Supporting Information

Gerashchenko et al. 10.1073/pnas.1120799109

SI Materials and Methods

Preparation of Lysates. Hydrogen peroxide (0.2 mM final concentration) was added to 400 mL of yeast culture, and the culture was incubated further for either 5 or 30 min. A 50-mL aliquot was taken rapidly and pelleted by centrifugation for 1 min at 3,400 × *g* at 4 °C; then the pellet was frozen immediately in liquid nitrogen. This aliquot was used for mRNA isolation. The rest of the yeast culture was treated with 0.1 g/L cycloheximide, incubated for 3 min with shaking, and centrifuged at 3,400 × *g* for 4 min. The pellet was resuspended in 3 mL of ice-cold polysome lysis buffer [20 mM Tris·HCl (pH 8.0), 140 mM KCl, 5 mM $MgCl_2$, 0.2g/L cycloheximide, 1% Triton-×100] and recentrifuged. The supernatant was removed, and the pellet was treated with 1.2 mL of the polysome lysis buffer along with an equal amount of glass beads. The resulting mix was vortexed rigorously five times for 1 min with 1-min breaks. The aqueous fraction was collected and clarified by centrifugation for 10 min at 20, 000 × *g*. The final yeast lysate containing intact ribosomes was flash frozen in liquid nitrogen.

Ribosome Fractionation and RNA Extraction. A 50-U aliquot of the cell extract (OD_{260}) was treated with 1,000 U of *Escherichia coli* RNase I (Ambion) and incubated for 1 h at room temperature with gentle shaking. The sample volume was brought to 1 mL by adding polysome gradient buffer [20 mM Tris·HCl (pH 8.0), 140 mM KCl, 5 mM $MgCl_2$, 0.2g/L cycloheximide, 0.5 mM DTT]. Sucrose gradients (10–50% wt/wt) were prepared in SW41 ultracentrifuge tubes (Beckman) using a freeze-thaw method (1). RNase-digested and control samples were loaded onto gradients and spun for 3 h at 35,000 rpm and 4 °C in a SW41 rotor (Beckman). Gradients were fractionated at 1 mL/min using the Brandel gradient fractionation system coupled with the BioRad UV detector, which continually monitored OD_{254} values. As a chase solution, 60% (wt/wt) sucrose was used, and fractions representing the monosome peak were pooled in one tube. Each sample was filtered through an Amicon-100 microcentrifugator (Millipore) for 10 min at 10,000 × *g*. The release buffer [20 mM Tris·HCl (pH7.0), 2 mM EDTA, 40 U/mL Superase-In (Ambion)] was added to the retentate until the volume reached 0.5 mL, and each sample was incubated further for 10 min on ice and then was filtered again. Flow-through fractions containing the majority of footprints were collected, and RNA was purified by hot acid phenol extraction and precipitated by ethanol with glycogen as a coprecipitant. Pellets were solubilized in 10 μL of water and analyzed on 15% Tris/borate/EDTA (TBE)-urea polyacrylamide gels (Invitrogen). The bands around 28–32 nt were cut off, and RNA was eluted in 300 μL of the elution buffer containing 20 mM Tris·HCl (pH 7.0), 2 mM EDTA, 0.5 M ammonium acetate, and 2 μL Superase-In, precipitated, and resuspended in 8 μL of water. After addition of 1 μL of T4 kinase A buffer and 1 μL of T4 kinase (Fermentas), the mixture was incubated for 60 min at 37 °C, inactivated for 5 min at 80 °C, and ethanol-precipitated.

