Supplemental information

Ribosomes modulate transcriptome abundance

via generalized frameshift

and out-of-frame mRNA decay

Yujie Zhang, Lilit Nersisyan, Eliska Fürst, Ioannis Alexopoulos, Carlos Santolaria, Susanne Huch, Claudio Bassot, Elena Garre, Per Sunnerhagen, Ilaria Piazza, and Vicent Pelechano

Ribosomes modulate transcriptome abundance via generalized frameshift and out-of-frame mRNA decay.

Yujie Zhang[1], Lilit Nersisyan[1,2,3], Eliska Fürst[4], Ioannis Alexopoulos[1], Carlos Santolaria[1], Susanne Huch[1], Claudio Bassot[4], Elena Garre[5], Per Sunnerhagen[6], Ilaria Piazza[4], Vicent Pelechano[1,*]

[1]SciLifeLab, Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Solna, 171 65, Sweden

[2]Armenian Bioinformatics Institute, Yerevan, Armenia.

[3]Institute of Molecular Biology, National Academy of Sciences of Armenia, Yerevan, Armenia.

[4]Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC Berlin), Berlin, Germany.

[5]Department of Laboratory Medicine, Institute of Biomedicine, Sahlgrenska Academy, Sahlgrenska Center for Cancer Research, University of Gothenburg, 41390 Gothenburg, Sweden.

[6]Department of Chemistry and Molecular Biology, University of Gothenburg 40530 Gothenburg, Sweden

*Lead contact: vicente.pelechano.garcia@ki.se
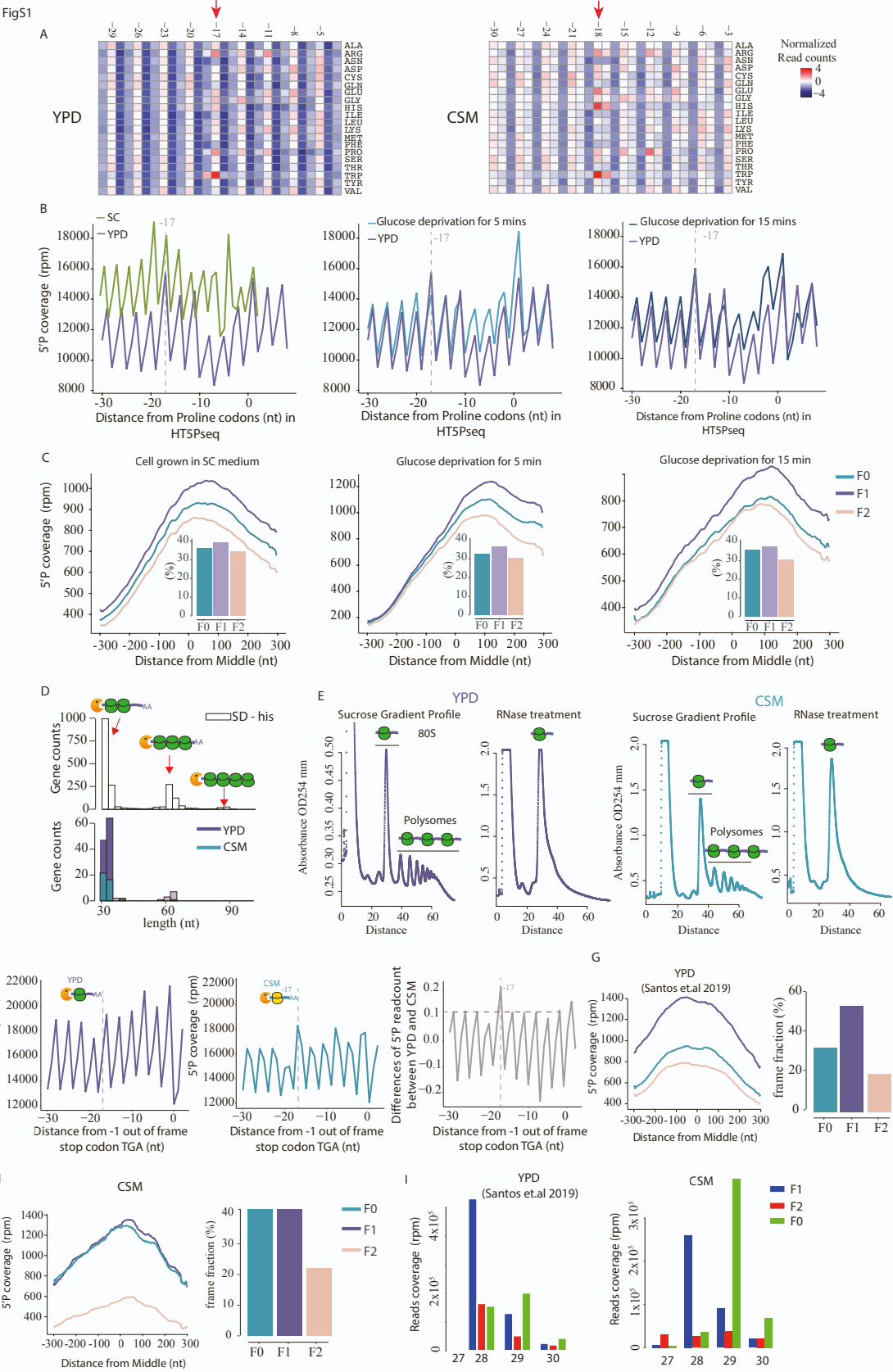
SUPPLEMENTARY FIGURES

