Profiling Condition-Specific, Genome-Wide Regulation of mRNA Stability in Yeast

Barrett C. Foat*, S. Sean Houshmandi†, Wendy M. Olivas†, and Harmen J. Bussemaker*‡

*Department of Biological Sciences, Columbia University, New York, New York 10027, U.S.A.
†Department of Biology, University of Missouri–St. Louis, St. Louis, Missouri 63121, U.S.A.
‡Center for Computational Biology and Bioinformatics, Columbia University, New York, New York 10032, U.S.A.

Abstract

The steady-state abundance of an mRNA is determined by the balance between transcription and decay. While regulation of transcription has been well studied both experimentally and computationally, regulation of transcript stability has received little attention. We developed a novel algorithm, MatrixREDUCE, that discovers the position-specific affinity matrices for unknown RNA-binding factors and infers their condition-specific activities, using only genomic sequence data and steady-state mRNA expression data as input. We identified and computationally characterized the binding sites for six mRNA stability regulators in *Saccharomyces cerevisiae*, which include two members of the Pumilio-homology domain (Puf) family of RNA-binding proteins, Puf3p and Puf4p. We provide computational and experimental evidence that regulation of mRNA stability by these factors is dynamic and responds to a variety of environmental stimuli.

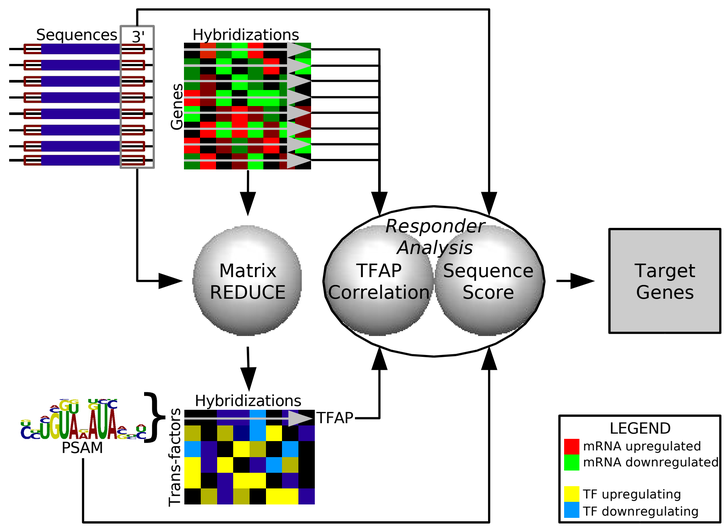

Figure 1: MatrixREDUCE data flow. Using microarray data for 757 pairs of experimental conditions and the downstream 200bp of every yeast ORF as input to MatrixREDUCE, we identified position-specific affinity matrices (PSAMs) corresponding to putative mRNA stability regulators and inferred trans-factor activity profiles (TFAPs) across all microarray conditions. For each discovered PSAM, we then ranked every gene by the sequence score for its 3' sequence and the correlation between its mRNA expression profile and the respective TFAP. Optimal thresholds were determined for the sequence scores and the TFAP correlations. This "responder analysis" produced a list of putative target genes that are likely to contain functional binding sites for each trans-factor modeled by a PSAM.

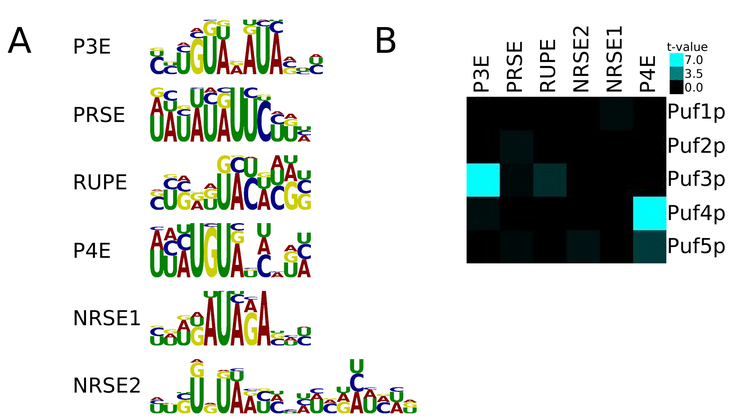

Figure 2: (**A**) Weight matrix logos for the six best supported 3' PSAMs. The PSAMs were discovered directly from the large set of steady-state mRNA expression data and sequences downstream of every ORF. (**B**) The matches to PSAMs in 3' sequences of genes can be correlated with trans-factor binding data to identify trans-factors that correspond to specific PSAMs. The degree of correlation with binding data (for Puf1p, Puf2p, Puf3p, Puf4p, and Puf5p) is represented by shades of cyan.