Library Construction for Footprint Sequencing. Polyadenylation of RNA footprints was performed by adding 0.5 U of polyA polymerase (New England Biolabs) in a total volume of 5 μL and incubating the mixture for 15 min at 37 °C. The enzyme was inactivated by heating the mixture at 80 °C for 10 min. The whole reaction mix was used for reverse transcription. Superscript III (Invitrogen) polymerase was used according to manufacturer's instructions in a total reaction volume 12 μL. The RT-library primer was used for each individual sample. Finally, 0.5 μL of 2 M sodium hydroxide was added to hydrolyze RNA from RNA-DNA duplexes, and the sample was incubated for 30 min at 98 °C. Then, 0.5 μL of 2 M HCl was applied to neutralize the solution. Upon the addition of an equal volume of TBE-sample buffer (Invitrogen), the reverse-transcription mixture was loaded onto a 10% TBE-urea gel (Invitrogen). The band corresponding to the elongated RT-library primer was cut, and DNA was eluted in 300 μL of 20 mM Tris·HCl (pH 7.0). An important step for efficient enrichment of ribosomal footprints was the subtractive hybridization of contaminating rRNA fragments. For this step, the biotinylated DNA oligonucleotide "bioAntiRiboPrime" (Table S5) was attached to streptavidin-activated magnetic beads (Invitrogen) as recommended in the manufacturer's manual. Ribosomal footprints eluted from the gel were incubated with these beads, and nonribosomal fragments that did not bind to the beads were collected and ethanol-precipitated. They served as substrates for CircLigase II (Epicentre) in a 10-μL reaction mix. Circularized ribosomal footprints were used as a template for the final library-amplification step. PCR conditions were set as follows: 0.5 μL of Phusion polymerase (New England Biolabs), 1 μL of 10 mM dNTP, 1 μL of CircLigase II (Epicentre), 10 μL of HF buffer (New England Biolabs), and 10 pmol of custom ill-Cluster3 and ill-Cluster4 primers compatible with Illumina sequencers (Table S5) in a 50-μL mixture. Annealing took place at 70 °C for 15 s, and elongation took place at 72 °C for 10 s. Several reaction tubes were set up to be removed from the PCR machine after 12–18 cycles. The product yield was analyzed on 8% nondenaturing TBE polyacrylamide gels to select samples (based on PCR conditions) before the appearance of nonspecific bands. The library was cut from the gel, eluted in 20 mM Tris·HCl (pH 7.0), ethanol-precipitated, and sequenced on the Illumina GLx2 or HiSeq2000 platforms.

mRNA Extraction. Frozen aliquots were thawed and lysed in 400 μL of lysis buffer (mRNA DIRECT kit; Invitrogen). A 250-μL aliquot of magnetic beads and two rounds of purification were implemented according to the manufacturer's protocol.

mRNA Sequencing Library Construction. mRNA was fragmented by alkaline solution [2 mM EDTA, 100 mM Na_2CO_3 (pH 9.2)], the fragments were loaded onto a 15% TBE-urea gel, and the 28- to 32-nt region was cut from the gel. Further steps in library preparation were identical to those used for ribosomal footprints, the only difference being that barcoded RT-library 1–4 primers were used that allowed multiplexing of samples for sequencing (Table S5). The subtractive hybridization step was omitted. The PCR annealing temperature was set to 60 °C with ill-Cluster3 and ill-Cluster5 primers.

Bioinformatics Analyses. In-house Perl scripts were used to prepare reference databases. We created several references using the Saccharomyces Genome Database as a starting point. The largest reference ("Functional") included all cDNAs except for transposons and dubious genes. Among these cDNAs, the genes with a high degree of sequence similarity were combined into single records. This dataset was used for differential gene-expression and translation studies. Additionally, 100 nt from the 5′ end of each gene were deleted to avoid bias caused by the region with elevated footprint density. Another reference ("noRepeat") included only unique gene sequences to which footprints could be aligned unambiguously. It was used when the nucleotide position-sensitive features of translation were examined. Alignment of sequencing reads was performed by Bowtie software v.0.12.7 (2) allowing two mismatches per read. Alignment against 5′ UTR was done with one mismatch allowed. Because every read bears a polyA tail at the end,

we omitted all "A" from the 3′ ends of sequences before aligning. Reads shorter than 23 nt after polyA removal were discarded.