FigS1

Figure S1. Co-translational mRNA decay reveals genome-wide -1 ribosome frameshift, Related to Figure 1 (A) Heatmap for metagene analysis displaying the 5′P reads coverage for each amino acid for YPD (left) and CSM (right). 5′P reads coverage is calculated by reads per million and the heatmap is row normalized. (B) Metagene analysis displaying the 5′P reads coverage for proline codons (CCG) in SC (in green, left), glucose deprivation for 5 mins (in blue, middle) and glucose deprivation for 15 mins (in dark blue, right) using HT-5PSeq. 5′P reads coverage for proline codons in YPD (in purple) used as control in all conditions. Dotted lines at −17 corresponding to the *in-frame* 5′ end of protected ribosome located at the A site. (C) Relative 5′P coverage with average of 20 codons for each frame in SC (left), glucose deprivation for 5 mins (middle), glucose deprivation for 15 mins (right) from the middle of genes. The fractions calculated for each frame for metagene analysis (excluding the first and last 50 nt for each gene), are shown in the histogram. (D) Number of genes where ribosome collisions are detected by FFT analysis (see methods for details). A disome provides a signal at 30nt, a trisome at 60 nt and a tetrasome at 90nt. We show values for YPD, CSM and minimal media without histidine (SD-His). SD-His is shown as a positive control for inducing disomes[8]. Higher level of ribosome collisions are observed in YPD than in CSM growth. (E) Sucrose gradient analysis of polysomes and ribosomes after RNase I digestion in YPD (left) and CSM (right). Ribosomes were separated by ultracentrifugation in a linear 10-50% sucrose density gradient. (F) Metagene analysis displaying the 5′P reads coverage for out-of-frame stop codons (*i.e.* TGA) in YPD (left), CSM (center) and normalized peak by dividing the relative peaks in CSM with in YPD (right) (See in Methods section). (G) as C, but for ribosome profiling in YPD. Data obtained from Santos *et.al* [59]. (H) as C, but for ribosome profiling in CSM generated in this study. (I) The distributions (read counts in rpm) of frames in ribosome profiling for fragment sizes ranging from 27 to 30 nt are shown for YPD (left) and CSM (right, merged three replicates).
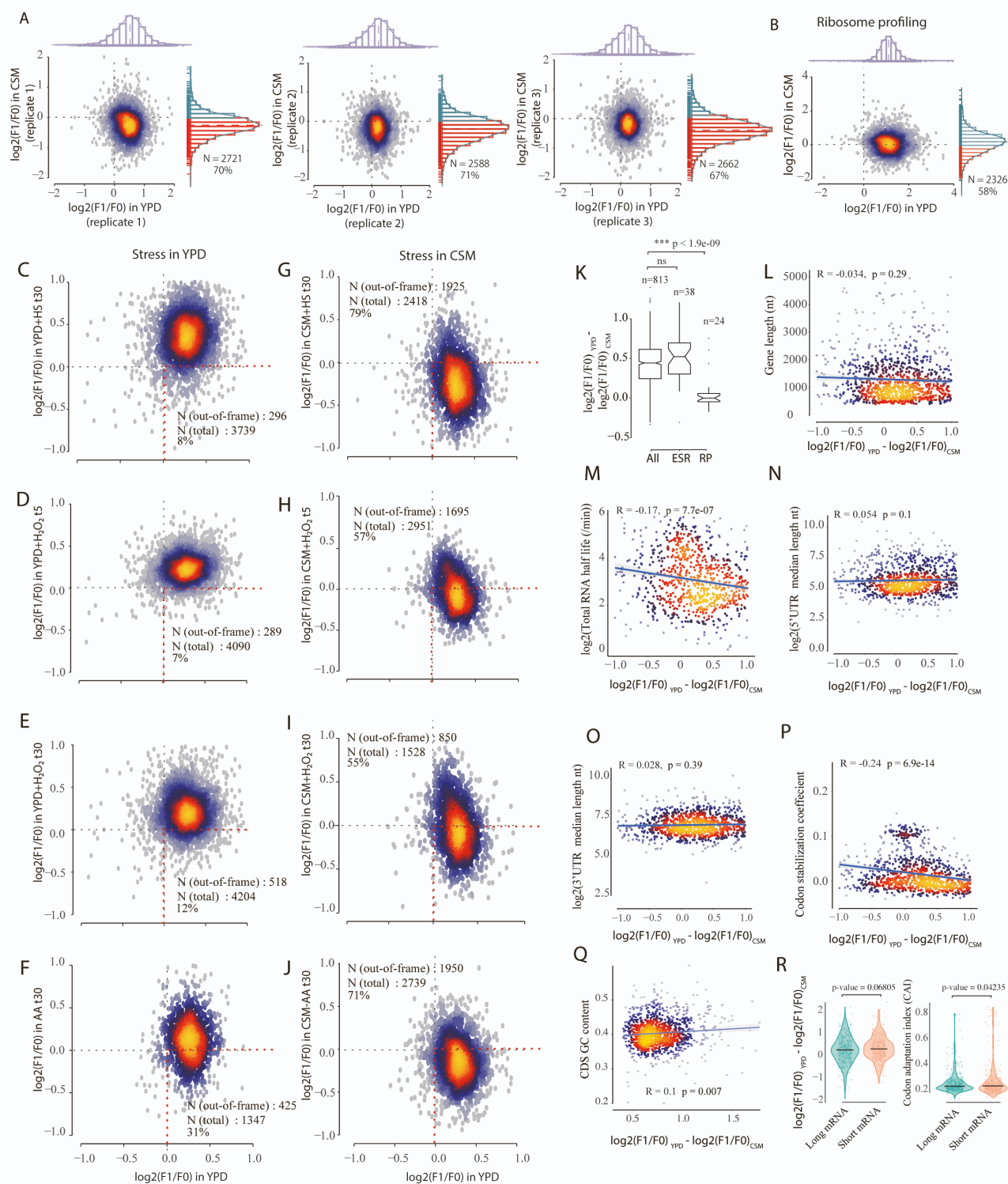
Figure S2

