_r) = C_{gr}(0)e^{\alpha t_r} - C_{gr}(0)e^{-\lambda_g t_r} = C_{gr}(0)\left[e^{\alpha t_r} - e^{-\lambda_g t_r}\right] \tag{4}$$

The very same equation can also be deduced from a comparison of synthesis and decay processes due to the the newly synthesized mRNA fraction

$$\frac{dA_{gr}(t_r)}{dt} = \mu_g N_r(t_r) - \lambda_g A_{gr}(t_r) = \mu_g N_r(0)e^{\alpha t_r} - \lambda_g A_{gr}(t_r) \tag{5}$$

with a constant synthesis rate μ_g. The solution of this differential equation yields

$$A_{gr}(t_r) = ce^{-\lambda_g t_r} + \frac{\mu_g N_r(0)e^{\alpha t_r}}{\alpha + \lambda_g} \tag{6}$$

with an initial value $A_{gr}(0) = 0$, and so

$$0 = c + \frac{\mu_g N_r(0)}{\alpha + \lambda_g} \tag{7}$$

This finally leads to

$$A_{gr}(t_r) = \frac{\mu_g N_r(0)}{\alpha + \lambda_g}\left[e^{\alpha t_r} - e^{-\lambda_g t_r}\right] \tag{8}$$

with

$$C_{gr}(0) = \frac{\mu_g N_r(0)}{\alpha + \lambda_g} \tag{9}$$

which can be used to express the synthesis rate μ_g by

$$\mu_g = m_g\left[\alpha + \lambda_g\right] \tag{10}$$

We now have to relate the measured levels of $L_{gr}(t_r)$, $U_{gr}(t_r)$ and $T_{gr}(t_r)$ to the levels of the mRNA fractions $A_{gr}(t_r)$, $B_{gr}(t_r)$ and $C_{gr}(t_r)$. Ideally, these fractions would respectively equal each other. There are however discrepancies that are due to RNA extraction efficiencies, amplification steps in the biochemical protocol and scanner calibration of the flourescence readouts (see Section 12.4 and Figure 10). The amount $L_{gr}(t_r)$ of labeled mRNA for instance is proportional to the amount of labeled mRNA $A_{gr}(t_r)$ at the time t_r of sampling,

$$L_{gr}(t_r) = a_r A_{gr}(t_r)\ , \tag{11}$$

with an unknown array-specific constant a_r. Analogously, the measured amounts $T_{gr}(t_r)$ and $U_{gr}(t_r)$ analogical depend on the actual amounts $C_{gr}(t_r)$ and $B_{gr}(t_r)$ respectively via

$$T_{gr}(t_r) = c_r C_{gr}(t_r)\ , \tag{12}$$

and

$$U_{gr}(t_r) = b_r B_{gr}(t_r) = b_r\left(C_{gr}(t_r) - A_{gr}(t_r)\right) \tag{13}$$

where c_r and b_r are array-specific scaling factors.
There are also discrepancies that are due to 4sU/Biotin labeling efficiency (see Figure (13)). mRNAs which contain less than 500 Uridine residues (approx. 72% of all mRNAs) are not captured efficiently since approximately only every 200^{th} Uridine residue is replaced by 4sU and afterwards attached to a Biotin molecule. Let l_{gr} represent the fraction mRNAs of gene g in sample r that are biotinylated. We assume that all biotinylated mRNAs are captured by the Streptavidin beads. Let p_r be the probability that during the labeling process of sample r, a Uridine is replaced by 4sU and afterwards attached to a Biotin molecule. Denote by $\#u_g$ the number of Uridine residues present in the mRNA corresponding to gene g, a number known from the literature. We can calculate l_{gr} as

$$l_{gr} = l(p_r, u_g) = 1 - (1 - p_r)^{\#u_g} \tag{14}$$

l_{gr} is thus the probability that at least one Uridine is replaced by 4sU and afterwards attached to a Biotin molecule. With this circumstance in mind, we have to correct equations (11),(12) and (13) to that effect. Hence we have the dependencies:

$$L_{gr}(t_r) = l_{gr} a_r A_{gr}(t_r)\ , \tag{15}$$

$$T_{gr}(t_r) = c_r C_{gr}(t_r)\ , \tag{16}$$

and

$$U_{gr}(t_r) = b_r\left(C_{gr}(t_r) - l_{gr}A_{gr}(t_r)\right). \tag{17}$$

13.2 Normalization and Parameter Estimation (steady state case)

Our model contains the parameters $\Theta = \{\alpha,\, p_r,\, a_r,\, b_r,\, c_r,\, \lambda_g,\, \mu_g \mid r \in R,\, g \in G\}$. It contains an implicit normalization procedure, because the two sources of experimental bias are part of the model (the parameters a_r, b_r, c_r account for multiplicative bias introduced via sample preparation and array scanning (Section 12.4), and p_r models the labeling bias (Figure 13)). We propose a 5-step procedure for the identification of the parameters Θ.

1. Estimation of α, the growth rate of the cells. Since the doubling times of the cells are usually known or can be measured accurately, α is given by $\alpha = \log 2/(\text{cell cycle length})$. Cell cycle length is set to 150 min.

2. Estimation of p_r. The estimation of the sample-related parameters $\{p_r,\, a_r,\, b_r,\, c_r \mid r \in R\}$ is done on the basis of the reliable genes $G^{reliable}$ (cf. Section 12.3). The quotient of observed total and labeled mRNA levels can be written as

$$\frac{L_{gr}}{T_{gr}} = \frac{l_{gr} a_r A_{gr}(t_r)}{c_r C_{gr}(t_r)} = l_{gr} \frac{a_r}{c_r} \left[1 - e^{-t_r(\alpha+\lambda_g)}\right] \tag{18}$$

   The first equation follows by (15) and (16), the second by (2) and (4). We can visualize this dependence conveniently by plotting u_g versus $\log \frac{L_{gr}}{T_{gr}}$ (see Figure 13). If all decay rates were equal, all points would lie on the graph given by the relationship of u_g versus $\log l_{gr} + \log \frac{a_r}{c_r}$. The scatter around this graph is caused by measurement errors and differences in decay rates (see Figure (13),(13)). We can also calculate the quotient

$$\frac{U_{gr}}{T_{gr}} = \frac{b_r}{c_r} \left(1 - l_{gr} \left[1 - e^{-t_r(\alpha+\lambda_g)}\right]\right) \tag{19}$$

   This equation follows by (16) and (17). We will predominantly use equation (18) for the estimation of p_r. Taking logs in Equation (18) and rearranging terms, we obtain

$$\log \frac{L_{gr}}{T_{gr}} = \log \frac{a_r}{c_r} + \log l(p_r, u_g) + \log \left[1 - e^{-t_r(\alpha+\lambda_g)}\right] \tag{20}$$

   Assuming $p > 1/700$ implies that for $\#u_g > 700$ say, the approximation $l_{gr} = l(p_r, u_g) \approx 1$ is almost exact, as one can easily see from the Bernoulli approximation $(1-p_r)^{\#u_g} \approx 1 - \#u_g \cdot p_r = 0$ and $l_{gr} = 1 - (1-p_r)^{\#u_g} \approx 1$. Hence Equation (18) simplifies to

$$\log \frac{L_{gr}}{T_{gr}} = \log \frac{a_r}{c_r} + \log \left[1 - e^{-t_r(\alpha+\lambda_g)}\right] \quad \text{for} \quad \#u_g > 700 \tag{21}$$

   If we additionally assume that the distribution of decay rates do not depend on the number of the uridines, the right-hand side in (21) becomes a constant plus some error term with expectation zero. Thus, we estimate $asymptote_r$ by letting

$$asymptote_r = \text{median} \left\{ \log \frac{L_{gr}}{T_{gr}} \mid g \in G^{reliable},\ \#u_g > 700 \right\} \tag{22}$$

   Given equation (22), it is relatively easy to compute a good estimate of p_r by finding an optimal fit to (20) (see Figure(13)), for all $g \in G$ with $\#u_g < 500$. So we optimize the value of p_r, $r \in R$, by minimizing the l_1-loss function

$$p_r^{est} = \underset{q \in (0,\,1)}{\text{argminloss}}(q) \quad . \text{ with} \quad \text{loss}(q) = \sum_{g \in G,\, \#u_g < 500} \left| \log \frac{L_{gr}}{T_{gr}} - \log l(q, u_g) - asymptote_r \right| \tag{23}$$

   Here, 500 is an upper bound that ensures that the measurements are still responsive to changes in u_g.

3. Estimation of $\frac{a_r}{c_r}$ and $\frac{b_r}{c_r}$. Notice that for the purpose of half-life estimation it is sufficient to determine the quotients $\frac{a_r}{c_r}$ and $\frac{b_r}{c_r}$ instead of the individual constants a_r, b_r and c_r. We simply multiply equations (18) and (19) by the inverse of those quotients and add them up to obtain

$$\frac{c_r}{a_r} \frac{L_{gr}}{T_{gr}} + \frac{c_r}{b_r} \frac{U_{gr}}{T_{gr}} = 1 \quad \text{or} \quad T_{gr} = \frac{c_r}{a_r} L_{gr} + \frac{c_r}{b_r} U_{gr} \tag{24}$$

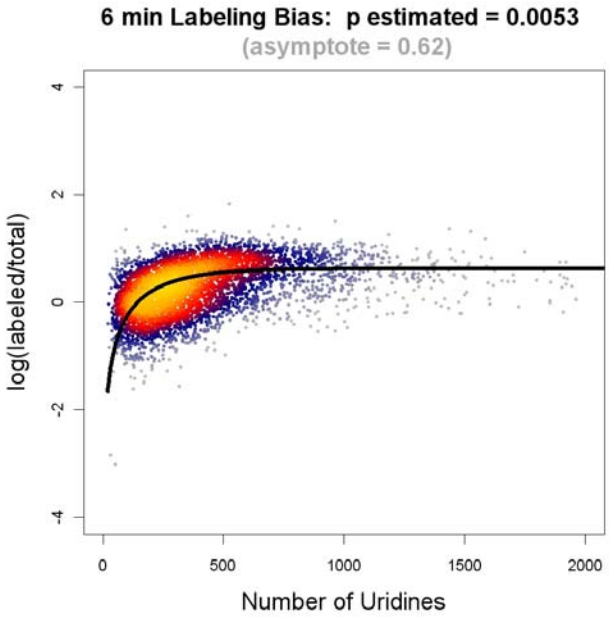


Figure 13: Scatterplot shows the dependence of equation (20) for the 6 min measurements. The number of Uridines is plotted versus the log-ratio of L_{gr} and T_{gr}. The black line shows that the estimate for $\log l(p_r^{est}, u_g) + asymptote_r^{est}$ fits the data. $p_r = 0.0053$ means that approximately only every 200^{th} Uridine residue is replaced by 4sU and afterwards attached to a Biotin molecule.

Equation 24 describes a plane $\{(T_{gr}, L_{gr}, U_{gr}) \mid T_{gr} = \frac{c_r}{a_r} L_{gr} + \frac{c_r}{b_r} U_{gr}\}$ in a3-dimensional Euclidean space. For error-free measurements, two observations (T_{gr}, L_{gr}, U_{gr}) would be enough to determine the two coefficients $\frac{a_r}{c_r}$ and $\frac{b_r}{c_r}$. We encounter variables having measurement errors on both sides of Equation (24). We therefore perform a total least squares regression of T_{gr} versus L_{gr} and U_{gr}, which accounts for a Gaussian error in the dependent variable (T_{gr}) and, in contrast to ordinary linear regression, also in the independent variables (L_{gr}, U_{gr}). The total least squares regression minimizes the orthogonal distance of the datapoints to the inferred plane as opposed to a linear regression, which minimizes the distance of T_{gr} to the inferred linear function of L_{gr} and U_{gr}. We use a robust version of total least squares regression.After the first run, we remove the data points with the 5% largest residues to avoid the potentially detrimental influence of outlier values on the parameter estimation process (see Figure (14)). We remark that we also tried to normalize the intensities of the respective fractions to the in vitro transcribed Bacillus subtilis spike ins, which were routinely added according to ourexperimental protocol, but this resulted in very inconsistent scaling of the labeled and total fractions (data not shown).

4. Estimation of λ_g, $g \in G^{reliable} \cup G^{rpg}$. First, solve equation (18) for an estimate λ_g,

$$\lambda_g = -\alpha - \frac{1}{t_r} \log \left[1 - \frac{1}{l(p_r, u_g)} \frac{L_{gr}}{T_{gr}} \frac{c_r}{a_r} \right]. \tag{25}$$

All measured samples r are combined to yield an estimate

$$\lambda_g^{est} = \text{median} \left\{ -\alpha - \frac{1}{t_r} \log \left[1 - \frac{1}{l(p_r, u_g)} \frac{L_{gr}}{T_{gr}} \frac{c_r}{a_r} \right] \mid r \in R \right\}, \tag{26}$$

which in turn is used to calculate the half-life estimates

$$hl_g^{est} = \frac{\log(2)}{\lambda_g^{est}}. \tag{27}$$

5. Estimation of μ_g, $g \in G^{reliable} \cup G^{rpg}$ (in molecules per cell and cell cycle). For each replicate experiment r, the number m_g of mRNA transcripts per cell of gene g is proportional to its total mRNA

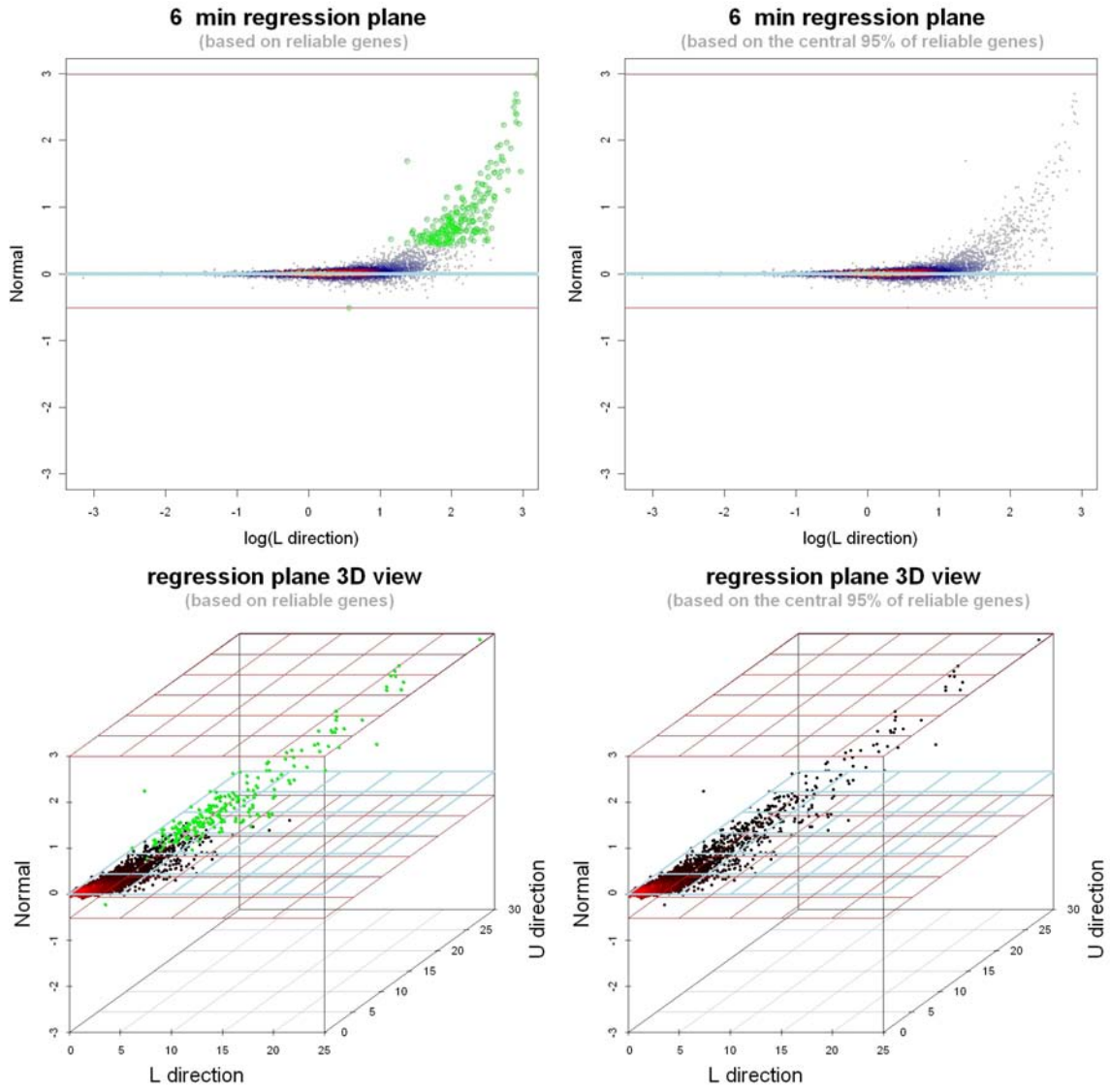


Figure 14: The upper plots show the two rounds of total least squares regression. The resulting plane is shown exactly from the side and is colored blue. The x-axis of those plots is chosen in the direction of the labeled fraction and in logarithmic scale. The y-axis is the normal of the plane. The second round (shown on the right-hand side) is performed without the 5% largest residues of the first round, depicted in green and shown in both plots on the left-hand side. It is noteworthy that these are mostly ribosomal protein genes G^{rpg}. This is a second justification for excluding these genes from parameter estimation, which is done in the very beginning. The lower plots show the plane which is fitted into the data in a 3D view. Red lines or planes indicate maximal residues.