Calculation of Translation Efficiency. Translational efficiency (TE) is a measure of how well translated a particular gene is relative to its mRNA abundance. TE can be defined as the number of footprints divided by the number of mRNA-seq reads normalized to gene length and total number of reads, i.e., footprint in reads per kilobase per million mapped reads (rpkm)/mRNA rpkm. A higher TE value represents greater potency of mRNA for translation. TE was used to examine translationally regulated genes. If a gene had a $\log_2$ (TE change) above 1.5 or below 1.5, it was considered up- or down-regulated, respectively. Fig. S3*B* shows the fraction of false positives at the selected threshold.

Inferring Translation Rate from Sequencing Data. Sequenced footprints represent pieces of mRNA trapped in the active translating ribosomes. A higher number of footprints aligned to a gene sequence implies a higher yield of the corresponding protein. This assumption is more reliable for genes with more even footprint coverage. Significant deviation from evenness may indicate ribosomal pauses in certain locations; such pauses complicate the inference of protein production. In this study, we observed higher density of footprints at the beginning of mRNAs; therefore, we discarded 100 nucleotides from the 5′ end of every gene to minimize unevenness of footprint coverage along transcripts.

Differential Gene Translation Analysis. All experimental samples were collected in duplicate. Based on the correlation between the replicates, we set up an rpkm threshold of 10 for the genes whose translation and transcription could be determined reproducibly (Fig. S4 *A* and *B*). The gene was considered regulated if its rpkm value changed more than 2.6-fold (1.4 in log2 scale). This threshold eliminated most of false-positive hits (Fig. S4*D*).

Comparing Translation Changes with Transcription Changes. In an ideal situation, assuming that transcript abundance is the only determinant for protein translation, changes in transcript abundance would be followed by the same changes in footprint abundance. In reality such coordinated changes never happen, as illustrated in Fig. 4*B*. Axis values are calculated as footprint change versus transcript change between two experimental conditions. Footprint change is defined as log2[(Footprints in peroxide-treated sample, rpkm)/(Footprints in initial sample, rpkm)]. Transcript change is defined in a same way for mRNA-seq reads.

Codon Translation Analysis. In an ideal situation, ribosomal footprints should be 28 nt in length. However, RNase I, which was used to degrade unprotected mRNA segments, occasionally left extra nucleotides or cut off extra nucleotides. By plotting a distribution of the footprint length, we found that RNase creates footprints mostly are 27–29 nt in length (Fig. S4*C*). A footprint can be aligned to the reference ORFs, and the position of its 5′ end relative to the reading frame can be obtained. If the 5′ end of a footprint matched the exact border of a codon, we considered it "ideal." If the 5′ end of a footprint matched the position of a codon ±1 nt, we deleted or added the first nucleotide, respectively. Thus, we minimized the error of ribosome position determination and defined which codon was located in the A site.

To estimate differences in TE among various codons (61 codons in total), we used following procedure. First, predicted occupancy was calculated for each type of codon as its frequency in mRNA sequence, normalized to gene expression (translation) and length (assuming that all codons are translated at the same rate). These values were compared with the observed frequencies. As a measure of difference, we used the following formula [(Observed) − (Predicted)]/(Predicted), which gave us an estimate of how the use of a particular codon compared with the predicted value.

Frameshift Analyses. The regions 50 nt downstream of stop codons of every gene were examined for the presence of ribosomal footprints. Footprint mapping similar to gene-coverage analysis was used to select possible frameshift extensions over readthrough events. Footprint reads were assigned to all possible reading frames and counted. During counting, reads were used as is; i.e., we did not add or subtract nucleotides from the 5′ ends. Candidates with signs of translation in different frames downstream of their stop codons were checked manually to exclude dubious cases and to define the frameshift regions more precisely.

Selecting Proteins with Potential N-Terminal Extensions. Some genes have ribosome profiling (Ribo-seq) footprints mapped to their 5′ UTRs in close proximity to annotated start codons. We marked proteins as potential bearers of N-terminal extensions if they satisfied three conditions. First, they were represented by at least 50 rpkm Ribo-seq counts 45 nt upstream of known ORFs. Second, the majority of Ribo-seq footprints mapped to these regions were in the same reading frame as the annotated proteins. Third, there were no stop codons in this frame 45 nt upstream of the annotated start codon (Table S3).