Figure S2. Genome-wide out-of-frame mRNA decay is environmentally regulated, Related to Figure 2. (A) Scatter plot comparing the frameshift index (log₂(F1/F0)) of

individual genes in YPD (x-axis) and CSM (y-axis) among three replicates. Histograms marked with red highlight genes that exhibited a mainly out-of-frame decay. Namely those genes with $\log_2(F1/F0) < 0$ in each replicate. (B) as (A), but for ribosome profiling. (C-F) as A, but for stress in YPD, *i.e.* (C) heat shock for 30 mins; (D) H_2O_2 (0.2 mM) exposure for 5 mins; (E) H_2O_2 (0.2 mM) exposure for 30 mins; (F) transfer from YPD to CSM medium lacking amino acids for 30 mins; (G-J) as C-F, but for cells grown long term in CSM. (K) Boxplot comparing frameshift index change for genes been detected among all stresses (All, N = 813), environmental response genes (ESR, N =38) and ribosomal protein genes (RP, N = 24) in Fig2C. Statistical analysis was performed using two-sided Wilcoxon rank-sum tests. (L-P) Spearman correlations between frameshift index change (x-axis) with RNA features, i.e. (L) gene length (nt); (M) Total RNA half-life in log scale (from Xu *et al* [60]) ; (N) 5′UTR length (from Pelechano *et.al* [61]) ; (O) 3′ UTR length (from Pelechano *et.al* [61]) ; (P) Codon stabilization coefficient (from Presnyak *et al* [10]). (Q) GC contents in CDS (from Latorre *et al* [62]). (R) Comparison of frameshift index (left panel) and codon adaptation index (right panel) between long and short genes within region 300 to 600bp. Long mRNAs are defined as the longest 20% (≥1934 nt, N=432) and Short RNA as the shortest 20% (≤752 nt, N=432).
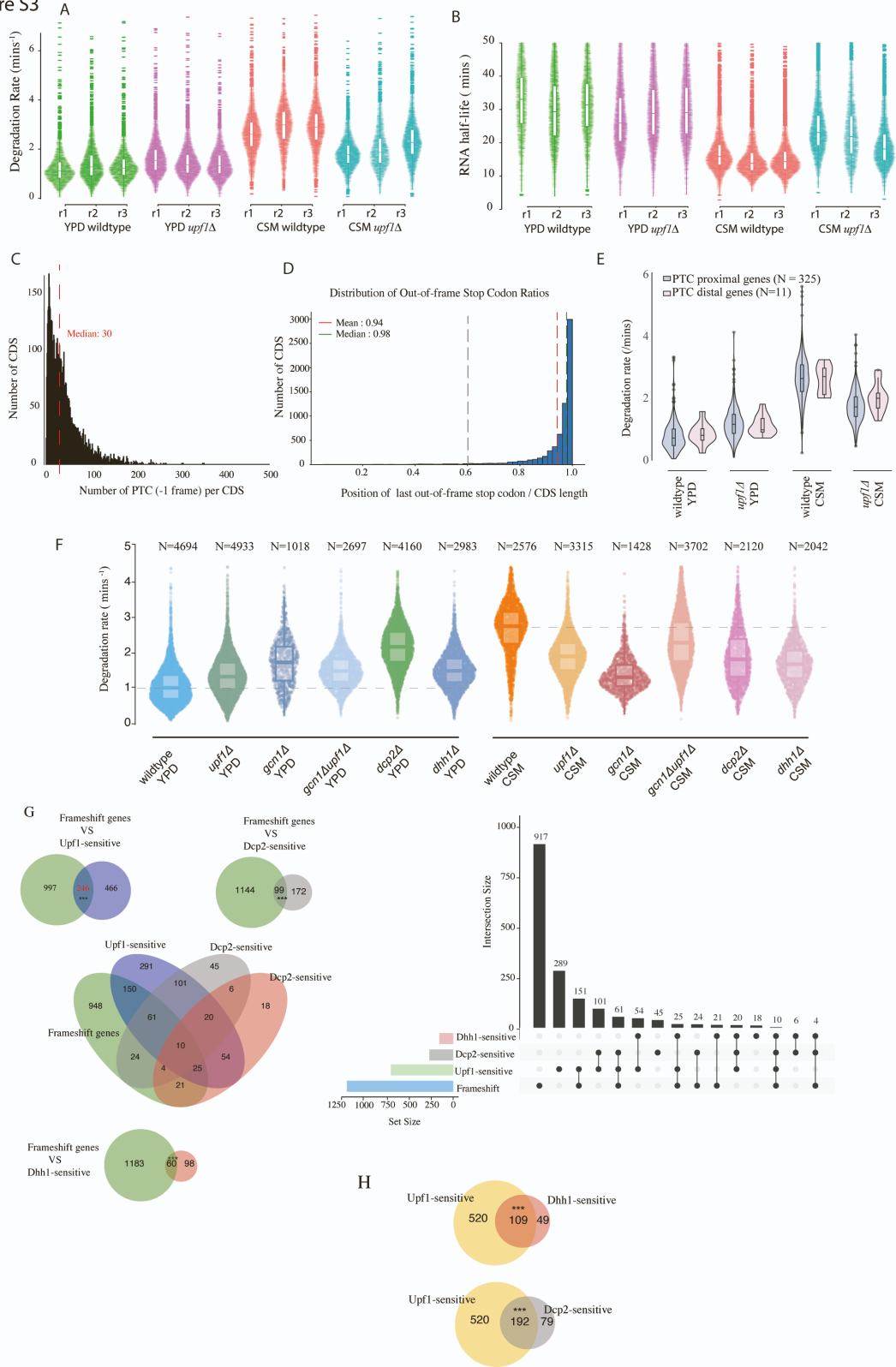
Figure S3. NMD and frameshifting-dependent decay in low nutrients conditions, Related to Figure 3. (A) Violin plot comparing the degradation rate (mins $^{-1}$) between