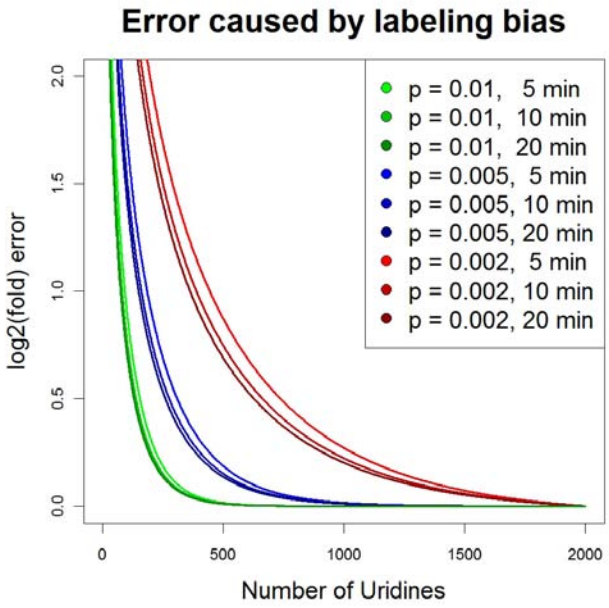


Figure 15: Plot u_g vs. bias (log2 est.decay/true decay) curves, each curve corresponding to a given half-life (5, 10 or 20 min) and labeling efficiency ($p = 0.002$, $0,005$ or 0.1).

intensity value T_{gr}, $m_g = d_r T_{gr}$. Assuming a total number of $\#mRNAs = 15.000$ mRNAs per cell [Hereford, 1977], this means that

$$15.000 = \#mRNAs = \sum_{g \in G} m_g = d_r \cdot \sum_{g \in G} T_{gr} \ , \tag{28}$$

thus

$$d_r = \frac{15.000}{\sum_{g \in G} T_{gr}} \ . \tag{29}$$

Together with equation (10), we may estimate μ_g as

$$\mu_g = m_g \left[\alpha + \lambda_g^{est}\right] \cdot CCL = \text{median}\left\{d_r T_{gr} \mid r \in R\right\} \left[\alpha + \lambda_g^{est}\right] \cdot CCL \ . \tag{30}$$

where the cell cycle length CCL is set to 150 min.

13.3 Improvement over existing methods

There have been attempts to obtain genome-wide synthesis rates and half-lives from RNA labeling in other organisms, like mouse and human [Dölken *et al.*, 2008, Friedel *et al.*, 2009]. The statistical analysis presented here is improves over existing approaches in at least two major points:

1. The labeling bias is corrected. Assuming a labeling efficiency of $p = 0.005$ as has been estimated in our data, (Figure 18) shows that omitting labeling bias correction results in half-life estimates that are severely systematically biased towards longer half-lives; 20% of the mRNA half-lives in our experiment would have been biased more than twofold (Figure 15). The labeling bias is smaller for higher labeling efficiencies. Therefore we re-analysed the mouse data in [Dölken *et al.*, 2008]. The diagnostic scatterplot for the assessment of the labeling bias (main text Figure 2A) showed that a bias is also present in their data (see Figure 16), and it was estimated to $p = 0.002$ by our method. Consequently, mRNAs of a length less than 500 (60% of all genes) had an average bias of at 1.7-fold. Labeling bias correction is done as point 2 of our estimation procedure.

2. Estimation of the normalization constants is done with total least squares regression. For normalization purposes, the parameters of the regression plane, namely the quotients $\frac{a_r}{c_r}$ and $\frac{b_r}{c_r}$ in Equation (24) need to be estimated. The straightforward idea is to perform a linear regression (without intercept) of

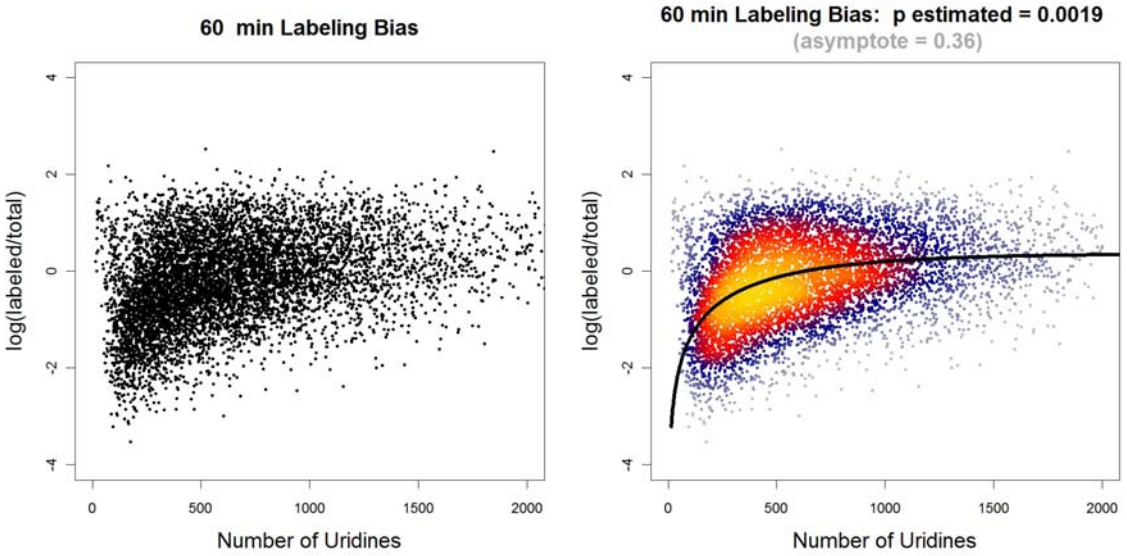


Figure 16: Both scatterplots show exactly the same data: the dependence of Equation (20) for the 60 min measurements. The number of Uridines is plotted versus the log-ratio of L_{gr} and T_{gr}. Whereas it is hard to recognize a labeling bias in the plot to the left, such is clearly visible in the right hand side plot. The points of the scatteplot are colored according to the (estimated) point density in that region, which turns out to be a valuable information. The plot has been generated with the R package LSD [Schwalb *et al.*, 2010]. The black line shows the estimate for $\log l(p_r^{est}, u_g) + asymptote_r^{est}$. The labeling bias parameter $p_r = 0.0019$ implies that approximately every 400^{th} Uridine residue is replaced by 4sU and afterwards attached to a Biotin molecule.

T_{gr} versus L_{gr} and U_{gr} [Dölken *et al.*, 2008, Friedel *et al.*, 2009]. Then, the quotients $\frac{a_r}{c_r}$ and $\frac{b_r}{c_r}$ are obtained as the inverse of the regression coefficients. In our situation however, we have to deal with considerable experimental noise in the dependent *and* the independent variables in Equation (24). The linear regression model assumes that only the dependent variable T_{gr} is error-prone, whereas the independent variables L_{gr} and U_{gr} are known without error; an assumption which is clearly not met. As we mentioned in (Section 12.3), the ribosomal protein genes (G^{rpg}) have expression levels that are way higher than the overall gene expression levels, thus they are outlier points, and any regression method that tries to minimize the squared estimation error (like linear regression and total least squares) is very sensitive to outliers. Thus, a robust version of total least squares needs to remove outliers beforehand the regression, which has not been done in previous studies. Our outlier removal procedure is described in (Section 13.2).

13.4 Simulation Study (steady state case)

To examine the estimation procedure described above, we simulated data for 5000 "genes" by providing a random half-life distribution and a random sample of corresponding size from the uridine numbers of the S.cerevisiae genes. The half-lives are assumed to have a right skewed distribution. This assumption is based on literature results. Half-lives are then drawn randomly from a log normal distribution whose logarithm has mean equal to 14 and standard deviation equal to 0.3.

$$H_r \sim \mathcal{LN}(14, 0.3) \tag{31}$$

The "true" amount of log(total mRNA) at timepoint $t = 0$ is drawn randomly from a normal distribution with mean equal to 7.5 and standard deviation equal to 1. This gives a microarray-typical intensity distribution in log-scale. The following steps are now in analogy with our model:

$$C_{gr}(t_r) = C_{gr}(0)e^{\alpha t_r} \tag{32}$$

where α is given by $\alpha = \log 2/(\text{cell cycle length})$. Cell cycle length is set to 150 min. t_r is chosen to be the timepoint of favor. $T_{gr}(t)$ does now arise from $C_{gr}(t)$ by adding noise:

$$T_{gr}(t_r) = C_{gr}(t_r) + x_{gr} \quad \text{with} \quad x_{gr} \sim \mathcal{N}(0, 0.125 C_{gr}(t_r)) \tag{33}$$

where $\mathcal{N}(0, 0.125C_{gr}(t_r))$ denotes a normal distribution with mean equal to 0 and standard deviation equal to $0.125C_{gr}(t_r)$. This is consistent with the fact that the variance of measured data increases with intensity. We set

$$c_r := \frac{\text{median}(T_{gr}(t_r))}{\text{median}(C_{gr}(t_r))} \tag{34}$$

which can be derived from equation (16). The "true" amount of newly synthesised mRNA, called $A_{gr}(t)$ in our model, is constructed as follows:

$$A_{gr}(t_r) = C_{gr}(t_r)(e^{\alpha t_r} - e^{-\lambda_g t_r}) \tag{35}$$

where $\lambda_g = \log(2)/H_{gr}$. To obtain L_{gr} we include a labeling bias l_{gr} as in (14), and add noise:

$$L_{gr}(t_r) = l_{gr}A_{gr}(t_r) + y_{gr} \quad \text{with} \quad y_{gr} \sim \mathcal{N}(0, 0.125 l_{gr}A_{gr}(t_r)) \tag{36}$$

p_r is set to 0.005. This value was typically observed in our wild type experiments (see section 12.1). We set

$$a_r := \frac{\text{median}(L_{gr}(t_r))}{\text{median}(l_{gr}A_{gr}(t_r))} \tag{37}$$

Further we rescale L_{gr} to the same range as T_{gr}:

$$L_{gr} = \frac{L_{gr} \cdot \text{median}(T_{gr})}{\text{median}(L_{gr})} \tag{38}$$

This strategy simulates amplification steps in the biochemical protocol and scanner calibration. Finally, we need the "true" amount of unlabeled mRNA:

$$\widetilde{B}_{gr}(t_r) = C_{gr}(t_r) - l_{gr}A_{gr}(t_r) \tag{39}$$

and again we add noise

$$U_{gr}(t_r) = \widetilde{B}_{gr}(t_r) + z_{gr} \quad \text{with} \quad z_{gr} \sim \mathcal{N}(0, 0.125\widetilde{B}_{gr}(t_r)) \tag{40}$$

Furthermore we also rescale U_{gr} to the same range as T_{gr}:

$$U_{gr} = \frac{U_{gr} \cdot \text{median}(T_{gr})}{\text{median}(U_{gr})} \tag{41}$$

and

$$b_r := \frac{\text{median}(U_{gr}(t_r))}{\text{median}(\widetilde{B}_{gr}(t_r))} \tag{42}$$

We generated an artificial data set which consists of microarray measurements for $t = 6$ and 12 minutes, each timepoint measured in duplicates. The "true" coefficients, that are to be recovered in (24) by total least squares regression, are now just the quotients of the individual constants a_r, b_r and c_r. We then applied our model to estimate the "true" half-lives with our procedure by estimating $\frac{a_r}{c_r}, \frac{b_r}{c_r}, p_r$ and λ_g, see (Figure (17)). We also used our simulated data to recompute the "true" half-lives with the former used first-order exponential decay model as described in [Dölken *et al.*, 2008]. Briefly: As it is default in GCRMA, quantile-normalization was performed on all "simulated arrays". To overcome the problem of similar magnitudes between the different mRNA fractions, a weighted linear regression of T_{gr} versus L_{gr} and U_{gr} (with no intercept) is fit to obtain the coefficients to rescale the L_{gr} and U_{gr} fractions. Weights were chosen to $1/(T_{gr} + \text{median}(T_{gr}))$. Cell cycle length was also set to 150 minutes. The ratio of U_{gr} to T_{gr} was then used to calculate the half-lives from the 6 min duplicates. Nevertheless, the bias introduced by the 4sU/Biotin labeling efficiency was not corrected for, see (Figure (18)).

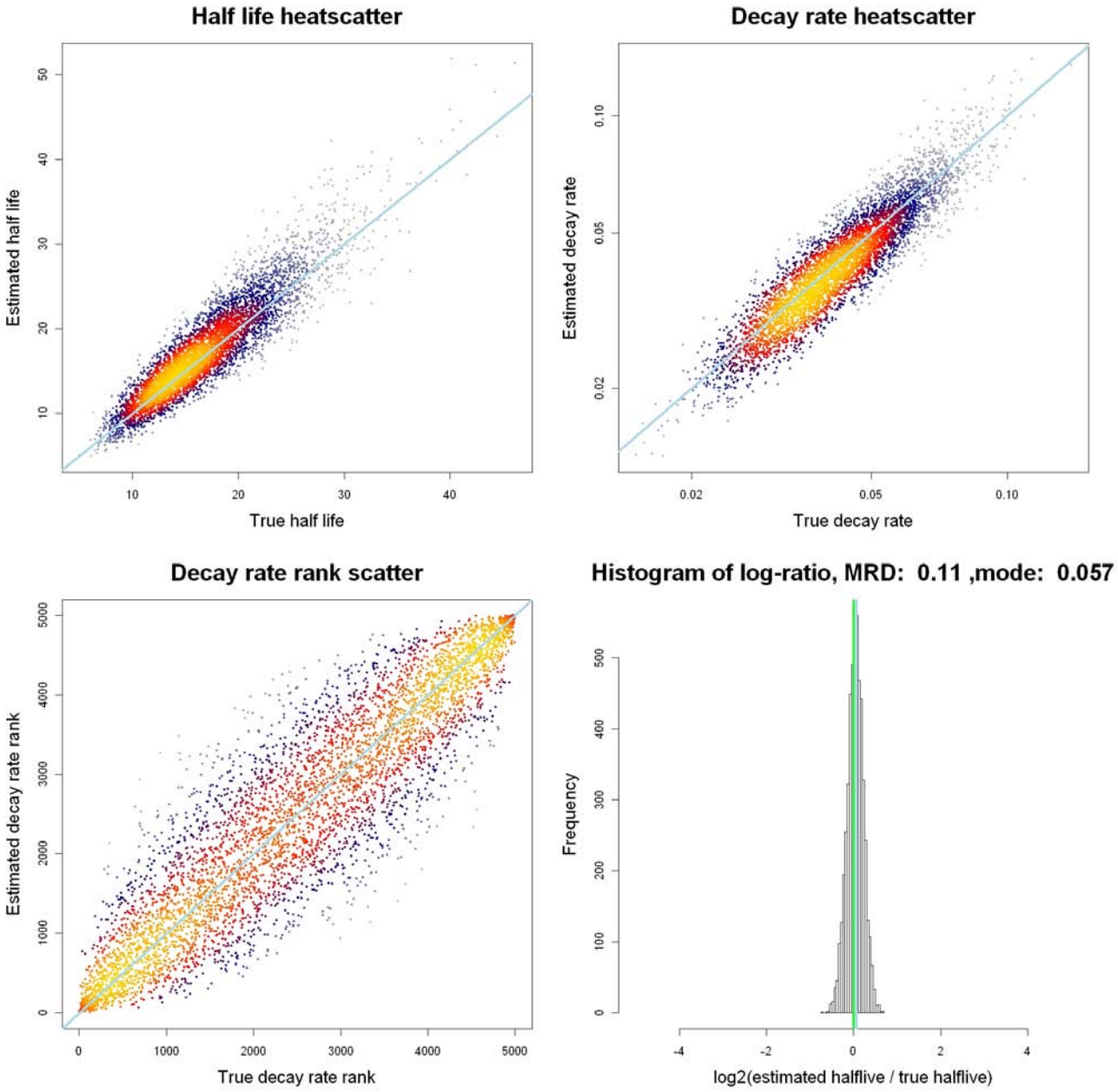


Figure 17: Figure compares the estimated vs. the true half-lives or the estimated vs. the true decay rates in scatterplots of their values or ranks. The lower right plot shows the log-ratio of the estimated vs. the true half-lives in a histogram. The mode is the maximum of the corresponding density indicated by the blue line. As a measure for a systematic deviation from zero (depicted in green) we calculated the MRD (mean relative deviation). The MRD is defined by mean $\left(\frac{|\lambda_g^{est}-\lambda_g^{true}|}{\lambda_g^{true}}\right)$, λ_g^{est}: estimated half-life, λ_g^{true}: true half-life. Plots were produced with the R package LSD [Schwalb *et al.*, 2010].