Gene Ontology Category Description	P3E	PRSE	P4E	NRSE1	NRSE2
C mitochondrion					
C mitochondrial ribosome					
C mitochondrial matrix					
C mitochondrial large ribosomal subunit					
C mitochondrial small ribosomal subunit					
F tRNA ligase activity					
P aerobic respiration					
C mitochondrial inner membrane					
P electron transport					
P oxidative phosphorylation					
P ATP synthesis coupled electron transport					
C mitochondrial electron transport chain					
P ribosome biogenesis and assembly					
P rRNA processing					
C nucleolus					
C small nucleolar ribonucleoprotein complex					
C cytosolic ribosome					
C cytosolic small ribosomal subunit					
C cytosolic large ribosomal subunit					
P establishment and/or maintenance of cell polarity					
P cytoskeleton organization and biogenesis					
C endomembrane system					
P transcription from RNA polymerase I promoter					
F translation factor activity, nucleic acid binding					
P RNA localization					
P nucleocytoplasmic transport					

P-value
$10^{-7} < P \le 10^{-3}$
$10^{-15} < P \le 10^{-7}$
$P \le 10^{-15}$

Figure 3: Gene Ontology analysis. The enrichment of the PSAM targets in each category was quantified using the cumulative hypergeometric distribution. The color scale corresponds to Bonferroni-corrected P-values. The ontology to which a category belongs is indicated by C (subcellular component), P (biological process), or F (molecular function). Only a representative selection of all significant functional categories is shown. The significantly enriched functional categories for each PSAM were generally consistent with their inferred activity profiles (Fig. 4).

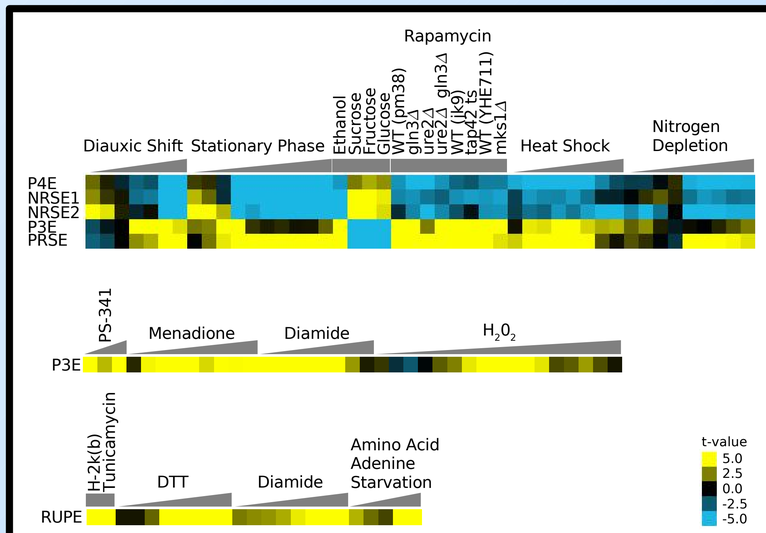

Figure 4: Trans-factor activity profiles (TFAPs). The direction and degree of the correlation between PSAM match scores and mRNA expression data is represented by a color scale ranging from bright blue (strong negative correlation) to bright yellow (strong positive correlation). These correlations correspond to changes in activity of the mRNA stability-regulating trans-factors whose binding affinities are modeled by the PSAMs. The shallow triangles represent the progression of time within time course experiments.

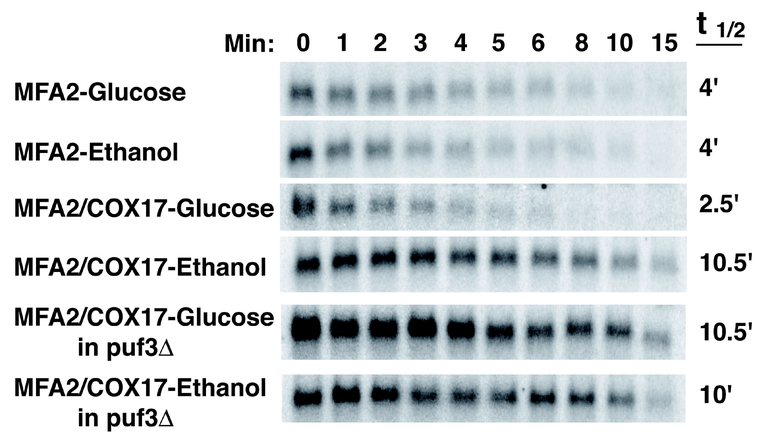

Figure 5: Regulation of Puf3p activity in response to a change in carbon source. Shown are Northern blot analyses of the decay of MFA2 mRNA or the hybrid MFA2/COX17 mRNA expressed from wild-type or puf3Δ yeast grown in media containing 2% glucose or 2% ethanol. Minutes following transcriptional repression are indicated above the set of blots, with the half-lives ($t_{1/2}$) as determined from multiple experiments.