wildtype (BY4741) and NMD mutant (*upf1Δ*) both in YPD and CSM across all three biological replicates. Only reads containing at least 1 T>C conversions were considered as labelled reads. Only coding mRNA with at least 20 total reads were considered for RNA degradation calculations. (B) as (A), but displayed as RNA half-life (in mins). (C) Distribution of the number of premature termination codons (PTC) in frame -1 per CDS in the *S. cerevisiae* genome. The average distance between two PTC in the –1 frame is 32bp (D) Distribution of the position of the last -1 out-of-frame stop codon versus CDS length. Assuming a uniform probability of frameshift along the CDS length, 94% of frameshifts will result in recognizing a PTC. (E) Degradation rate according to the likelihood that a frameshift event will lead to the recognition of downstream PTC. Gene groups as in Fig 3G. (F) Violin plot comparing the median degradation rate (mins^{-1}) for wildtype (BY4741), *upf1Δ*, *gcn1Δ*, *gcn1Δ upf1Δ*, *dcp2Δ*, *dhh1Δ* both in YPD and CSM across biological replicates (n >=2). Only reads containing at least 1 T>C conversions were considered as labelled reads. Only coding mRNA with at least 20 total reads were considered for RNA degradation calculations. (G) Analysis of overlap between genes exhibiting mutant specific