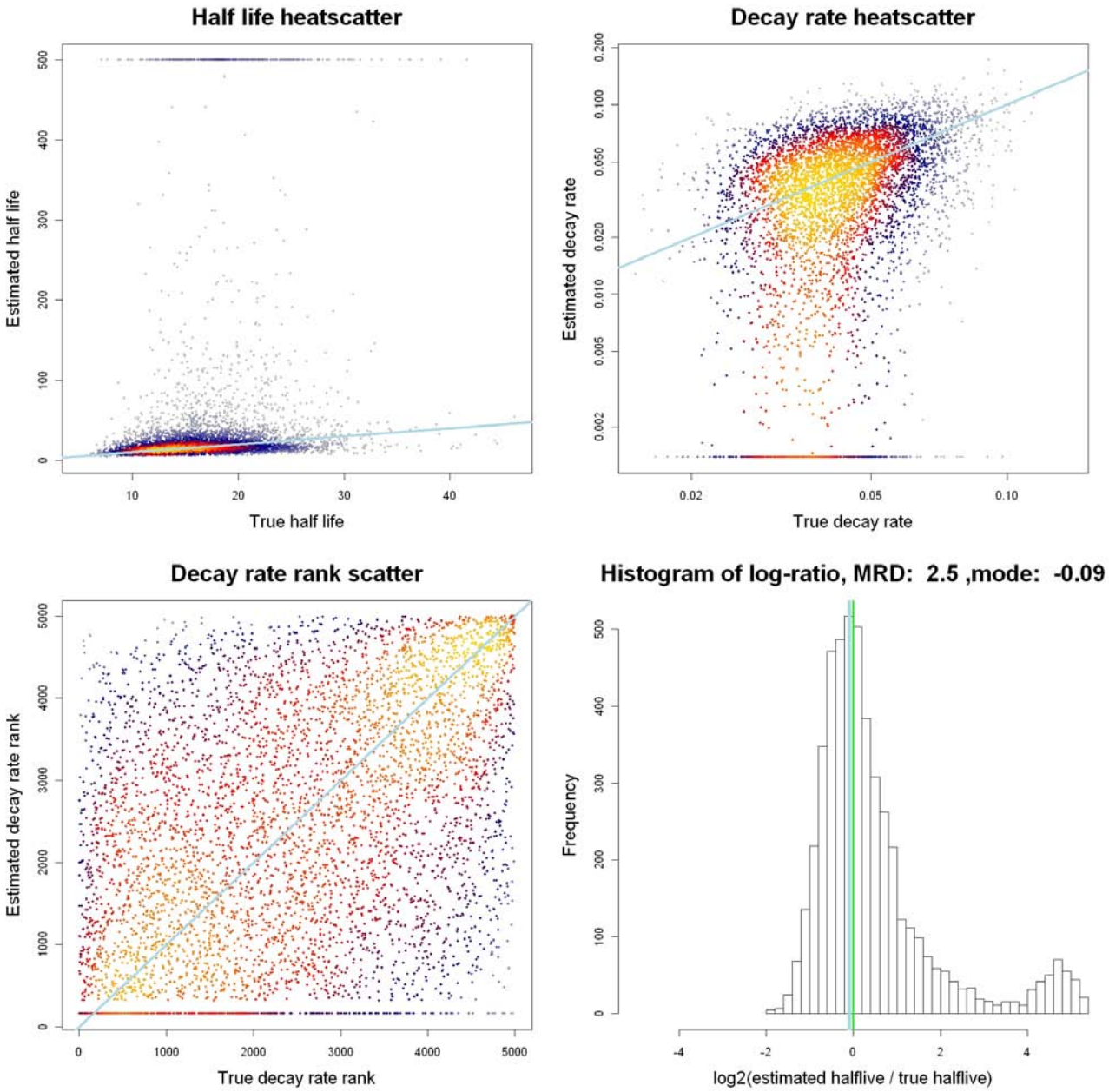


Figure 18: Figure compares the estimated vs. the true half-lives or the estimated vs. the true decay rates in scatterplots of their values or ranks. The lower right plot shows the log-ratio of the estimated vs. the true half-lives in a histogram. The mode is the maximum of the corresponding density indicated by the blue line. As a measure for a systematic deviation from zero (depicted in green) we calculated the MRD (mean relative deviation). The MRD is defined by mean $\left(\frac{|\lambda_g^{est} - \lambda_g^{true}|}{\lambda_g^{true}}\right)$, λ_g^{est}: estimated half-life, λ_g^{true}: true half-life. The points in the scatterplots should be located directly and in a symmetric manner around the blue line to represent a good fit. The histogram is broader than in the results of our procedure, and it additionally shows a systematic error on the right-hand side. Plots were produced with the R package LSD [Schwalb *et al.*, 2010].

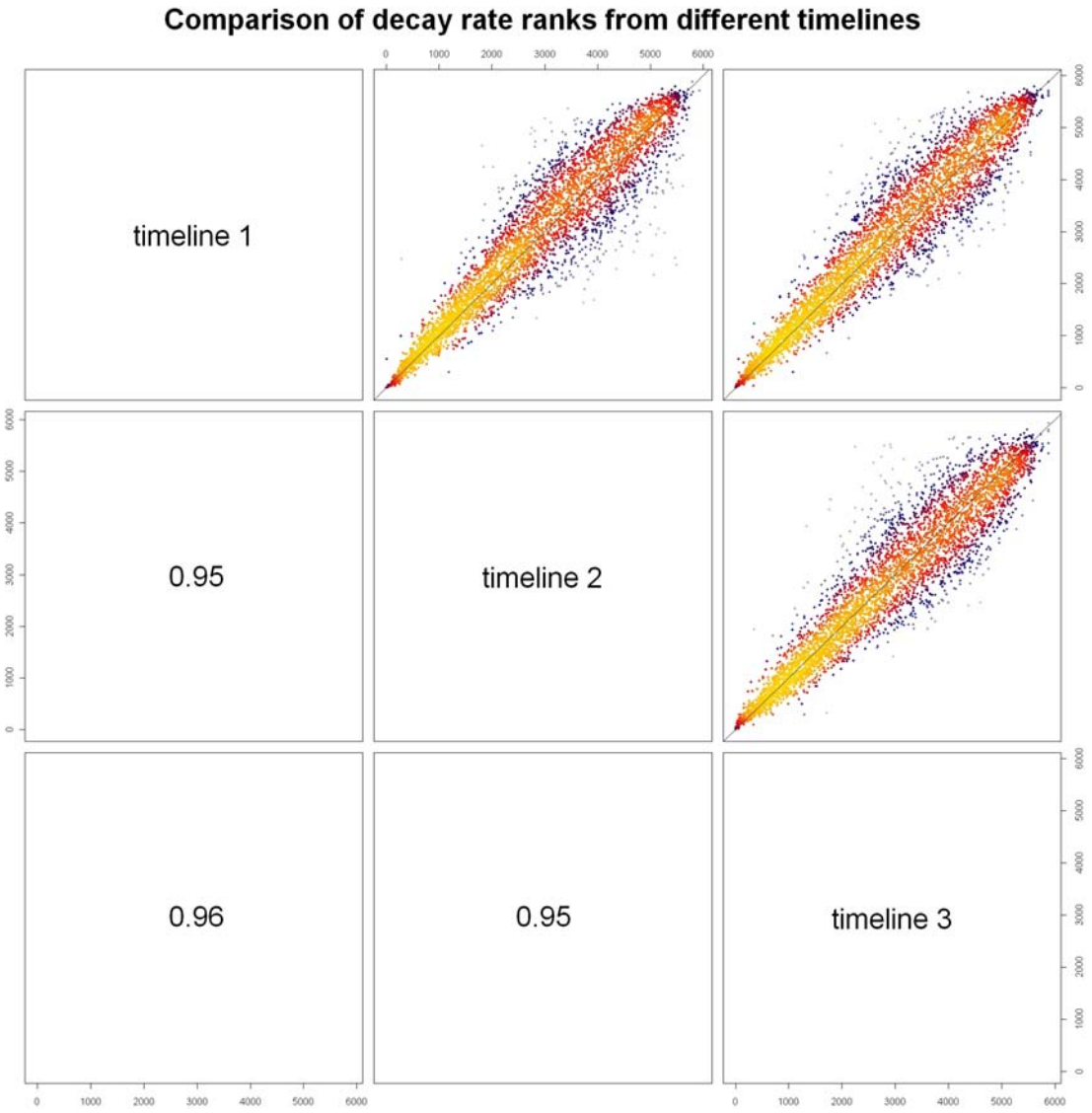


Figure 19: Ranks of estimated decay rates obtained from three different groupings of all measured arrays (termed timelines) compared in pairwise scatterplots. The lower panel shows the respective spearman correlations of the decay rates. Plots were produced with the R package LSD [Schwalb *et al.*, 2010].

13.5 Robustness and Reproducibility (steady state case)

To validate our estimation procedure we used all measured arrays and divided them in three groups (termed timelines). This means we assembled all the first, second and third replicates and used each group to estimate the half-lives by themselves. Even though they partly differ by small factors ($\sim 1.1, 1.5, 1.7$), they correlate well, see (Figure 19).
We also used all measured arrays to calculate the half-lives for all time points each. As expected they also differed by small factors, and as before they correlated well. To investigate this behavior, we altered the regression factors on purpose. This approach resulted in half-lives with identical ranks, which differed by a scaling factor due to the alteration. This shows that the estimation of the regression factors is the sensitive step in our procedure. And so we can not for sure tell the absolute half-life of a gene, but we can clearly rely on the rank of it.
We also simulated data using the estimated half-lives and the array intensities of four measured arrays at time-point 0 to represent $C_{gr}(0)$. The advantage of this approach is, that the correlation between half-lives and array intensities can be chosen arbitrarily. The artificial half-lives and expression data were equipped with a correlation structure a) of zero and b) identical to that in the real experiments. Since the correlations that were estimated from the simulated data agreed very well with the pre-set correlations, this demonstrates that the correlation observed in the real data application is not an artefact of the estimation procedure.

14 The Dynamics of mRNA Synthesis- and Decay

14.1 The dynamic Model

In order to use DTA for a time-resolved analysis of the osmotic stress response (see main text figure 3A and Section 12.2), our static approach (Section 13.1) has to be adapted. The main reason for this is that mRNA levels can no longer be assumed constant, i.e., synthesis and degradation are not necessarily in a dynamic equilibrium. Moreover, we did not measure the unlabeled mRNA fraction for all of the time points. Our aim here was to develop a cost- and time-saving method which works well with only the labeled and the total mRNA fractions. Prior to the actual normalization, all mRNA fractions were rescaled such that $median(T_{gr}) = 1$, to obtain the so called "median centered" data:

$$F_{gi} \text{ was replaced by } F_{gi} / \text{ median}\{T_{gi} \mid g \in G\} \text{ for all } F \in \mathcal{F}^{osmotic}, \tag{43}$$

where the index i indicates the time window $i \in I = \{\text{0-6, 6-12, 12-18, 18-24, 24-30, 30-36 min}\}$ of 4sU-labeling (see main text, Figure 3A). Total least squares regression (the first step in our steady-state DTA estimation procedure (section 13.2)) was not applicable in the context of our osmotic stress experiment (section 12.2). To circumvent this fact, the following approach was used: A set of 480 "stable" genes G^{stable} was defined, whose mRNA stability during the osmotic stress response is considered virtually unaltered. This selection was based upon ranks, since they are more robust than folds. Values of the labeled (L) resp. total (T) fractions were ranked for each of the time windows $i \in I$. Rank gains were calculated relative to the ranks in the $i =$ 0-6 min time window as a reference. 480 genes showed a rank gain below 500 for all arrays and were gathered to build G^{stable}. This cutoff is arbitrary, but it can be chosen even stricter without qualitatively effecting the results (data not shown). As λ_g is a monotonic function of $\frac{A_{gi}}{C_{gi}}$ equation 18, and assuming the decay rate of mRNA g, $g \in G^{stable}$ unchanged, we have

$$\text{median}\left\{\frac{A_{g,i}(t_i)}{C_{g,i}(t_i)} \mid g \in G^{stable}\right\} = \text{median}\left\{\frac{A_{g,i+1}(t_{i+1})}{C_{g,i+1}(t_{i+1})} \mid g \in G^{stable}\right\} \tag{44}$$

According to equation (18), we write (44) as

$$\text{median}\left\{\frac{c_i}{a_i}\frac{L_{g,i}(t_i)}{l_{g,i}T_{g,i}(t_i)} \mid g \in G^{stable}\right\} = \text{median}\left\{\frac{c_{i+1}}{a_{i+1}}\frac{L_{g,i+1}(t_{i+1})}{l_{g,i+1}T_{g,i+1}(t_{i+1})} \mid g \in G^{stable}\right\} \tag{45}$$

If we define

$$median_i^{stable} = \text{median}\left\{\frac{L_{g,i}(t_i)}{l_{g,i}T_{g,i}(t_i)} \mid g \in G^{stable}\right\}, \tag{46}$$

Equation (45) reads as

$$\frac{c_{wt}}{a_{wt}} median_{wt}^{stable} = \frac{c_i}{a_i} median_i^{stable} \quad \text{for all} \quad i \in I \tag{47}$$

Here, the index wt indicates the wild type experiment. The quotient $\frac{c_{wt}}{a_{wt}}$ can be derived from the steady state version of DTA (see Section (13.2), Equation(24)). In this manner we can obtain stable estimates for $\frac{a_i}{c_i}$, $i \in I$. The labeled fraction is then normalized linearly,

$$L_{g,i}^{normalized} = \frac{c_i}{a_i} L_{g,i} \ , g \in G, \ i \in I$$

In contrast to the steady state case of DTA (Section 13.1), the assumption of constant mRNA levels does not hold in the case of a perturbation with a global impact on transcription. Neither the total mRNA amount nor any single mRNA population is growing at a constant rate of α. We can account for this by introducing mRNA-specific and time-dependent "growth rates" $\alpha_{g,i}$, which can also be negative. The gene-specific mRNA growth rates $\alpha_{g,i}$ at time t_i will replace the global mRNA growth rate α which was introduced in equation (1). For the estimation of these growth rates, we first need a stable estimate of the time course of the total mRNA amount of gene g. It is obtained by fitting a cubic smoothing spline to the intensity time course $T_{g,i}$. The resulting spline function is a robust estimate $\widetilde{T}_{g,i}$ of the total mRNA amount at t_i, and it additionally provides an estimate of its derivative $f_{g,i}$ at t_i. Due to the short labeling time, it is justified to model the local dynamic behavior of the total mRNA at a certain time t_i by an exponential function,

$$T_{g,i}(t_i + t) \approx T_{g,i}(t_i) \cdot e^{\alpha_{g,i}} \text{ for small } t; \tag{48}$$

where $T_{g,i}(t_i)$ is the total mRNA amount of gene g in sample i at time t_i. The parameter $\alpha_{g,i}$ is the logarithmic derivative of the right-hand side of Equation (48); consequently, an estimate of $\alpha_{g,i}$ can be obtained from the logarithmic derivative of the spline function at t_i,

$$\alpha_{g,i} = f_{g,i} / \widetilde{T}_{g,i} \ , \tag{49}$$

In analogy to equation (25), we obtain a decay rate estimate

$$\lambda_{g,i} = -\alpha_{g,i} - \frac{1}{t_L} \log \left[1 - \frac{1}{l(p_i, u_g)} \frac{L_{g,i}^{normalized}}{T_{g,i}} \right] \tag{50}$$

Here, $t_L = 6$ min is the labeling period (which was used for all samples in the time series). The synthesis rate is then calculated as in (Section 13.2, Equation 30).