1. Fourcroy P, et al. (1981) Polyribosome analysis on sucrose gradients produced by the freeze-thaw method. *J Biochem Biophys Methods* 4(3–4):243–246.

2. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10:R25.

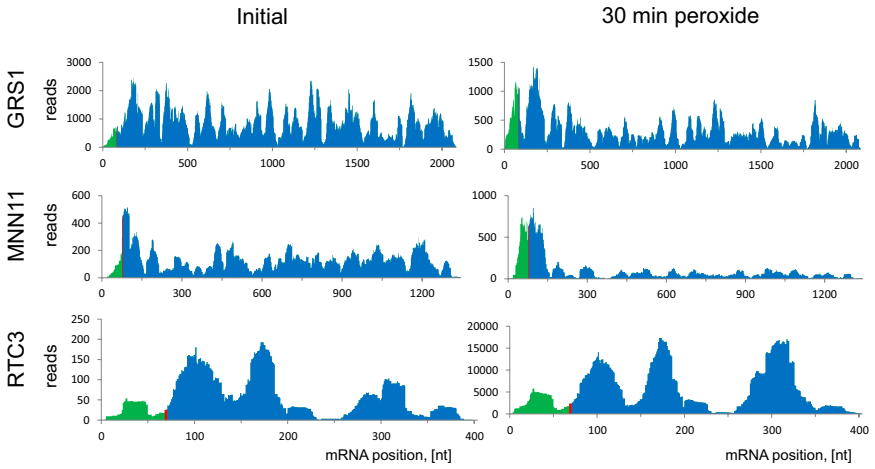

Fig. S1. Examples of proteins with N-terminal extensions. These proteins were selected from Table S3 to represent different scenarios of the N-terminal extension/ORF interplay. This figure compares the translation of proteins in control and 30-min hydrogen peroxide treatment samples. The entire 5′ UTR and ORF of the gene were used to generate the coverage density map. The 5′-UTR part of the mRNA is shown in green, the AUG start codon in red, and the annotated gene in blue.

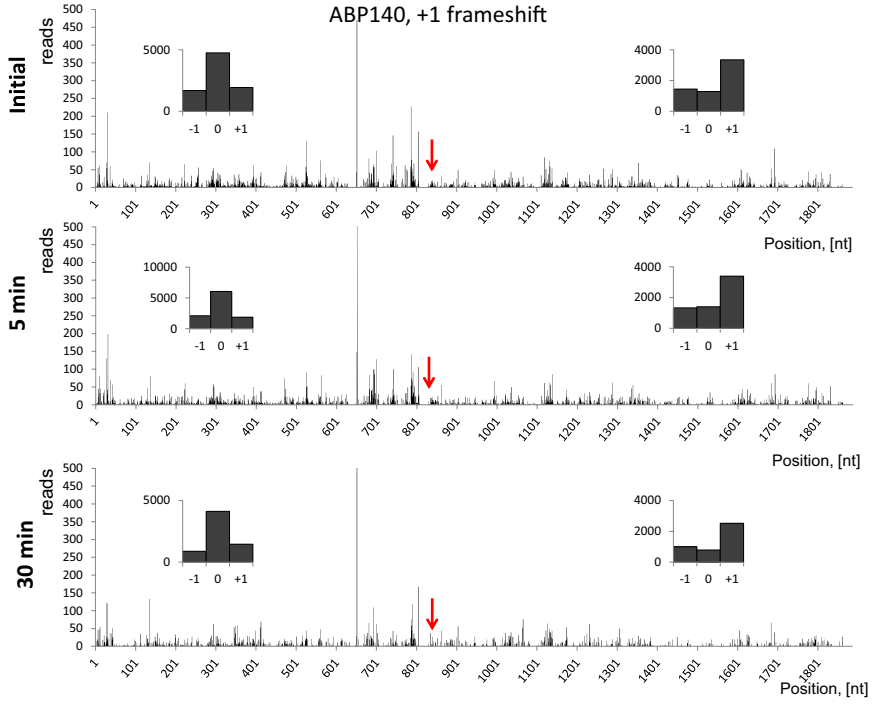

Fig. S2. Validation of the frameshift in ABP140 gene. The 5′ ends of footprints were mapped to the genomic sequence of antizyme. The red arrow indicates the frameshift position. (*Insets*) Histograms show footprint counts, matching one of three possible frames either to the left or to the right of the frameshift. The "0" frame is the frame with the annotated start codon. The greatest number of footprints matched the "0" frame before the frameshift and the "+1" frame after the frameshift. Thus, we observed a change of frame, leading to the translation of a longer protein.