environment-dependent mRNA decay regulation and genes showing environment-dependent frameshifts (log2(F1/F0) $_{YPD}$ > 0.2 and log2(F1/F0) $_{CSM}$ < -0.2) (shown in green) (left). The intersection of all factor-sensitive transcripts is shown on the right. Specific environment -sensitive transcripts were defined based on their degradation rate between wild-type and mutant strains under different growth conditions: (wildtype/mutant)$_{CSM}$ > 1.2 and (wildtype/mutant)$_{YPD}$ < 0.8. Upf1-sensitive (purple), Dcp2-sensitive (grey), and Dhh1-sensitive (red) transcripts were identified using these criteria. Statistical significance of overlap between each factor-sensitive group and frameshift genes was assessed using Fisher's exact test (n = 5,375) *** p < 0.001. (H) Overlap between Upf1- and Dcp2-sensitive genes, defined by their degradation rate ratios as in G. Statistical significance was assessed using Fisher's exact test (n = 2722) *** p < 0.001.
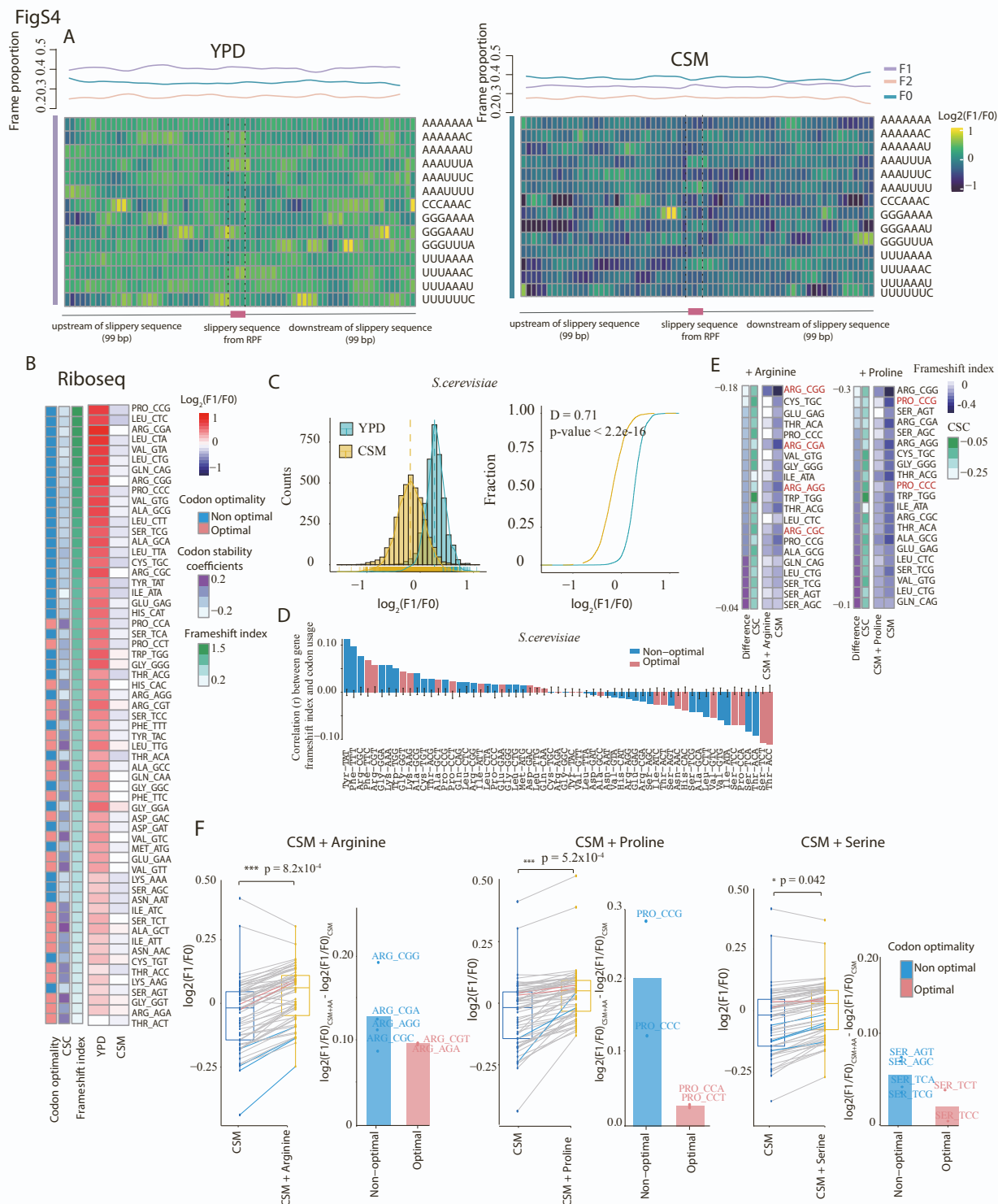
Figure S4. Codon optimality controls out-of-frame mRNA decay, Related to Figure 4. (A) Heatmap comparing 5PSeq coverage across three frames (F0, F1 and F2) with respect to slippery sequences in YPD (left) and CSM (right). Slippery sequence regarding to programmed ribosomal frameshift (PRF) were obtained from PRFdb[23]