15 Half-life Estimation via Quantitative PCR

15.1 Experimental Design

We used rt-qPCR for the validation of the DTA measurements. The following genes were selected for PCR quantification for the reason indicated behind each gene: act1, tub2 (housekeeping genes), rdn1 (gene with high transcript abundance), ctt1, gpd1, stl1 (known salt stress responders), kss1, sfg1 (genes with very long estimated wild type half-lives). First, we repeated the osmotic stress experiment as described in the methods and took samples of total and labeled mRNA (after a labeling period of 6 min) of the wild type and at $t =$ 12, 30 and 36 min (Supplemental figure S3). qPCR of the selected genes served as a check of the accuracy of the DTA measurements (see figure 5). Secondly, we performed a transcriptional shutoff experiment, adding 1,10-Phenanthroline in the wild type and after 6 min of osmotic stress. The mRNA decay of the selected genes was then determined with qPCR in a classical way by an mRNA decay time series taken at $t = 0$, 2.5, 6, 10, 16 min after transcription shutoff (Supplemental figure S3). Note that the time points of the decay series were determined such that for the bulk of genes that have a half life of about 11 min, the mRNA amounts in the decay time series are approximately equally spaced. This maximizes the estimation accuracy of the method. PCR quantification was always done for two biological samples yielding three technical replicates each.

	12 min	30 min	36 min
L	0.55	0.92	0.94
T	0.45	0.96	0.97

Table 5: Correlations of mRNA folds derived from DTA (x-axis) vs. PCR (y-axis). Upper/lower panel shows the labeled/total fraction respectively. Left/middle/right column corresponds to the 12/30/36 min vs. wild type fold.

15.2 The Decay Model

qPCR experiments determine the abundance of an mRNA by the number of amplification cycles that are needed to reach a certain threshold concentration. Let $ct = ct(g,t,p,r)$ denote the *ct* (cycle time) value of gene $g \in G$ in biological or technical replicate $r \in R$ that has been measured on plate $p \in P$ after a decay time of t. We assume gene-specific exponential mRNA decay rates λ_g. The initial mRNA amount in a sample may vary by a replicate-specific factor α_r, and by a plate-specific factor β_p. Those quantities relate via

$$2^{-ct(g,t,p,r)} = c_0 \cdot e^{-\lambda_g t} \cdot \alpha_r \cdot \beta_p \tag{51}$$

with some unknown constant c_0. Let the data be given by a sequence of tuples $(ct_j, g_j, t_j, p_j, r_j)_{j\in J}$. Assuming a Gaussian error on the *ct*-value measurements, and taking logs in Equation(51), our model can be cast as

$$ct_j = c_0 - t_j\lambda_{g_j} + \alpha_{r_j} + \beta_{p_j} + \epsilon_j \quad \epsilon_j \underset{iid}{\sim} \mathcal{N}(0,\sigma^2) \ , \quad j \in J \tag{52}$$

for some variance σ^2. We seek to find a good parameter fit $\{(\lambda_g)_{g\in G}, (\alpha_r)_{r\in R}, (\beta_p)_{p\in P}\}$ via maximum likelihood estimation, which is equivalent to solving the least squares regression problem

$$ct_j \sim c_0 + \sum_{g\in G}(-t_j\delta_{g=g_j})\cdot\lambda_g + \sum_{r\in R}\delta_{r=r_j}\cdot\alpha_r + \sum_{p\in P}\delta_{p=p_j}\cdot\beta_p \ , \quad j \in J \tag{53}$$

Note that this system is underdetermined, and each of the three parameter sets (λ_g),(β_r),(α_p) can only be determined up to some additive constant. As a consequence, it is in principle not possible to determine the absolute decay rates with this approach, however it is well suited for the estimation of relative decay rates $(\lambda_{g_1} - \lambda_{g_2})$. Usually, each PCR plate contains control RNAs of known concentration that capture plate effects, but are neither subject to decay nor does their amount vary between replicates. In our setting, it is easy to incorporate the control measurements $(ct_{k,p_k})\, k \in K$ into the model by adding

$$ct_k \sim c_0 + \sum_{p\in P}\delta_{p=p_k}\cdot\beta_p \ , \quad k \in K \ , \tag{54}$$

to the regression problem, one for each control measurement (ct_k, p_k), $k \in K$.

15.3 Results of the PCR experiments, Comparison to DTA

In a first simulation study, we convinced ourselves that the procedure described in (Section 15.2) is able to recover the decay rates properly, up to some additive constant (Figure20). We then applied the method to the rt-qPCR measurements (Figure 21). It is possible to treat the *ct*-values of the PCR experiment like the (logarithmic) microarray measurements of the unlabeled mRNA fraction, by swapping signs and adding an arbitrary constant to rescale all values to a positive range. Relevant are only relative differences between two values (this corresponds to their ratio in absolute scale), because absolute values underlie amplification steps in the biochemical protocol and can differ by unknown factors. The next section will show these results and compare them to the corresponding PCR results.
In order to do so, we rescaled the mean half-life of the gene set used for PCR to a common value. The amount of labeled and total mRNA as quantified by both methods agreed quite well (Table 5), as well as the estimated decay rates for the wild type (Figure 22). The decay rates at $t = 12$ min showed only weak correlation. Since the PCR measurements are obtained by an invasive method, we do not expect them to agree better than the decay rates from the literature (Supplemental figure S5), which also correlated poorly.

Note that the absolute decay rates cannot be estimated from the PCR profiles, therefore some mRNA amounts are seemingly increasing in the plot (they are assigned negative decay rates).

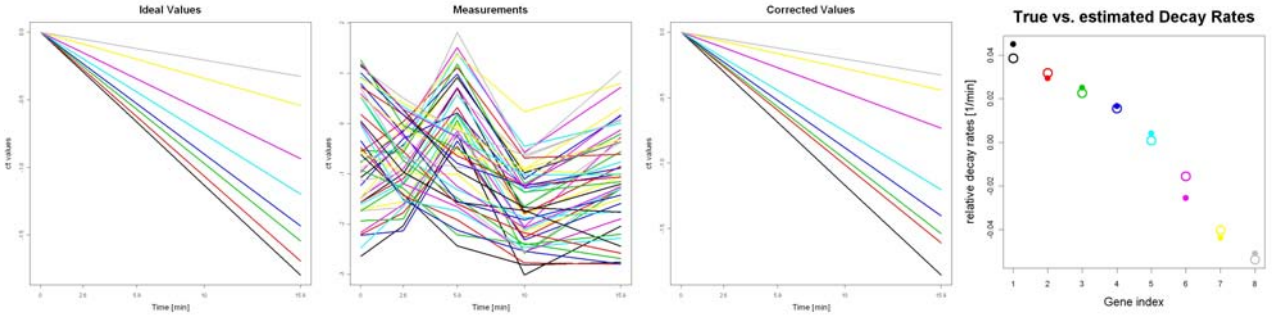


Figure 20: Results of the simulation study. The first plot in the series shows ideal mRNA decay time courses of 8 "genes" with different decay rates (one colored line corresponds to one gene. Data is displayed on a half-logarithmic scale. x-axis: time, y-axis: log(mRNA amount)). The second plot shows realistic data on the same scale as the first picture, including measurement noise and plate effects. The third plot shows the decay profiles as they have been reconstructed by our method. Normalizing both the true and the estimated decay profiles to the same mean decay rate, the fourth plot demonstrates that the estimation error is small (circles: true decay rates, dots: estimated decay rates).

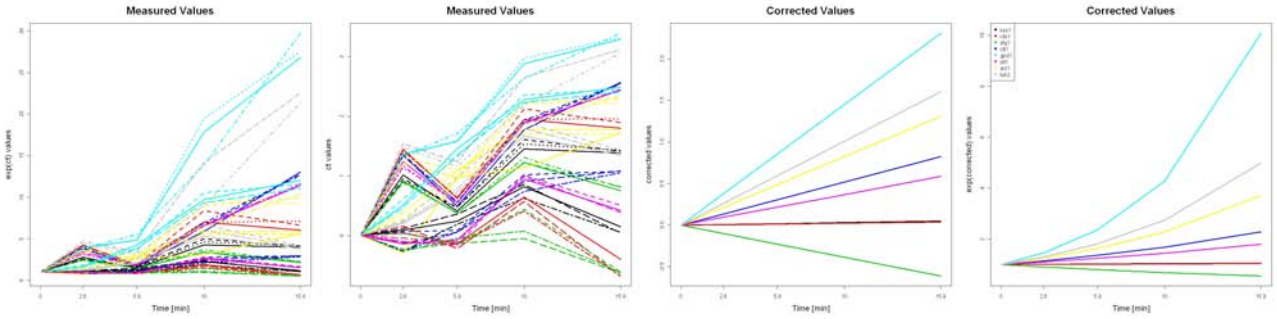


Figure 21: Results of the qPCR experiment. Time courses of the 8 genes selected for PCR (each color corresponds to one gene) were measured in 2 biological samples, each of which was measured in triplicate. The two leftmost plots show the raw data as in Figure20 (on the absolute resp. the half-logarithmic scale). The two rightmost plots are the fitted decay profiles, derived from the estimated decay rates (on the half-logarithmic resp. absolute scale).

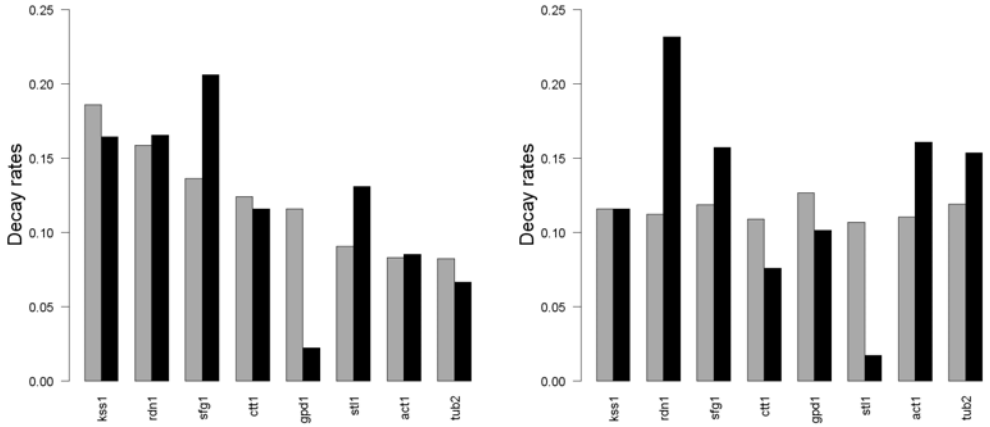


Figure 22: Bar plots of the DTA (grey) and PCR (black) decay rate estimates, obtained as described in Section 13.2(DTA) resp. Section 15.2(PCR). The left plot shows the wild type, the right plot shows the situation after 12 min of osmotic stress.

Part II
Dynamics of Polymerase II binding and its relation to mRNA Synthesis during osmotic Stress

16 Pol II ChIP-chip

16.1 Pol II ChIP-chip experiment

For genomic occupancy profiling by ChIP-chip we used S. cerevisiae strain BY4741 containing a TAP tag on the Pol II subunit Rpb3. Yeast cells were grown in YPD medium until exponential phase (OD600 ~ 0.8). ChIP-chip was performed for biological replicates at timepoint 0 (taken from [Mayer *et al.*, 2010]) and at 12 and 24 min after the addition of 0.8 M NaCl. Additionally chromatin samples, where the chromatin immunoprecipitation step was omitted (termed genomic Input), were subjected to array analysis and served as a reference. All DNA extracts were hybridized to high-density custom-made tiling array (Affymetrix). A listing of all produced arrays can be found in (Supplementary Table T3).

16.2 Preprocessing of ChIP-chip data

The high-density custom-made tiling arrays used in the experiments contain approx. 3 million perfect match and mismatch probe pairs tiled through the complete yeast genome at a 4 bp tiling resolution. Only perfect match probes were used for normalization and analysis. The Bioconductor package Starr [Gentleman *et al.*, 2004, Zacher *et al.*, 2010] was used for data read-in, processing and all further analyses. Quality assessment of each measured array was done by inspection of raw image, density-plots, boxplots, scatter-plots and MA-plots in order to avoid processing awed arrays. This confirmed that there were no manufacturing defects, etc. (images not shown). It revealed that the first replicate of 12 min measurements showed a lower intensity range than the other replicates. The scatterplots and pairwise correlations (Figure 23) however revealed, that apart from a global intensity shift this array did not behave qualitatively different than the other arrays. We therefore included all arrays into the subsequent analysis steps. All time points were log2 transformed and spererately normalized using loess normalization [Yang *et al.*, 2002]. The median over the replicates was calculated and the genomic input sample was substracted to remove local genomic and sequence-dependent bias. To ensure comparability of the three time points regarding the absolute scale, the normalized ratios were subsequently scale normalized to a common median intensity level. The Median Absolute Deviations (MADs) for the three time points were calculated to 1.03, 0.96, 1.00 for time points 0, 12, 24 respectively, and give a robust measure of the variability of the underlying ratio distributions [Smyth & Speed, 2003].
Initially, we excluded all genes, which are annotated as dubious or silenced by the Stanford Public Database (SGD™: Saccharomyces Genome Database) [Cherry *et al.*, 1998], and only kept those annotated as verified or uncharacterized (5743 genes). To align gene profiles across entire transcripts, only genes with available TSS (transcription start site) and pA (poly(A) site) assignments from RNA-seq experiments by [Nagalakshmi *et al.*, 2008] were taken into account (4366 genes). Genes with TSS (pA) measurements downstream (upstream) of the annotated ATG (Stop) codon were excluded. To remove possible wrongly annotated TSSs and pAs, we only included genes with TSS (pA) annotations showing a distance less than 200 bp to the corresponding downstream (upstream) ATG (Stop) codon (3465 genes). Furthermore, we restricted our analysis to transcripts with a length of more than 1000 bp. This leaves 2407 genes to build our geneset of interest G^{chip}. The mean profiles for each transcript were calculated as described in the following. First, the profiles were aligned from 500 bp upstream to 500 bp downstream at each TSS, resp. pA. The region between TSS and pA of the transcripts were scaled to their median length:
Let len be the median length of all these regions, p the vector of intensities at each position of length l, p^* the scaled vector and p^*_i the i^{th} position in p^*. Then p^*_i is defined as

$$p^*_i = \begin{cases} mean\{p_j \mid \lceil \frac{l\,j}{len} \rceil = i\} & if\ l > len \\ p_{round(1+\frac{(i-1)\,(l-1)}{(len-1)})} & else \end{cases}.$$

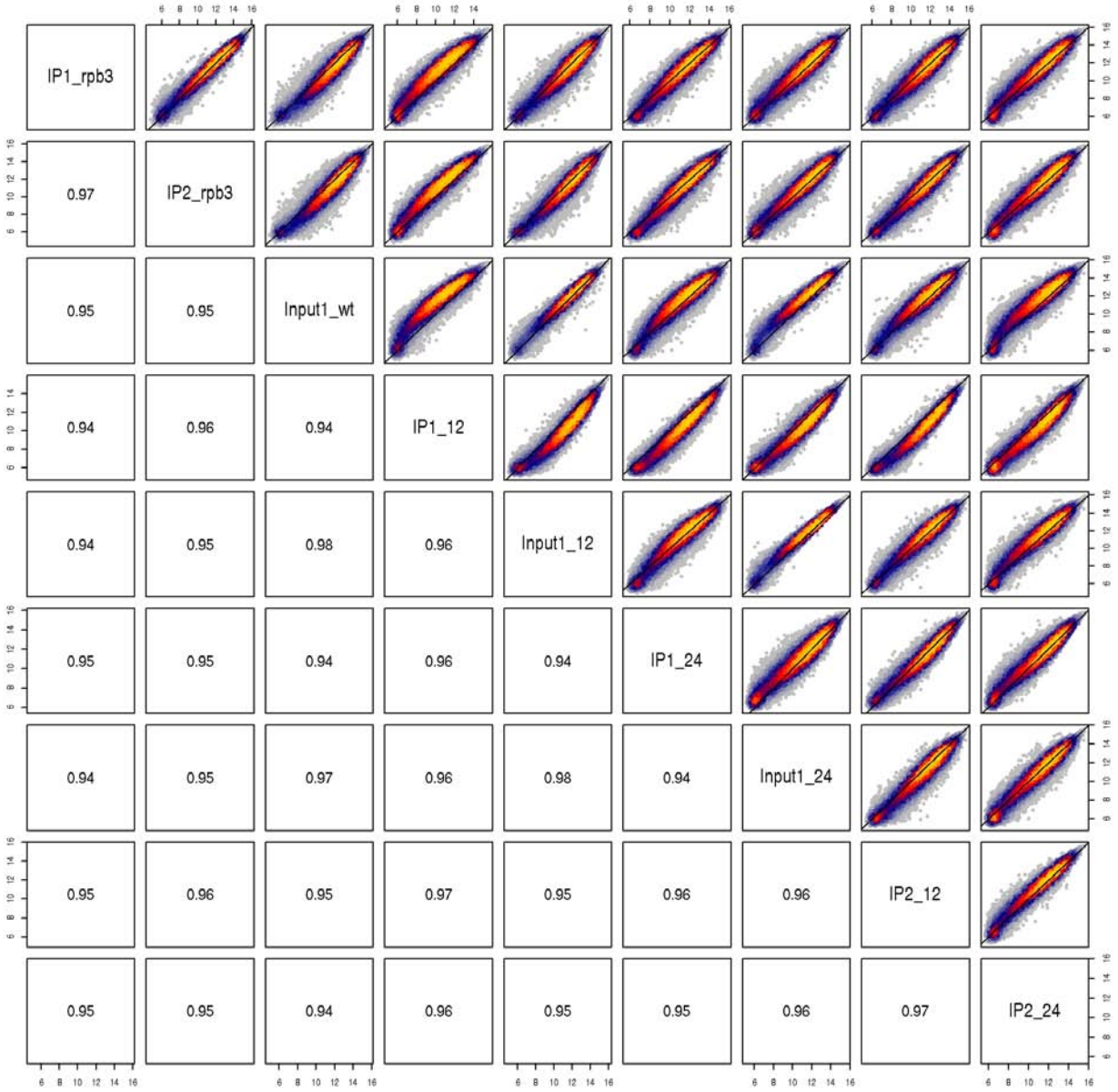


Figure 23: Pairwise scatterplots of the probe intensities. The lower panel shows the respective pearson correlations.