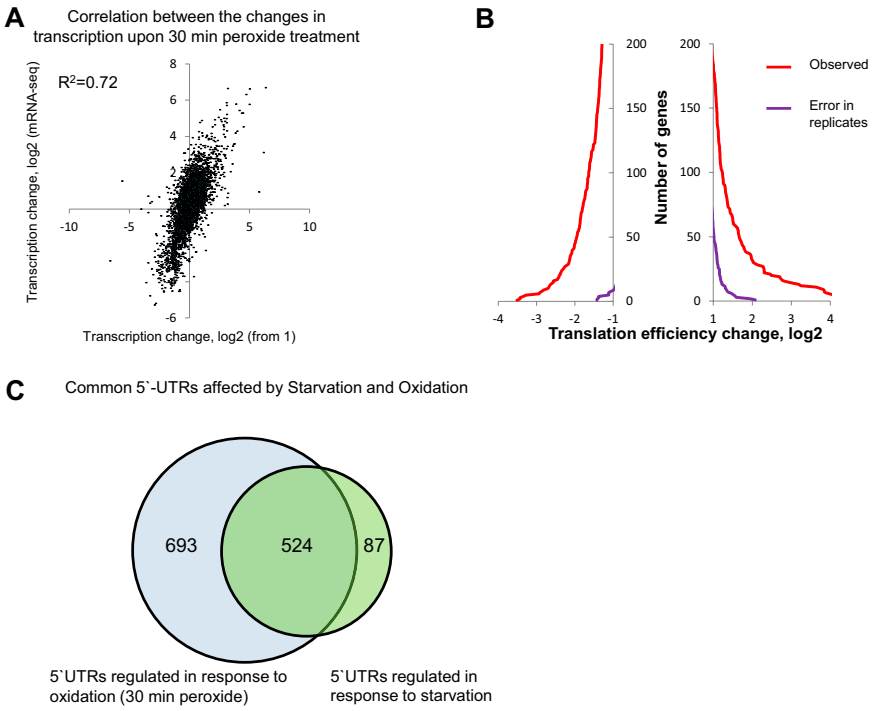

Fig. S3. (*A*) Comparison of gene expression in our RNA-seq data and the Gasch et al. (1) microarray data at the 30-min time point. Peroxide concentration used in our study was 0.2 mM and in the Gasch et al. study was 0.32 mM. Microarray data were taken from the online supplement of ref. 1. (*B*) Estimation of the error rate for TE change. The purple line shows how many genes at a certain threshold would be assigned mistakenly if two biological replicates were compared. The red line shows a number of genes in which the TE changed from the initial state to 30-min peroxide time point. (*C*) Comparison of ribosome occupancies at the 5′ UTRs affected by oxidative stress and starvation. Data for starvation were calculated by the procedure used to calculate oxidative stress. Raw sequencing files were taken from ref. 2.

1. Gasch AP, et al. (2000) Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell* 11:4241–4257.
2. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324:218–223.