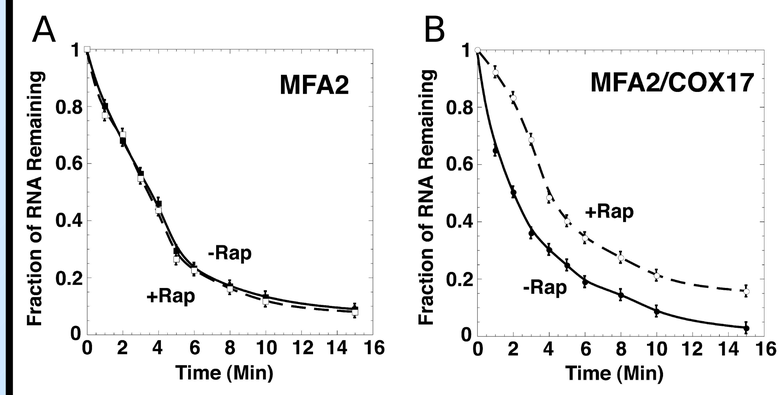

Figure 6: Regulation of Puf3p activity in response to rapamycin. (**A**) Data from Northern blot analyses of MFA2 or (**B**) the hybrid MFA2/COX17 mRNA decay are plotted, with minutes following transcriptional repression on the x-axis and the fraction of RNA remaining as compared to the steady-state RNA level at time 0 on the y-axis. Decay was monitored with or without rapamycin treatment for 60 minutes prior to transcriptional repression as follows: (**A**) MFA2 without rapamycin (closed square), MFA2 with rapamycin (open square), (**B**) MFA2/COX17 without rapamycin (closed circle), and MFA2/COX17 with rapamycin (open circle). Data points are averages of multiple experiments.

Methods

MatrixREDUCE Model

$$A_{gh} = C_h + \sum_{\phi \in M} F_{\phi h} N_{\phi g} + \epsilon_{gh} \qquad N_{\phi g} = \sum_{i=1}^{L_g} \prod_{j=0}^{L_\phi - 1} w_{\phi j b_{(i+j,g)}}$$

Here, A_{gh} is the $\log_2$ expression ratio for gene g between the two conditions in hybridization h; C_h is an intercept term for hybridization h; ϕ is one of the trans-factors in the model M; $F_{\phi h}$ is a regression coefficient that quantifies the change in post-translational activity of the trans-factor; $N_{\phi g}$ scores the occurrences of binding sites for factor ϕ in the regulatory region of gene g; and ϵ_{gh} represents the residuals of A_{gh}. For each nucleotide position j in the position specific affinity matrix (PSAM) for trans-factor ϕ, there is a weight $w_{\phi jb}$ for each nucleotide b. The weights range between zero and one. The score for a given position i in the sequence of gene g equals the product of the weights for the nucleotides found at positions $(i + j)$ with $0 \le j < L_\phi$, where L_ϕ is the length of the PSAM. The score $N_{\phi g}$ for the entire regulatory sequence of gene g equals the sum of the scores for each position i in the sequence of length L_g.

Responder Analysis

$$P = \sum_{i=X}^{A} \frac{\binom{G-A}{B-i}\binom{A}{i}}{\binom{G}{B}}$$

All genes were ranked by their expression-TFAP correlation and their sequence scores $N_{\phi g}$. The optimum rank thresholds were determined by calculating the P-value for every combination of ranks. G is the number of genes in the genome; A is the number of genes with a rank equal to or higher than the rank being tested for the first metric; B is the number of genes with a rank equal to or higher than the rank being tested for the second metric ($B \ge A$); and X is the number of genes in the overlap between both gene sets. Once the optimal P-value was determined, genes were considered target genes if they had ranks above threshold for both the sequence score and the TFAP correlation.

Conclusions

- Puf3p destabilizes transcripts coding for mitochondrial components in response to metabolite repressing sugars.
- The trans-factor that binds the novel positive response to starvation element (PRSE) regulates transcripts coding for components of the electron transport chain in response to nutrient conditions.
- Puf4p and the trans-factors that bind the novel negative response to starvation elements (NRSE1 and NRSE2) regulate transcripts encoding for proteins involved in cytosolic ribosome biogenesis in response to nutrient conditions.
- The trans-factor that binds the novel response to unfolded proteins element (RUPE) is involved in the adapting to conditions that would cause proteins to misfold.
- Puf3p, Puf4p, and the trans-factors that bind the PRSE, NRSE1, and NRSE2 are likely downstream of the Target of Rapamycin signaling pathway.
- Using only transcript nucleotide sequence and steady-state microarray data, our methods allow one to discover the PSAMs of mRNA stability regulators, to determine which mRNAs are most likely targeted by the trans-factors, and to identify the conditions under which these factors modulate mRNA stability.
- *Regulation of mRNA stability is not a special-case phenomenon but rather a pervasive regulatory mechanism that is used to rapidly adapt cellular processes to a changing environment.*