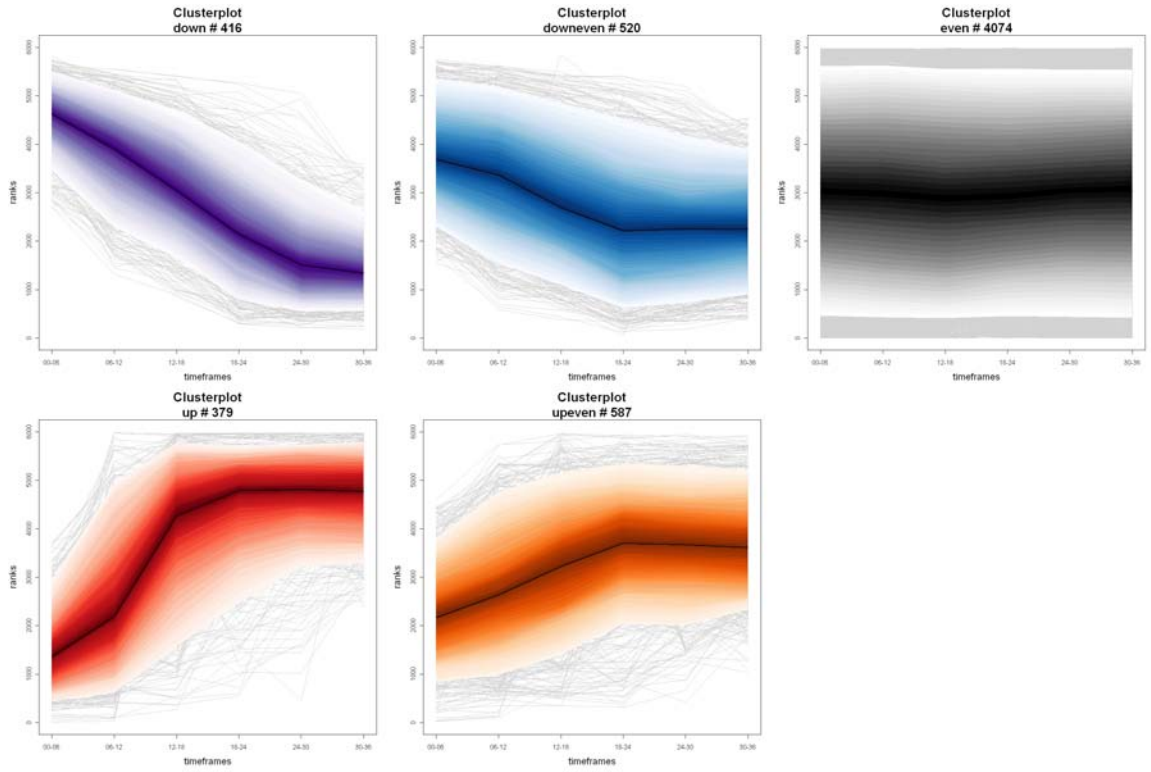


Figure 24: Figure shows rank behavior over time of those previously chosen clusters. The numbers above the plots give the size of the shown clusters. Plots were produced with the R package LSD [Schwalb *et al.*, 2010].

17 Rank-cluster selection

Apart from the global transcriptional shut off in the first minutes of the osmotic stress response (probably due to an immediate dissociation of most proteins from the chromatin by means of increased ion concentrations [Proft & Struhl, 2004]), one can observe gene regulation patterns consisting of induction resp. repression of certain transcripts compared to the whole transcriptome. This provides a view independent of a reaction of unprepared cells, which are thought to be initially unresponsive due to transcription. This can be addressed by considering the internal ranking of newly synthesized mRNAs over time to get an idea of "intra-differential" expressed transcripts (differential behavior referring to the bulk of "unchanged" genes), and hence regulation that is independent of changes in the global level. This whole procedure is based upon ranks, since they show much more robustness than typical folds, and are thus much more sensitive in terms of detecting significantly changed expression. To find these "intra-differential" transcripts, synthesis rates were ranked for each of the time windows 0-6, 6-12, 12-18, 18-24, 24-30 and 30-36 minutes. This whole approach is similar to the assumption that most genes do not respond to the stimulus that is set. rank gains are calculated to the 0-6 minutes time window as a reference.
Clusters are chosen due to the rank gains of the transcripts seen in the last timeframe, namely 30-36 minutes. The "up"-cluster contains transcript that show a rank gain of more than 2000 (corresponding to a 35% difference in rank percentiles; 2000 of 5743 genes). The "up-even"-cluster contains transcript that show a rank gain between 1000 and 2000. The "even"-cluster contains transcript that show a rank gain between -1000 and 1000. The "down-even"-cluster contains transcript that show a rank gain between -2000 and -1000. And the "down"-cluster contains transcript that show a rank gain lower than -2000. Genes in each cluster thus show similar behavior due to "intra-differential" expression. We use rank differences to detect really exceedingly strong effects, namely rank differences that span at least 35% of the rank percentiles of the whole distribution ("up"- and "down"-cluster). This means that any gene considered relevantly altered by us has to "jump over" the central region in the distribution (Figure 24).
Each of these clusters shows a specific behavior. It can be claimed that genes that belong to the "up"-cluster are strongly induced, since they are almost not expressed at the beginning and are among those genes at the end that are expressed the most. Vice versa for genes that belong to the "down"-cluster. Genes that belong

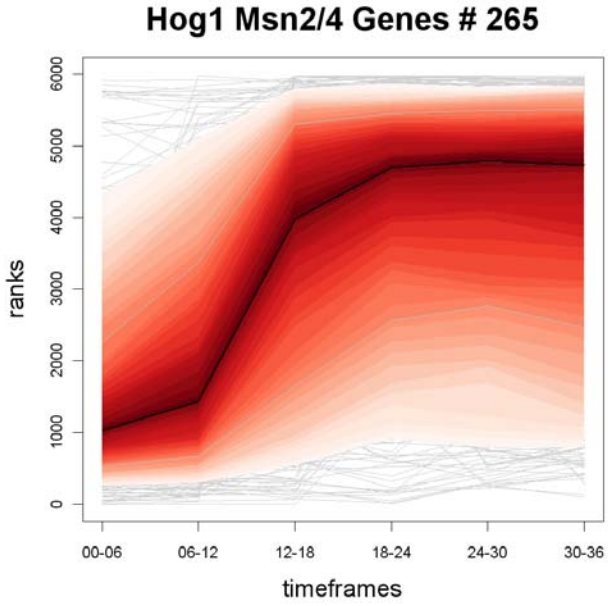


Figure 25: 265 Hog1 and/or Msn2/4 dependent genes observed by [Capaldi *et al.*, 2008] to be up-regulated more than 1.5 fold after 20 min. treatment of 0.4 M KCL. Plot was produced with the R package LSD [Schwalb *et al.*, 2010].

to the "up-even" or "down-even" -cluster are slightly induced or repressed respectively. Genes that belong to the "even"-cluster show no "intra-differential" behavior. To assess the rank variability, we calculated a MSD (mean standard deviation). Since this amounts to approximately 117 ranks among (wild type) replicate measurements, our previous cutoffs for cluster selection are justified. We further verified our calculated rank gains by highlighting known genesets from the literature to confirm their expected behavior during the osmotic stress response (most figures not shown) (Figure 25), and also compared them to typical folds (see main text figure 3C).

18 Correlation to Pol II ChIP-Chip data obtained under osmotic stress conditions

Clusters were correlated to Pol II occupancy obtained by our ChIP-chip experiment (see section 16.1) under wild-type and osmotic stress conditions (Figure 26). The data shown is the genewise mean from start to stop codon, processed by means of the Bioconductor Starr package [Gentleman *et al.*, 2004], see (section 16.2). Genes that were selected for gene profile alignment G^{chip} share 111, 242, 1534, 272 and 248 genes with the up , up-even, even, down-even and down cluster respectively.

19 Correlation to Proteomics

Clusters were correlated to quantitative proteomics data obtained under wild-type and osmotic stress conditions [Soufi *et al.*, 2009]. Although the overall relationship to our selected clusters is only slightly, there is an excellent correlation between the subset of osmotic stress up-regulated mRNAs and their corresponding Protein changes. This poor overall relationship may be attributed to the "uncoupled worlds" of transcription and translation (Figure 27).

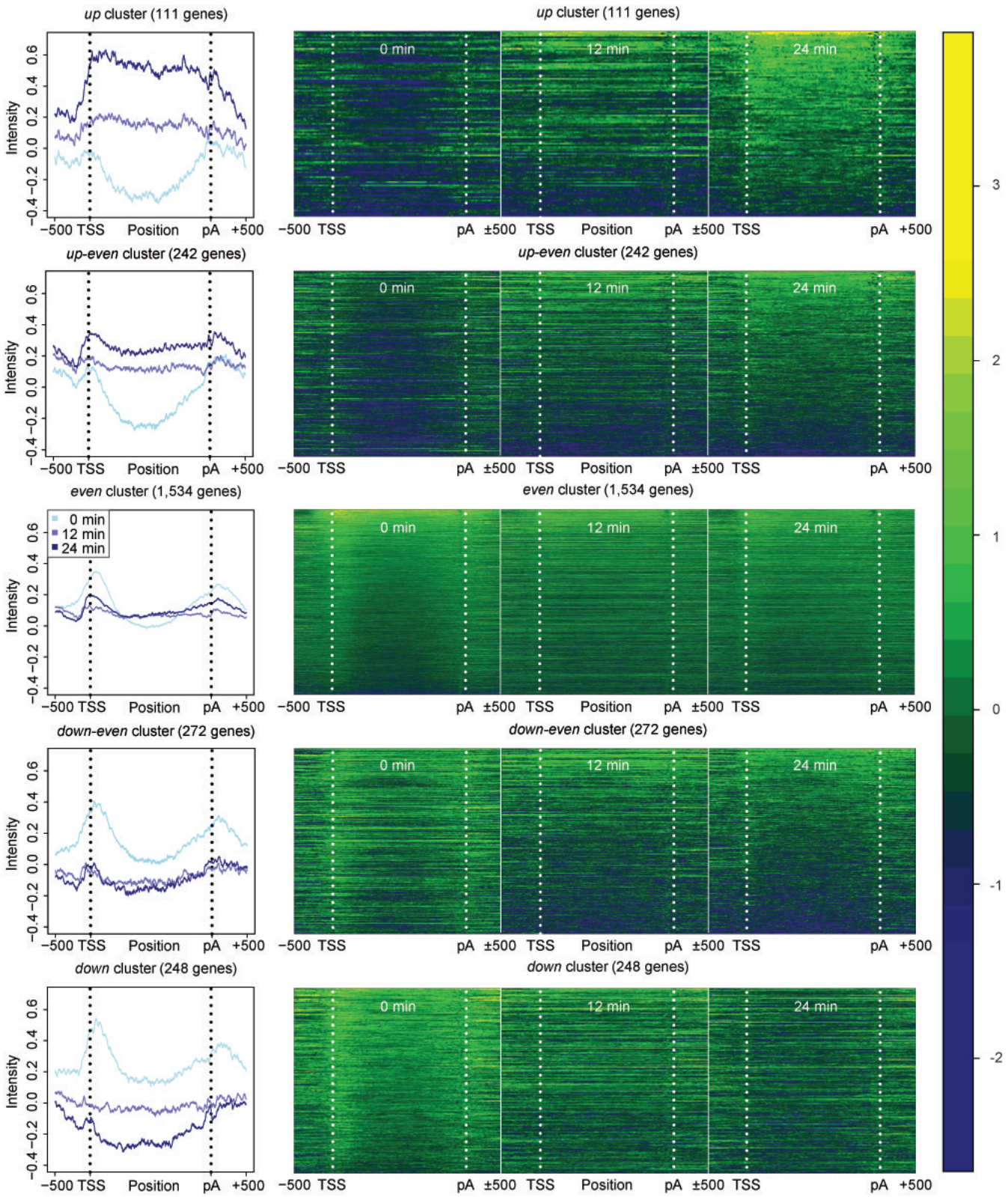


Figure 26: Left: Mean Pol II occupancy profiles of all selected clusters (see section 16.2,17). Profiles are obtained after 0, 12 and 24 min of osmotic stress (light blue, blue, dark blue lines). Vertical dotted lines are drawn at the TSS and the pA site. Right: Heatmaps of the Pol II profiles for all cluster at 0, 12, and 24 min. Each row corresponds to one gene. The vertical dotted lines mark TSS and pA of each gene. Pol II occupancy from low to high is coded with colors ranging from dark to bright.

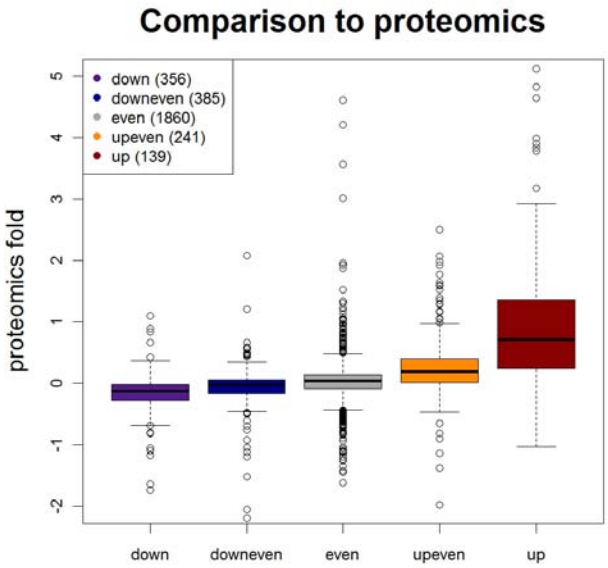


Figure 27: The data shown are the log2(folds) of the proteome after 20 min of osmotic stress. Each cluster is depicted as a boxplot (only the common genes are depicted). Colors correspond to initial color scheme.

20 GO enrichment analysis of selected clusters

GO enrichment analysis was performed with the R Bioconductor package "GOstats" [Falcon & Gentleman, 2007]. All yeast orfs were chosen as the "Gene Universe". This is the population (the urn) of the Hypergeometric test, which yields p-values and odd-ratios for overrepresentation of each GO term in the specified category among the GO annotations for the interesting genes. In this case, the selected clusters.