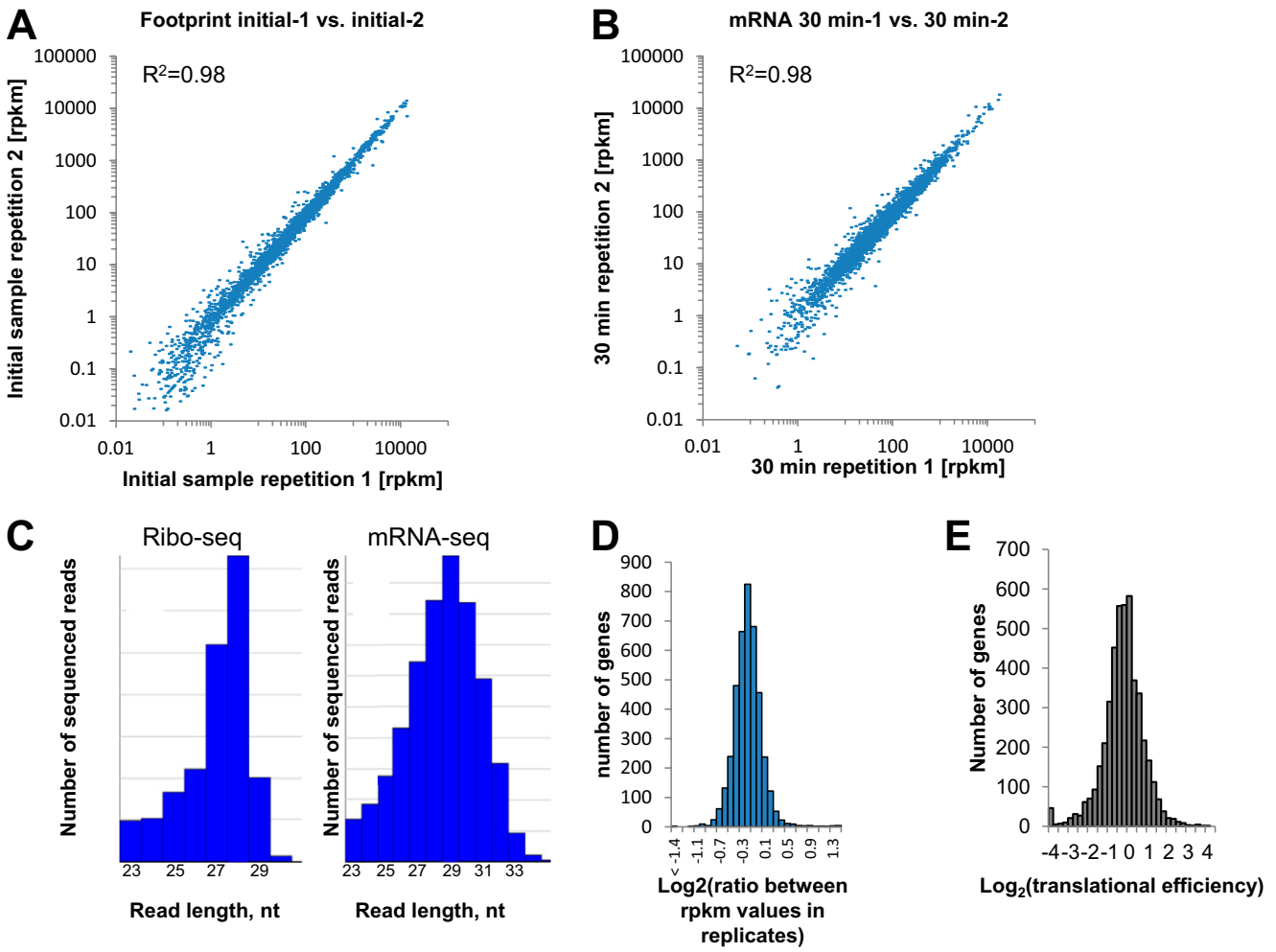

Fig. S4. (*A* and *B*) Comparison of gene expression in two replicates of footprints. *A* shows footprints, and *B* shows mRNA reads. Correlation coefficients are indicated in the figure. (*C*) Distribution of sequence reads by length in the control sample. (*Left*) Footprints. (*Right*) mRNA reads. Poly(A) tails were omitted from the reads. (*D*) Justification for threshold selection. The majority of differences between the two replicates fit in ±1 interval on the log2 scale. However, to minimize false-positive hits, we set up the ±1.4 interval as the threshold. This threshold allowed us to avoid most false positives in the 5-min peroxide treatment samples in which the overall count of regulated genes was low. (*E*) Histogram of TE shown as $\log_2$(number of footprints/number of reads from RNA-seq).