(Supplementary Data 4). The average proportion of each frame are showed with sliding window of 3 nucleotides extended to 99 nt both upstream and downstream of slippery sequence. (B) Heatmap comparing ribosome profiling coverage at $\log_2$ ratio of -15 (corresponding to F1) and -16 (corresponding to F0) relative to the A-site of codons in YPD (Data from Santos, *et.al* [59]) and CSM (from this study). The codons ordered by the differences of frameshift index between YPD and CSM. Codon stability coefficients (CSC) and Codon optimality were obtained from Presnyak *et al.*[10]. (C) Density plot comparing gene-specific frameshift index distributions between *S. cerevisiae* growing in YPD and in CSM using 5PSeq (left). A cumulative fraction plot with the frameshift index distribution is shown (right). Statistical analysis was performed using Kolmogorov–Smirnov test with P-value and D as effect size. (D) Correlation coefficients (Pearson) between frameshift index and codon compositions among frameshift genes ($\log_2$(F1/F0) control > 0.2 and $\log_2$(F1/F0) treatment < -0.2) in *S. cerevisiae*. Optimal codons and non-optimal codons are showed in red and blue, respectively. (E) Heatmap comparing the change in codon frameshift index for the top 30% of codons, when supplemented with arginine (from 50 to 85.6 mg/L) (left) and proline (from 0 to 85.6 mg/L) (right) as compared to CSM alone. The codons are ordered by the differences of frameshift index between CSM and the respective amino acid additions. (F) Boxplot comparing frameshift index change per codons between CSM and the respective amino acid additions (left). Barplot comparing the differences in frameshift index change between non-optimal and optimal codons for each amino acid addition (right). The codons highlighted in blue (non-optimal) and red (optimal) corresponding to respective amino acids (arginine, CGG/CGA/CGC/AGA/AGG/CGT; proline, CCT/CCC/CCA/CCG; serine, TCT/TCC/TCA/TCG/ AGT/AGC). Statistical analysis was performed using two-sided t tests.
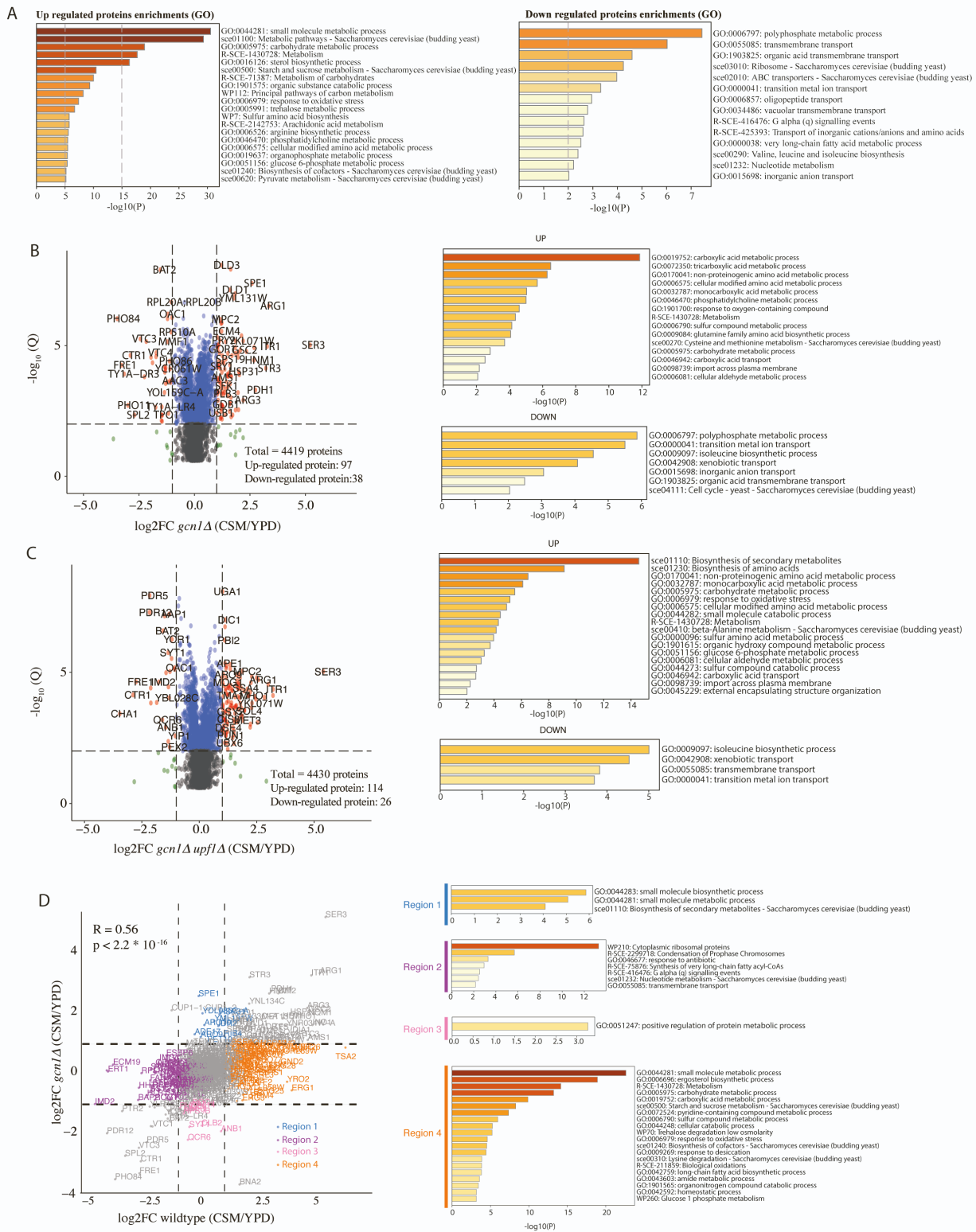
FigS5