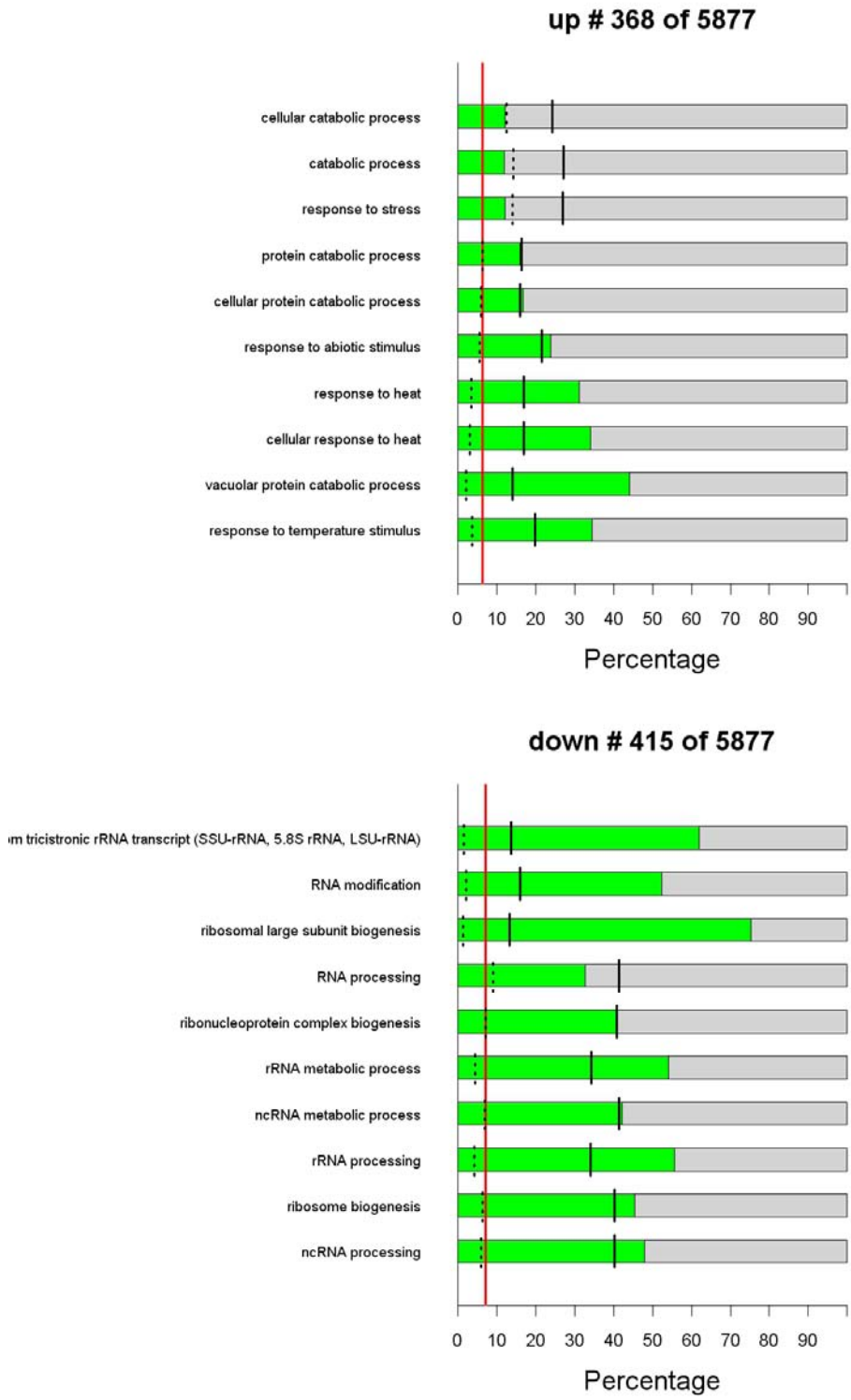


Figure 28: Top 10 GO categories of "up" and "down" mRNAs. Sorted by p-values. GO category annotations can be accessed in supplementary tables (6,7). Grey bars depict the size of each GO category scaled to 100%. The red line indicates the percental proportion of each category that is expected by chance. Green bars show the actual enrichment of each category in percent. Black lines indicate the number of enriched genes (green bars) in relation to the number of tested genes (# 368 resp. 415). Dashed lines show the size of each category in proportion to the whole number of genes (Gene Universe #5877).

	GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
1	GO:0009266	8.45e-37	9.56	13.275	73	212	response to temperature stimulus
2	GO:0007039	2.17e-32	13.57	7.389	52	118	vacuolar protein catabolic process
3	GO:0034605	7.37e-31	9.10	11.396	62	182	cellular response to heat
4	GO:0009408	2.33e-28	7.94	12.461	62	199	response to heat
5	GO:0009628	1.63e-27	5.73	20.664	79	330	response to abiotic stimulus
6	GO:0044257	9.73e-13	3.37	22.166	59	354	cellular protein catabolic process
7	GO:0030163	4.76e-12	3.19	23.607	60	377	protein catabolic process
8	GO:0006950	1.21e-11	2.43	51.534	99	823	response to stress
9	GO:0009056	1.73e-11	2.40	52.598	100	840	catabolic process
10	GO:0044248	1.82e-10	2.38	46.274	89	739	cellular catabolic process
11	GO:0050896	1.13e-09	2.09	71.008	118	1134	response to stimulus
12	GO:0033554	1.26e-09	2.46	36.067	73	576	cellular response to stress
13	GO:0044265	1.17e-08	2.48	29.931	62	478	cellular macromolecule catabolic process
14	GO:0005996	1.28e-08	3.89	9.580	30	153	monosaccharide metabolic process
15	GO:0044282	1.66e-08	3.94	9.142	29	146	small molecule catabolic process
16	GO:0006066	3.04e-08	2.99	16.468	41	263	alcohol metabolic process
17	GO:0044275	3.25e-08	5.15	5.322	21	85	cellular carbohydrate catabolic process
18	GO:0009057	6.34e-08	2.33	31.997	63	511	macromolecule catabolic process
19	GO:0006914	1.37e-07	3.71	8.892	27	142	autophagy
20	GO:0016052	1.74e-07	4.57	5.823	21	93	carbohydrate catabolic process
21	GO:0051716	2.13e-07	2.10	42.141	75	673	cellular response to stimulus
22	GO:0006006	2.61e-07	3.93	7.514	24	120	glucose metabolic process
23	GO:0046164	2.63e-07	5.20	4.508	18	72	alcohol catabolic process
24	GO:0044262	3.96e-07	2.57	20.037	44	320	cellular carbohydrate metabolic process
25	GO:0046365	4.52e-07	5.29	4.195	17	67	monosaccharide catabolic process
26	GO:0019323	9.38e-07	Inf	0.313	5	5	pentose catabolic process
27	GO:0019318	1.21e-06	3.45	8.704	25	139	hexose metabolic process
28	GO:0005975	1.30e-06	2.40	22.292	46	356	carbohydrate metabolic process
29	GO:0042732	5.34e-06	75.87	0.376	5	6	D-xylose metabolic process
30	GO:0009065	8.34e-06	15.24	0.877	7	14	glutamine family amino acid catabolic process

Table 6: GO enrichment analysis of "up" mRNAs. Sorted by p-values.

	GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
1	GO:0034470	6.92e-109	19.54	24.644	167	349	ncRNA processing
2	GO:0042254	2.44e-104	17.72	25.915	167	367	ribosome biogenesis
3	GO:0006364	2.74e-102	24.58	17.865	141	253	rRNA processing
4	GO:0034660	5.14e-101	15.67	28.811	172	408	ncRNA metabolic process
5	GO:0016072	9.47e-101	23.16	18.501	142	262	rRNA metabolic process
6	GO:0022613	8.73e-96	14.50	29.376	169	416	ribonucleoprotein complex biogenesis
7	GO:0006396	8.64e-80	10.24	37.072	172	525	RNA processing
8	GO:0016070	8.50e-58	5.56	83.042	224	1176	RNA metabolic process
9	GO:0042273	2.64e-49	46.21	5.155	55	73	ribosomal large subunit biogenesis
10	GO:0009451	8.62e-44	17.03	8.897	66	126	RNA modification
11	GO:0000462	1.89e-43	24.69	6.497	57	92	maturation of SSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)
12	GO:0030490	1.09e-42	23.34	6.638	57	94	maturation of SSU-rRNA
13	GO:0044085	1.61e-42	4.70	64.188	175	909	cellular component biogenesis
14	GO:0006807	2.38e-41	4.06	146.101	276	2069	nitrogen compound metabolic process
15	GO:0006139	2.35e-40	3.97	126.823	253	1796	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process
16	GO:0000460	5.25e-35	25.43	5.014	45	71	maturation of 5.8S rRNA
17	GO:0000466	5.25e-35	25.43	5.014	45	71	maturation of 5.8S rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)
18	GO:0010467	6.21e-32	3.37	130.283	242	1845	gene expression
19	GO:0000447	4.73e-27	42.48	2.825	30	40	endonucleolytic cleavage in ITS1 to separate SSU-rRNA from 5.8S rRNA and LSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)
20	GO:0000967	1.18e-26	91.20	2.118	26	30	rRNA 5'-end processing
21	GO:0034471	1.18e-26	91.20	2.118	26	30	ncRNA 5'-end processing
22	GO:0000469	1.46e-26	20.66	4.307	36	61	cleavages during rRNA processing
23	GO:0000478	5.41e-26	35.39	2.966	30	42	endonucleolytic cleavages during rRNA processing
24	GO:0000479	5.41e-26	35.39	2.966	30	42	endonucleolytic cleavage of tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)
25	GO:0000966	6.83e-26	72.95	2.189	26	31	RNA 5'-end processing
26	GO:0000472	1.53e-25	87.47	2.048	25	29	endonucleolytic cleavage to generate mature 5'-end of SSU-rRNA from (SSU-rRNA, 5.8S rRNA, LSU-rRNA)
27	GO:0000480	3.01e-25	111.69	1.907	24	27	endonucleolytic cleavage in 5'-ETS of tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)
28	GO:0042255	7.89e-23	14.70	4.872	35	69	ribosome assembly
29	GO:0042274	2.61e-20	17.02	3.743	29	53	ribosomal small subunit biogenesis
30	GO:0000027	5.01e-20	24.95	2.754	25	39	ribosomal large subunit assembly

Table 7: GO enrichment analysis of "down" mRNAs. Sorted by p-values.

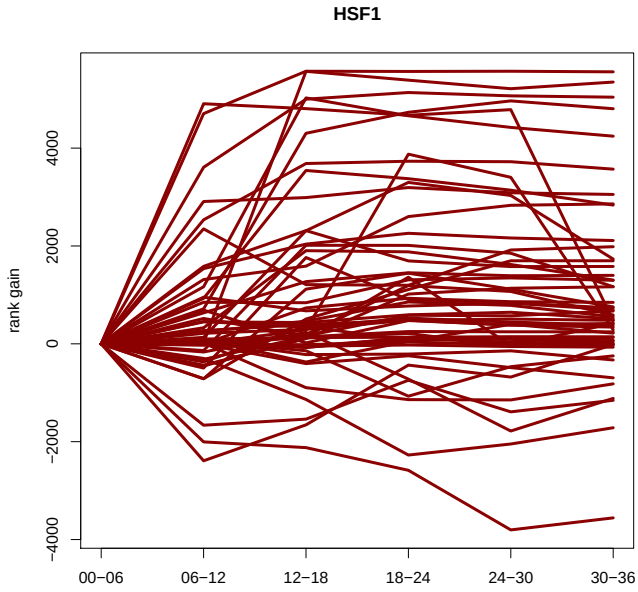


Figure 29: rank gain of target genes of TF HSF1. We observe three typical trends: 1) induced genes 2) repressed genes 3) genes whose expression does not change over time.

Part III
Gene regulation during osmotic stress

21 Transcription factor dynamics in osmotic stress

The majority of transcription factors (TFs) are low abundance transcripts (see main text figure 2B), and their transcription rates do not behave extraordinarily in osmotic stress compared to the overall transcription rates (Figure 30). This suggests that most TFs are regulated on the signaling level, by protein modifications, and not on the transcription level. This is not surprising, since we record the initial phases of the stress response. Such a fast response cannot be established by transcriptional regulation of TFs. It is much more likely that TFs reside in the cytoplasm in their inactive form until they are converted, e.g. by Hog1, into their active counterpart. It is therefore rarely possible to directly detect active TFs by their own expression rather than by the expression of their target genes.
Likewise the expression of the target genes of a TF do not show a common trend. Rather there are some genes that are induced, some repressed and some that do not show any major changes in expression. This leads us to conjecture that TF-TF interaction play a crucial role. (Figure 29) shows this phenomenon for the target genes of HSF1, a TF that is active during stress response (see main text figure 4A).
In order to discover the main players of transcriptional regulation during osmotic stress we first need to define a TF-target relation. There are several databases which offer more or less comprehensive lists of TF-target pairs or motifs [Teixeira *et al.*, 2006, Matys *et al.*, 2006, Bryne *et al.*, 2008]. It is important for our purposes to have a reliable list with a low number of false positive relations. Therefore, we decided to use the TF-target relation provided by [MacIsaac *et al.*, 2006]. TFs with less than 20 targets were excluded from subsequent analysis steps (84 TFs remaining). Among the set G of all genes, let G^{up} resp. G^{down} be the set of genes that are induced resp. repressed during osmotic stress, i.e. the genes in the *up* resp. *down* cluster (see section 17). We used an exact Fisher test for assessing whether the set of induced/repressed genes is significantly enriched in the target set of a certain TF, relative to the set of non-targets (see main text figure 4A). To correct for multiple testing we applied a Benjamini-Hochberg correction, selecting adjusted p-values below 5%, i.e., controlling the familywise error rate at the level of 5% [Benjamini & Hochberg, 1995]. This means that less than 5% of the rejected hypotheses are wrong, which in our case (9 TF) means less than one.
It is noteworthy, that even for the most evidently active TFs, only a fraction of the target genes were induced

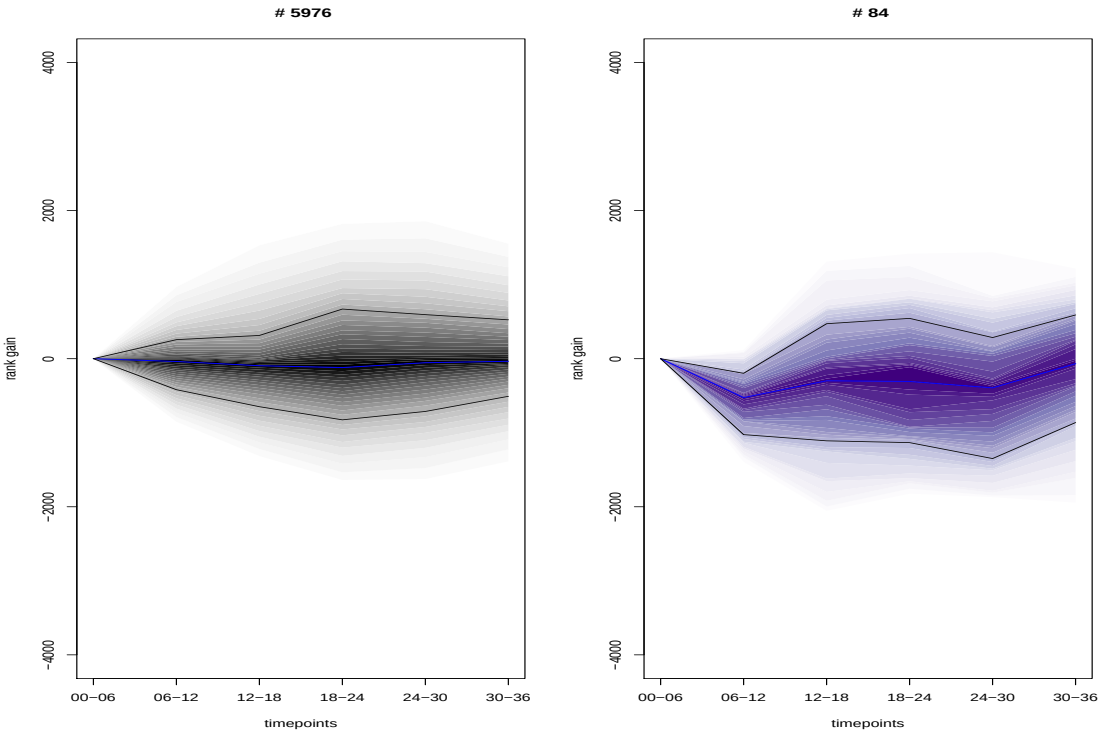


Figure 30: Clusterplot of the rank gain of all genes (5976 genes in grey, left) and genes coding for TF factors (84 genes in purple, right). Blue line indicates median rank gain, black lines show quartiles of the distribution, shown is the distribution between the 95% and 5% quantiles.

resp. repressed. Therefore, the medians of the TF targets' synthesis rates do not change considerably with time. Instead, a plot of the 90% and the 10% quantiles reveals activatory and repressive TFs (see main text figure 4B). Several possible explanations of this phenomenon need to be discussed: 1) There is a tremendous amount of false positive TF-target annotations in the MacIsaacs data set. 2) Different targets of one TF have different response characteristics, e.g., different (active) TF concentration thresholds for being switched on. 3) TF interactions, either cooperative or antagonistic, play a crucial role in target gene regulation.
ad 1): We consider the TF-target predictions of [MacIsaac *et al.*, 2006] reliable, as they are based on high quality ChIP-chip experiments [Harbison *et al.*, 2004] in different conditions and are combined in a stringent and conservative way. ad 2): Although various targets may have various response characteristics, this is unlikely to play such a dominant role in the early salt stress reaction. It is known that the dynamics of the Hog1 stress response varies with salt concentration [Muzzey *et al.*, 2009], hence the level of TF activation is likely to vary as well. Nevertheless, experiments at salt concentrations other than 0.8 M have shown a very similar behaviour concerning the set of regulated target genes (see, e.g., our comparison to the [Capaldi *et al.*, 2008] data, figure 25). To us, interactions between TF are the most plausible explanation for the phenomenon of non-responding target genes. The next section will therefore be devoted to the derivation of methods for assessing TF interactions.