Table S1. Statistics of deep-sequencing reads in Ribo-seq

Footprints	Initial-1	initial-2	5min-1	5min-2	30min-1	30min-2
Total reads	27,145,924	84,852,974	13,341,052	82,763,853	5,981,943	80,589,116
Genomic, nonrRNA	25,302,082	79,522,848	12,204,639	74,177,834	5,271,843	70,444,698
ORF_minus100nt, uniq	18,690,126	61,222,201	8,297,207	49,006,214	3,435,799	42,568,826
5′ UTR	61,769	228.496	176,003	867,375	120,516	1,241,515

Table S2. Statistics of deep-sequencing reads in mRNA-seq

mRNA	Initial-1	5min-1	5min-2	30min-1	30min-2
Total reads	22,560,757	18,283,784	13,424,316	20,910,828	19,871,495
Genomic, nonrRNA	20,707,193	17,434,262	12,398,186	18,250,816	19,301,893
ORF_minus100nt, uniq	12,211,073	9,849,232	7,614,102	11,834,969	12,257,517
5′ UTR	297,592	361,129	257,098	298,010	319,222

Table S3. Proteins with translated N-terminal extensions

Gene	Name
YDR077W	SED1
YHR179W	OYE2
YHR087W	RTC3
YPL154C	PEP4
YDL022W	GPD1
YIR037W	HYR1
YBR221C	PDB1
YGL039W	
YDR086C	SSS1
YKL103C	LAP4
YBR121C	GRS1
YOR039W	RPS12
YMR297W	PRC1
YNL064C	YDJ1
YJL183W	MNN11
YFR049W	
YGR146C	ECL1
YMR088C	VBA1
YPL183W-A	TAE4
YKL004W	AUR1
YKR052C	MRS4
YIL124W	AYR1
YDR043C	NRG1
YAL012W	CYS3
YKL138C	MRPL31
YER048W-A	ISD11
YER133W	GLC7
YPL170W	DAP1
YJL099W	CHS6
YOR335C	ALA1
YPR182W	SMX3
YLR332W	MID2

Table S4. Proteins with frameshifts induced by hydrogen peroxide treatment

Gene	Name
YKL157W	APE2
YPL224C	MMT2
YJR103W	URA8
YLR179C	Function unknown

Table S5. Primers used in library preparation

RT-library	pGATCGTCGGACTGTAGAACTCTθCAAGCAGAAGACGGCATACGATTTTTTTTTTTTTTTTTTVN
RT library1*	pCGTGATGATCGTCGGACTGTAGAACTCTGAACCTGTCGGTGGTCGCCGTATCATT/iSp18/CAAGCAGAAGACGGCATACGATTTTTTTTTTTTTTTTTVN
RT library2*	pTGGTCAGATCGTCGGACTGTAGAACTCTGAACCTGTCGGTGGTCGCCGTATCATT/iSp18/CAAGCAGAAGACGGCATACGATTTTTTTTTTTTTTTTTVN
RT library3*	pATTGGCGATCGTCGGACTGTAGAACTCTGAACCTGTCGGTGGTCGCCGTATCATT/iSp18/CAAGCAGAAGACGGCATACGATTTTTTTTTTTTTTTTTVN
RT library4*	pCTGATCGATCGTCGGACTGTAGAACTCTGAACCTGTCGGTGGTCGCCGTATCATT/iSp18/CAAGCAGAAGACGGCATACGATTTTTTTTTTTTTTTTTVN
Ill-cluster 3	CAAGCAGAAGACGGCATACGA
Ill-cluster 4	AATGATACGGCGACCACCGACAGGTTCAGAGTTCTACAG
Ill-cluster 5	AATGATACGGCGACCACCGA
bioAntiRiboPrime	bio\GAGGTGCACAATCGACCG

θ, abasic site (dSpacer); iSp18, internal spacer 18; p, phosphate.
*Primers with a barcode tag at the 5′ end.

Other Supporting Information Files

Dataset S1 (TXT)
Dataset S2 (TXT)