Figure S5. Proteomics change upon long-term amino acid limitation, Related to

Figure 5. (A) Gene ontology terms for up-regulated and down-regulated protein

abundance. Only statistically significant enrichments are showed (p-adj < 0.01). (B) Volcano plot comparing protein abundance change in *gcn1Δ* mutation between YPD and CSM in log2 fold change (x-axis) versus the statistical significance (-log10Q) (left). Gene ontology terms for up-regulated and down-regulated protein abundance (right). (C) as (B), but for *gcn1Δupf1Δ* mutant. (D) Comparison of protein abundance changes (fold change in CSM versus YPD) between wildtype (x-axis) and *gcn1Δ* (y-axis) strains.
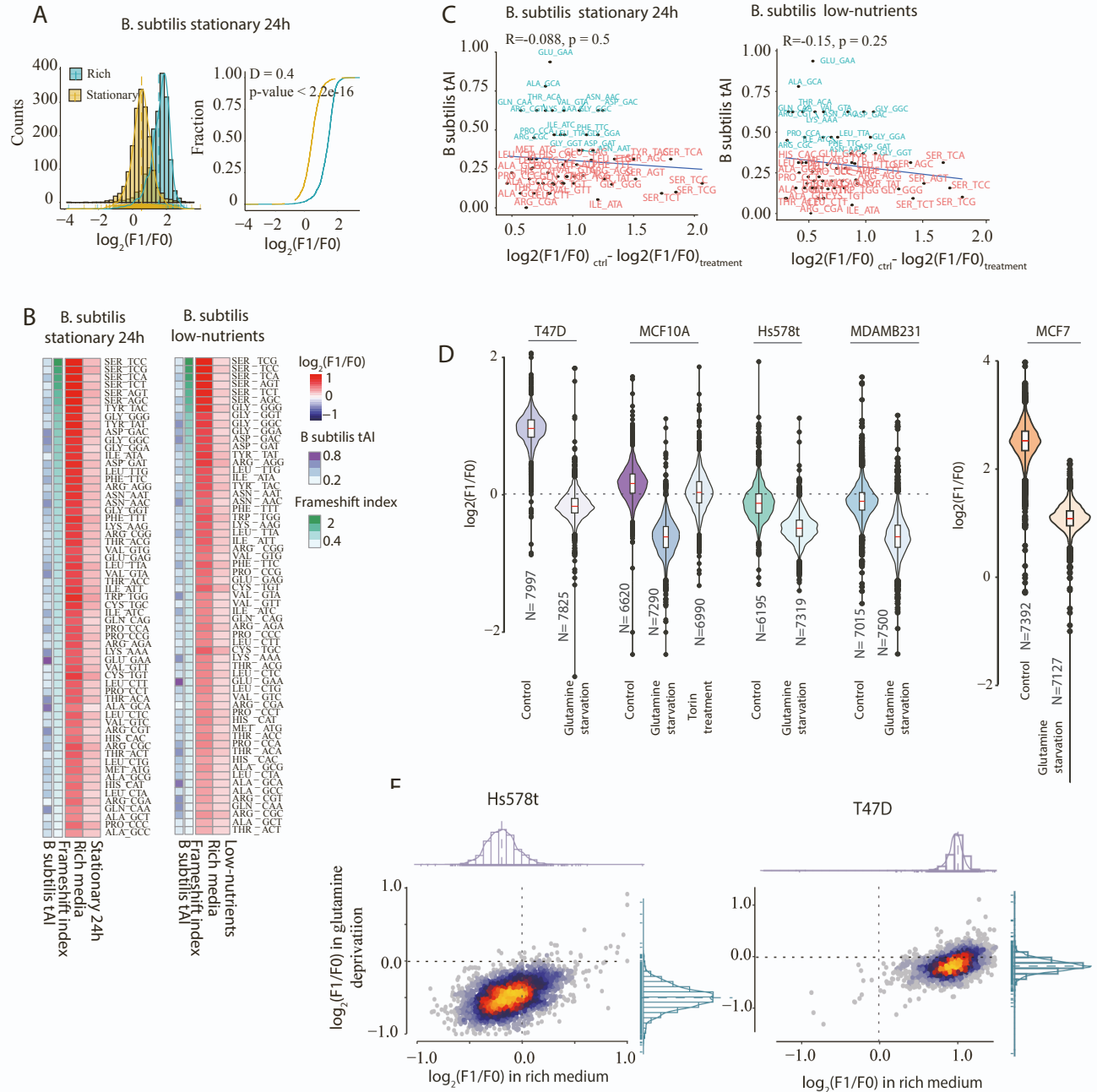
Figure S6. Out-of-frame co-translational decay of mRNA is evolutionary conserved, Related to Figure 6. (A) Density plot comparing frameshift index distributions between–*B. subtilis* growing in stationary for 24h and in rich medium (LB). Data obtained from Huch, *et.al.*[11] (B) Heatmap comparing 5PSeq coverage at $\log_2$ ratio of -14 (corresponding to F1) and -15 (corresponding to F0) relative to the A-site codons

for *B. subtilis* growing in LB and in stationary phase for 24h (left). *B. subtilis* growing in poor minimal growth medium (minimal growth medium) (Details in Method section). (C) Scatterplot comparing frameshift index (x-axis) and tRNA adaptation index (y-axis) for *B. subtilis* growing in stationary for 24h (upper) and *B. subtilis* growing in