22 Modeling of transcription factor interactions

22.1 General approach to modeling genetic interactions

Let G be the set of all genes for which our measurements are available. Traditional interaction screens measure the effects β_1,β_2 (e.g. growth relative to wild type) of two interventions at $T_1 \in G$ resp. $T_2 \in G$, say. A function $f = f(\beta_1, \beta_2)$ is assumed which predicts the effects of the combined intervention in T_1 and T_2 from the observed single interventions. Having measured the actual effect e_{12} of the combined intervention, the difference between measured and predicted effect, $\beta_{12} = e_{12} - f(\beta_1, \beta_2)$, is defined as the interaction effect. For a large set $T \subseteq G$ of genes, (the majority of) all pairwise interactions are measured and typically

displayed in a heatmap of the resulting $T \times T$ interaction matrix. Since there are several plausible choices for the function f [Mani *et al.*, 2008], which do affect the outcome, and since interaction scores are prone to large variation, individual interaction scores are generally not very reliable. However, two interventions T_1 and T_2 may be grouped together by similarity of their interaction profiles, i.e. by the vectors $S_j = (s_{jx})_{x \in T}$, $j = 1, 2$. Similarity is typically measured by correlation distance, $similarity(T_1, T_2) = 1 - corr(S_1, S_2)$. The products of the genes T_1, T_2 that are similar with respect to correlation distance are very likely to interact either physically or functionally [Krogan *et al.*, 2006, Pan *et al.*, 2006].

Our aim is to asses interactions of transcription factors, so let T be a set of transcription factors. The data consists of wild type and osmotic stress expression measurements (sections 12.1, 12.2). Let $TF1, TF2 \in T$, throughout the rest of this paper, denote their target sets resp. by $M_1, M_2 \subseteq G$. We do not perform single/double interventions at $TF1, TF2$, thus the situation is somewhat different from that described above. Assume that the activity of $TF1$ and $TF2$ has changed between the two environmental conditions. Let $\overline{M_j} = G \setminus M_j$, $j = 1, 2$. The set of all genes can be partitioned disjointly into the four sets

$$G = (M_1 \cap M_2) \cup (M_1 \cap \overline{M_2}) \cup (\overline{M_1} \cap M_2) \cup (\overline{M_1} \cap \overline{M_2})$$

The genes in $M_1 \cap M_2$ are affected by both activity changes, whereas the genes in $M_1 \cap \overline{M_2}$ resp. $\overline{M_1} \cap M_2$ are affected by $TF1$ resp. $TF2$ only. Finally, genes in $\overline{M_1} \cap \overline{M_2}$ are affected by neither of the two TFs. Hence these four gene sets can be used to separate baseline effects from individual effects of $TF1, TF2$, and from interaction effects of both. The next two sections will introduce two conceptually distinct approaches to quantify transcription factor interaction effects. It is a major success of our strategy that both methods agree remarkably well in terms of their profile similarity-based interaction predictions.

22.2 Logistic Regression Model

For each pair of transcription factors $TF1$ and $TF2$ with their target sets M_1 and M_2 respectively, we fit a logistic regression of the form:

$$\log\left(\frac{\text{P(g is induced)}}{\text{P(g is not induced)}}\right) \sim \beta_1 Ind(g \in M_1) + \beta_2 Ind(g \in M_2) + \beta_{12} Ind(g \in M_1 \cap M_2)\ , \quad g \in G$$

Here, $Ind(.)$ is the indicator function with values in $\{0, 1\}$. The response term is the log-odds for a gene g to be differentially expressed, $\beta_1$and β_2 measure the first-order effects while β_{12} is the coefficient of the interaction term. The response term is obtained by calling the *eBayes* function from the R package *limma* [Smyth, 2004]. As we did not have any replicates we used the last two time points from the osmotic stress experiment (synthesis rates) (section 12.2). This procedure is justified, because those time points are highly correlated (Pearson correlation $r = 0.95$).
It is also possible to use the derived value

$$\tilde{\beta}_{12} = 2 \cdot sign(\beta_{12}) \cdot (0.5 - pvalue(\beta_{12}))_+ \qquad \text{, with } x_+ = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{else} \end{cases}$$

as an interaction measure. This has the advantage of constraining the range of values to the range $[-1; 1]$ and, emphasizing significant p-values, leads to more stable results. The matrix obtained with this interaction measure is shown in (Figure 31).

22.3 Odds Ratio Ratio (ORR) model

For two transcription factors $TF1, TF2 \in T$ and any set of genes $S \subseteq G$, the enrichment of common targets of $TF1$ and $TF2$ in S can be measured with the help of the odds ratio (=determinant) OR_S of the 2×2 contingency table

$$\begin{pmatrix} |M_1 \cap M_2 \cap S| & |M_1 \cap \overline{M_2} \cap S| \\ |\overline{M_1} \cap M_2 \cap S| & |\overline{M_1} \cap \overline{M_2} \cap S| \end{pmatrix}$$

We expect cooperative transcription factors to have their common targets enriched in the set G^{up} of induced genes, compared to the enrichment in the set of non-induced genes, $G \setminus G^{up}$. The log-ratio of the odds-ratios $ORR = \log \frac{OR_{G^{up}}}{OR_{G \setminus G^{up}}}$ constitutes our interaction measure.
Since generally $|G^{up}| \ll |G \setminus G^{up}|$, we add pseudocounts to the contingency table by adding the second row to the first, after scaling with a factor α and adding the second column to the first likewise. Both odds-ratios are biased towards 1 with this procedure. We found $\alpha = 0.001$ to yield the best results.

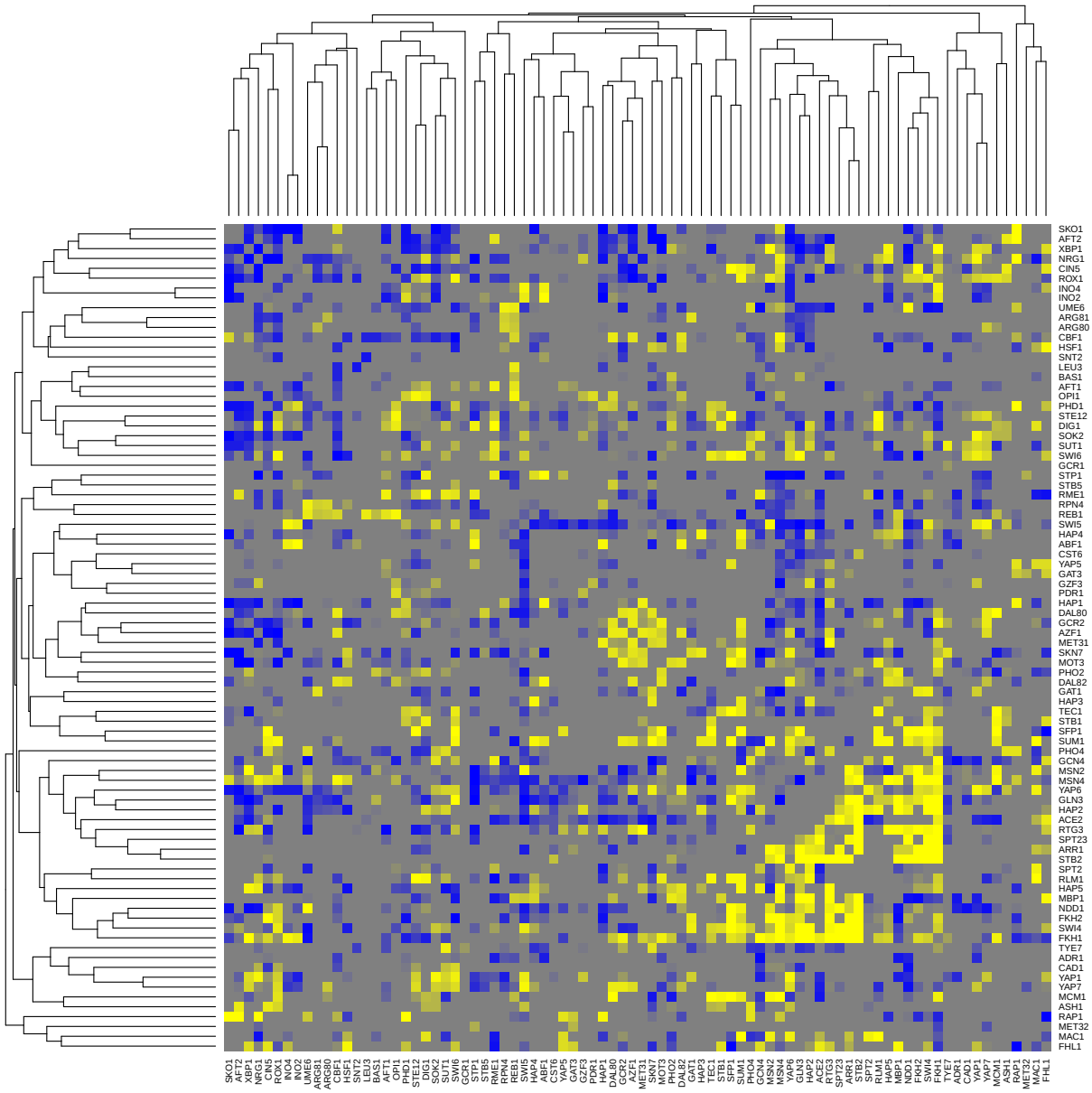


Figure 31: Synthetic interaction screen using the modified interaction term $\tilde{\beta}_{12}$ from the logistic regression as interaction term.

22.4 Quality control

We made sure that the effects we are seeing are not due to the target annotation. (Figure 32) shows a heatmap of the pairwise intersect sizes from our target annotation. The pairwise TF similarities calculated as the intersect sizes of their target sets yields a TF clustering which is completely different from that obtained in the synthetic interaction screen (see Figure 31).

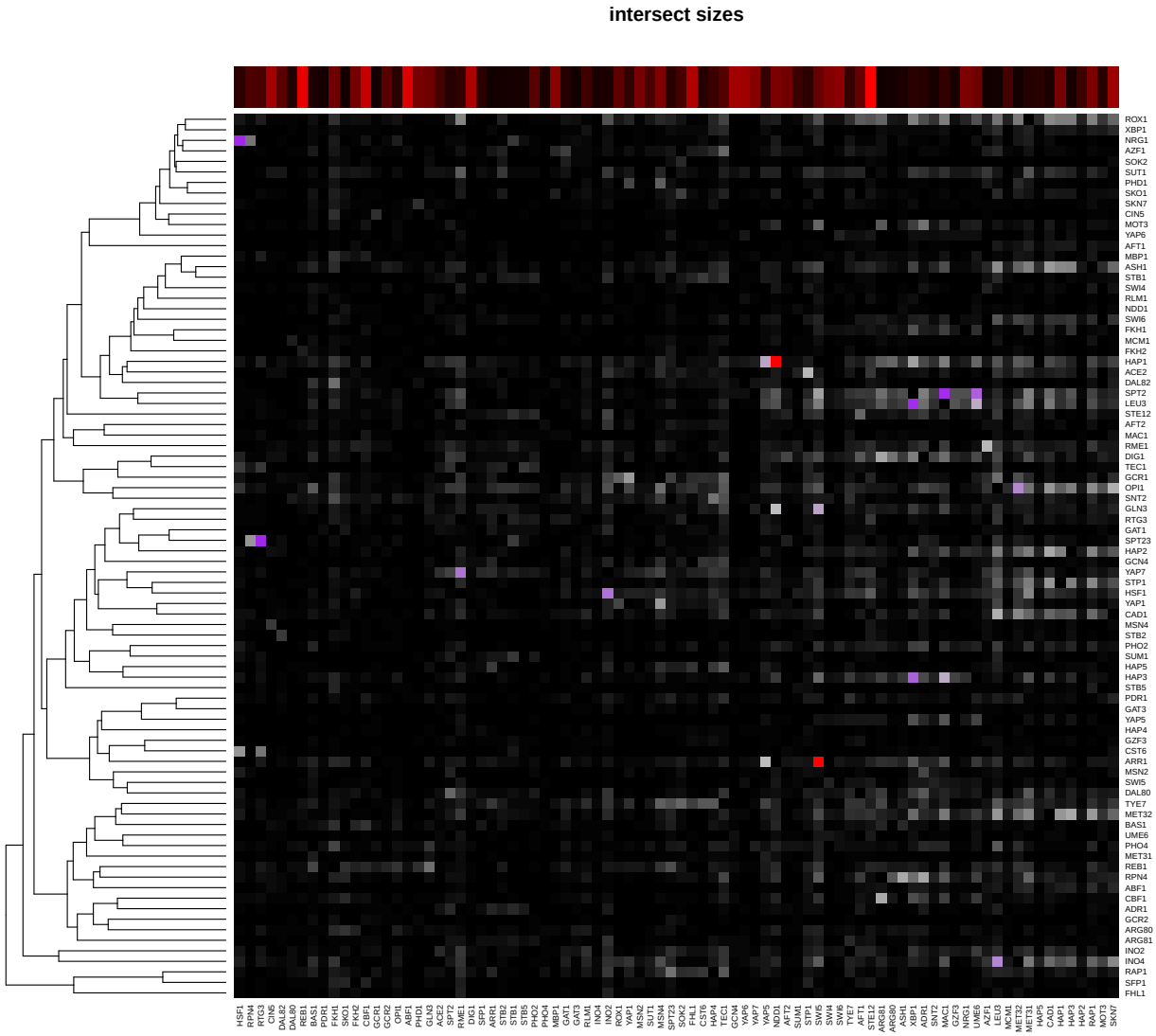


Figure 32: Heatmap of the pairwise intersect sizes of all transcription factor annotations. The bar on top shows the absolute size of the target gene annotation. Intersect profiles were clustered according to their correlation using average linkage clustering. This figure shows substantially different neighborhood relations than in (Figure 31).

Part IV
Salt sensitivity screen

Strains were spotted on YPD plates with 0.8 M sodium chloride at 30°C. We started with an OD_{600} of 0.1 and spotted a 1 : 10 dilution series of length 6.

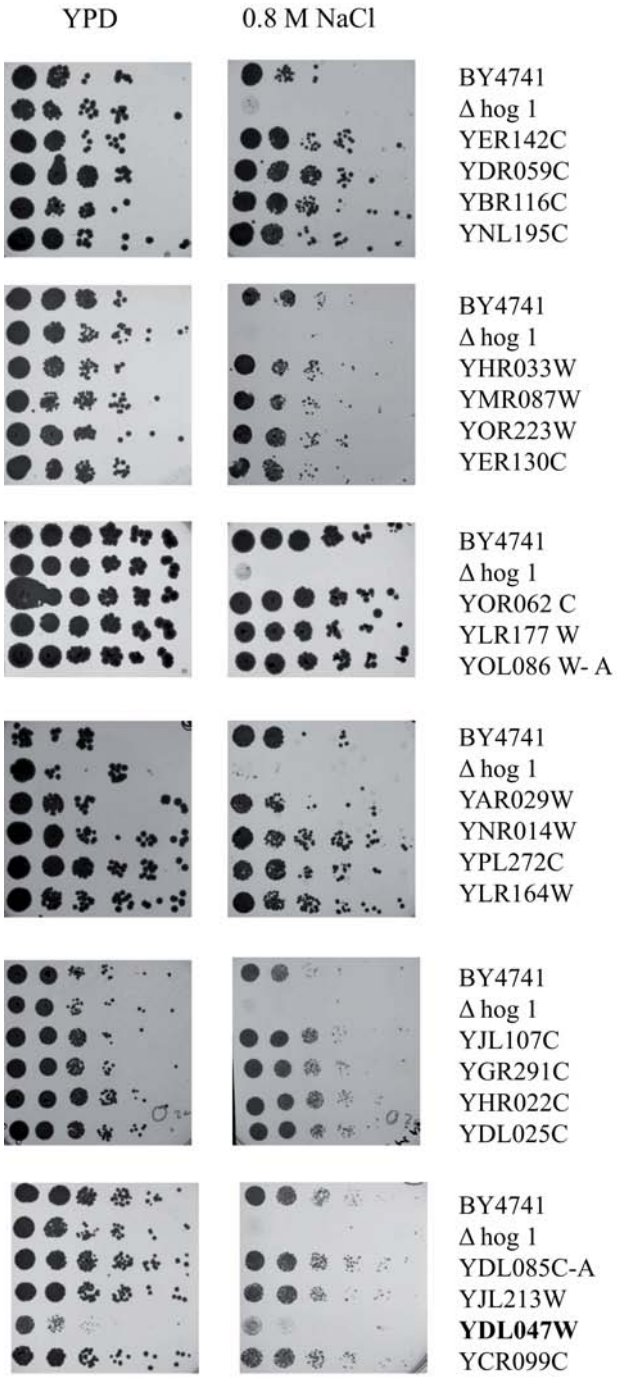


Figure 33: Dilution series of some of the "up"-cluster genes. Genes were selected for a particularly high rank gain and unknown biological function.