poor minimal growth medium (below). The correlation between the two datasets was evaluated using Spearman correlation analysis. Optimal codons and non-optimal codons are represented in red and blue, respectively. (D) Frameshift index distributions across different breast cancer cell lines MCF10A, MCF7, T47D, MDA-MB-231, and Hs578T measured by Ribosome profiling. Data obtained from Loayza-Puch,*et.al*[43]. A generalised frameshift can be observed for all cell lies after glutamine deprivation. The red line in box plot indicates the median gene frameshift index. (E) Scatter plot comparing the frameshift index ($log2(F1/F0)$) of individual genes in rich DMEM medium (x-axis) and glutamine-deprived conditions (y-axis). Namely those genes with $log2(F1/F0) < 0$ in each condition. The upper panel in corresponds to the Hs578t cell line, while the lower panel corresponds to the T-47D cell line.

SUPPLEMENTARY TABLES

Table S1, **Related to Figure 1.** Growth medium composition in CSM and SC.

Table S2, Related to Figure 2. Frameshift index in YPD, CSM and stress conditions**.**

Table S3, Related to Figure 3. RNA half-life measured by SLAM-Seq in wildtype and mutant.

Table S4, Related to Figure 4. Codon frameshift index measurement in 5PSeq and Riboseq

Table S5, Related to Figure 5. MS measured canonical protein abundance.

Table S6, Related to Figure 6. Gene-specific frameshift index in bacteria and human.

Table S7, Related to Figure 7. Cell growth in CSM and YPD.