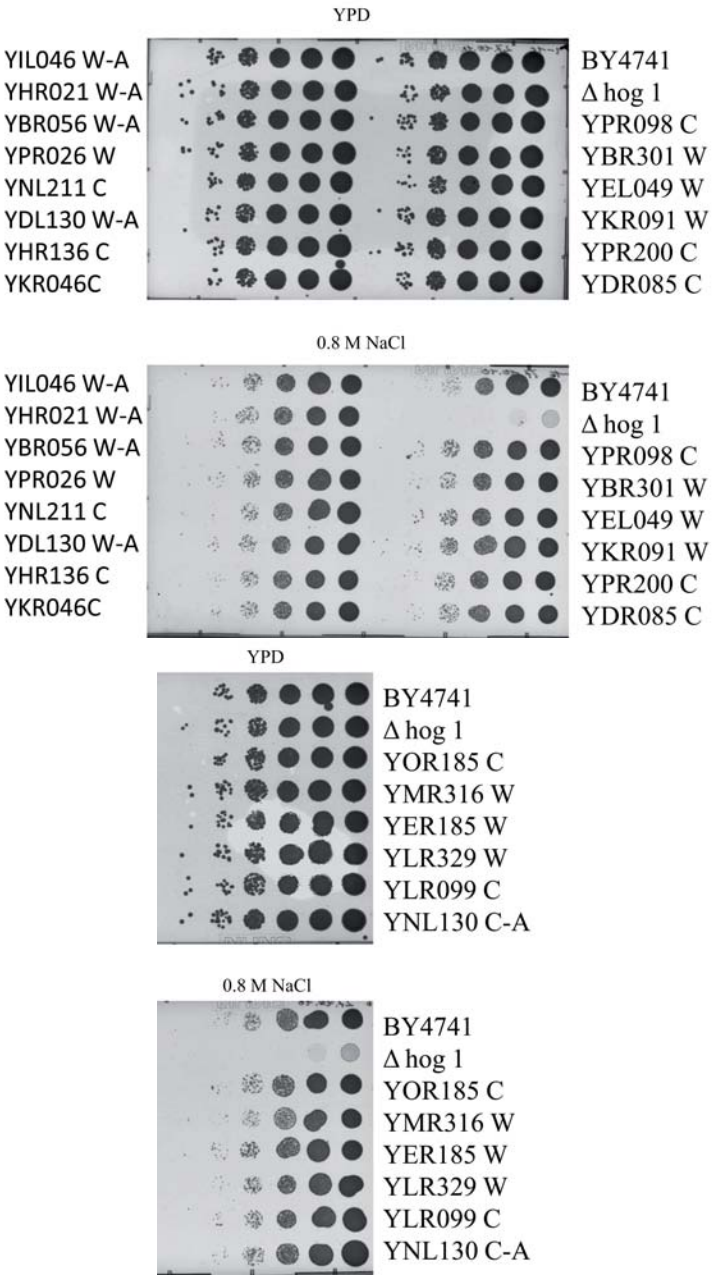


Figure 34: Dilution series of 20 of the marked differentially expressed genes in (Supplementary figure S8).

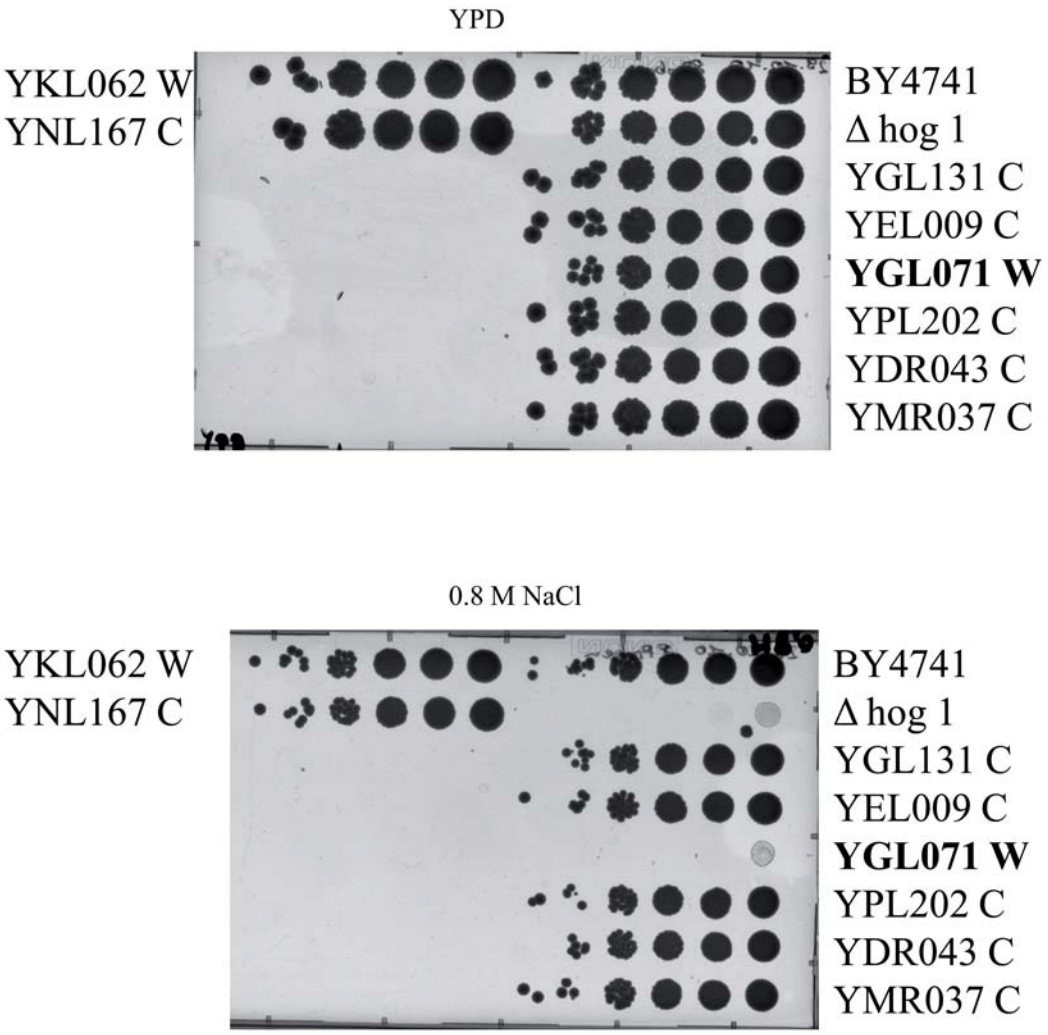


Figure 35: Dilution series of the transcription factors detected as active (main text Figure 4A).

References

[Benjamini & Hochberg, 1995] Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological) *57*(1), 289–300.

[Bryne *et al.*, 2008] Bryne, J. C., Valen, E., Tang, M.-H. E., Marstrand, T., Winther, O., da Piedade, I., Krogh, A., Lenhard, B., & Sandelin, A. (2008). Jaspar, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. Nucl. Acids Res. *36*(suppl_1), D102–106.

[Capaldi *et al.*, 2008] Capaldi, A. P., Kaplan, T., Liu, Y., Habib, N., Regev, A., Friedman, N., & O'Shea, E. K. (2008). Structure and function of a transcriptional network activated by the MAPK Hog1. Nature Genetics *40*(11), 1300–1306.

[Cherry *et al.*, 1998] Cherry, J. M., Adler, C., Ball, C., Chervitz, S. A., Dwight, S. S., Hester, E. T., Jia, Y., Juvik, G., Roe, T., Schroeder, M., Weng, S., & Botstein, D. (1998). SGD: Saccharomyces Genome Database.. Nucleic acids research *26*(1), 73–79.

[Dölken *et al.*, 2008] Dölken, L., Ruzsics, Z., Rädle, B., Friedel, C. C., Zimmer, R., Mages, J., Hoffmann, R., Dickinson, P., Forster, T., Ghazal, P., & Koszinowski, U. H. (2008). High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. RNA *14*(9), 1959–1972.

[Falcon & Gentleman, 2007] Falcon, S. & Gentleman, R. (2007). Using GOstats to test gene lists for GO term association. Bioinformatics *23*(2), 257–258.

[Friedel *et al.*, 2009] Friedel, C. C., Dölken, L., Ruzsics, Z., Koszinowski, U. H., & Zimmer, R. (2009). Conserved principles of mammalian transcriptional regulation revealed by RNA half-life.. Nucleic acids research *37*(17), e115+.

[Gentleman *et al.*, 2004] Gentleman, R., Carey, V., Bates, D., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J., & Zhang, J. (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biology *5*(10), R80+.

[Grigull *et al.*, 2004] Grigull, J., Mnaimneh, S., Pootoolal, J., Robinson, M. D., & Hughes, T. R. (2004). Genome-Wide Analysis of mRNA Stability Using Transcription Inhibitors and Microarrays Reveals Post-transcriptional Control of Ribosome Biogenesis Factors. Mol. Cell. Biol. *24*(12), 5534–5547.

[Harbison *et al.*, 2004] Harbison, C. T., Gordon, D. B., Lee, T. I., Rinaldi, N. J., Macisaac, K. D., Danford, T. W., Hannett, N. M., Tagne, J. B., Reynolds, D. B., Yoo, J., Jennings, E. G., Zeitlinger, J., Pokholok, D. K., Kellis, M., Rolfe, P. A., Takusagawa, K. T., Lander, E. S., Gifford, D. K., Fraenkel, E., & Young, R. A. (2004). Transcriptional regulatory code of a eukaryotic genome. Nature *431*, 99–104.

[Hereford, 1977] Hereford, L. (1977). Number and distribution of polyadenylated RNA sequences in yeast. Cell *10*(3), 453–462.

[Holstege *et al.*, 1998] Holstege, F. C., Jennings, E. G., Wyrick, J. J., Lee, T. I., Hengartner, C. J., Green, M. R., Golub, T. R., Lander, E. S., & Young, R. A. (1998). Dissecting the regulatory circuitry of a eukaryotic genome.. Cell *95*(5), 717–728.

[Krogan *et al.*, 2006] Krogan, N. J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A. P., Punna, T., PeregrÃÂn Alvarez, J. A. M., Shales, M., Zhang, X., Davey, M., Robinson, M. D., Paccanaro, A., Bray, J. E., Sheung, A., Beattie, B., Richards, D. P., Canadien, V., Lalev, A., Mena, F., Wong, P., Starostine, A., Canete, M. M., Vlasblom, J., Wu, S., Orsi, C., Collins, S. R., Chandran, S., Haw, R., Rilstone, J. J., Gandi, K., Thompson, N. J., Musso, G., St Onge, P., Ghanny, S., Lam, M. H. Y., Butland, G., Altaf-Ul, A. M., Kanaya, S., Shilatifard, A., O'Shea, E., Weissman, J. S., Ingles, C. J., Hughes, T. R., Parkinson, J., Gerstein, M., Wodak, S. J., Emili, A., & Greenblatt, J. F. (2006). Global landscape of protein complexes in the yeast saccharomyces cerevisiae. Nature *440*(7084), 637–643.

[MacIsaac *et al.*, 2006] MacIsaac, K., Wang, T., Gordon, D. B., Gifford, D., Stormo, G., & Fraenkel, E. (2006). An improved map of conserved regulatory sites for saccharomyces cerevisiae. BMC Bioinformatics *7*(1), 113+.

[Mani *et al.*, 2008] Mani, R., St, Hartman, J. L., Giaever, G., & Roth, F. P. (2008). Defining genetic interaction. Proceedings of the National Academy of Sciences *105*(9), 3461–3466.

[Matys *et al.*, 2006] Matys, V., Kel-Margoulis, O. V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K., Voss, N., Stegmaier, P., Lewicki-Potapov, B., Saxel, H., Kel, A. E., & Wingender, E. (2006). Transfac(r) and its module transcompel(r): transcriptional gene regulation in eukaryotes. Nucl. Acids Res. *34*(suppl_1), D108–110.

[Mayer *et al.*, 2010] Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Soding, J., & Cramer, P. (2010). Uniform transitions of the general RNA polymerase II transcription complex. Nature Structural & Molecular Biology *17*(10), 1272–1278.

[Muzzey *et al.*, 2009] Muzzey, D., Gómez-Uribe, C. A., Mettetal, J. T., & van Oudenaarden, A. (2009). A systems-level analysis of perfect adaptation in yeast osmoregulation.. Cell *138*(1), 160–171.

[Nagalakshmi *et al.*, 2008] Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., & Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing.. Science (New York, N.Y.) *320*(5881), 1344–1349.

[Nakao *et al.*, 2004] Nakao, A., Yoshihama, M., & Kenmochi, N. (2004). RPG: the Ribosomal Protein Gene database.. Nucleic acids research *32*(Database issue).

[Pan *et al.*, 2006] Pan, X., Ye, P., Yuan, D. S., Wang, X., Bader, J. S., & Boeke, J. D. (2006). A dna integrity network in the yeast saccharomyces cerevisiae.. Cell *124*(5), 1069–1081.

[Pelechano & Pérez-Ortín, 2010] Pelechano, V. & Pérez-Ortín, J. E. (2010). There is a steady-state transcriptome in exponentially growing yeast cells. Yeast *27*(7), 413–422.

[Proft & Struhl, 2004] Proft, M. & Struhl, K. (2004). MAP Kinase-Mediated Stress Relief that Precedes and Regulates the Timing of Transcriptional Induction. *118*(3), 351–361.

[Schwalb *et al.*, 2010] Schwalb, B., Dümcke, S., & Tresch, A. (2010). LSD Lots of Superior Depictions. to appear on R CRAN.

[Shalem *et al.*, 2008] Shalem, O., Dahan, O., Levo, M., Martinez, M. R., Furman, I., Segal, E., & Pilpel, Y. (2008). Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation. Molecular Systems Biology *4*.

[Smyth, 2004] Smyth, G. K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments.. Statistical applications in genetics and molecular biology *3*(1).

[Smyth & Speed, 2003] Smyth, G. K. & Speed, T. (2003). Normalization of cDNA microarray data.. Methods (San Diego, Calif.) *31*(4), 265–273.

[Soufi *et al.*, 2009] Soufi, B., Kelstrup, C. D., Stoehr, G., Fröhlich, F., Walther, T. C., & Olsen, J. V. (2009). Global analysis of the yeast osmotic stress response by quantitative proteomics. Molecular bioSystems *5*(11), 1337–1346.

[Teixeira *et al.*, 2006] Teixeira, M. C., Monteiro, P., Jain, P., Tenreiro, S., Fernandes, A. R., Mira, N. P., Alenquer, M., Freitas, A. T., Oliveira, A. L., & Sa-Correia, I. (2006). The yeastract database: a tool for the analysis of transcription regulatory associations in saccharomyces cerevisiae. Nucl. Acids Res. *34*(suppl_1), D446–451.

[Wang *et al.*, 2002] Wang, Y., Liu, C. L. L., Storey, J. D., Tibshirani, R. J., Herschlag, D., & Brown, P. O. (2002). Precision and functional specificity in mRNA decay.. Proceedings of the National Academy of Sciences of the United States of America *99*(9), 5860–5865.

[Wu *et al.*, 2004] Wu, Z., Irizarry, R. A., Gentleman, R., Murillo, F. M., & Spencer, F. (2004). A Model-Based Background Adjustment for Oligonucleotide Expression Arrays. Journal of the American Statistical Association *99*(468).

[Yang *et al.*, 2002] Yang, Y. H., Dudoit, S., Luu, P., Lin, D. M., Peng, V., Ngai, J., & Speed, T. P. (2002). Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Research *30*(4), e15.

[Zacher *et al.*, 2010] Zacher, B., Kuan, P., & Tresch, A. (2010). Starr: Simple Tiling ARRay analysis of Affymetrix ChIP-chip data. BMC Bioinformatics *11*(1), 194+.