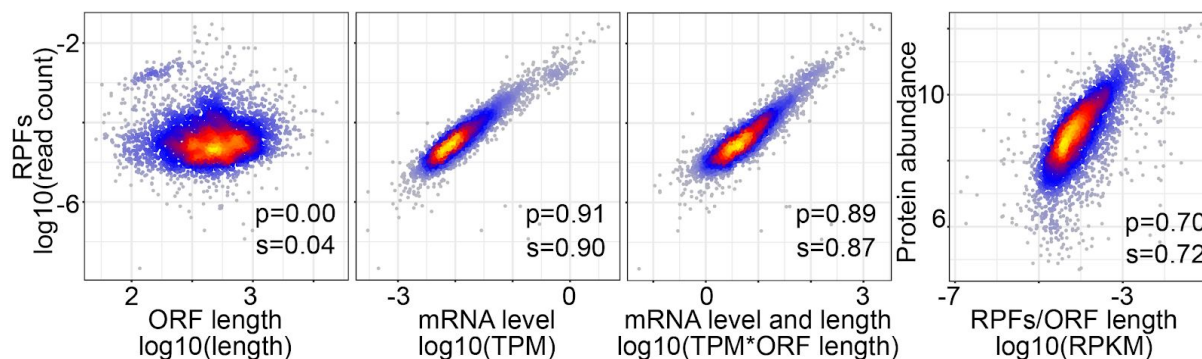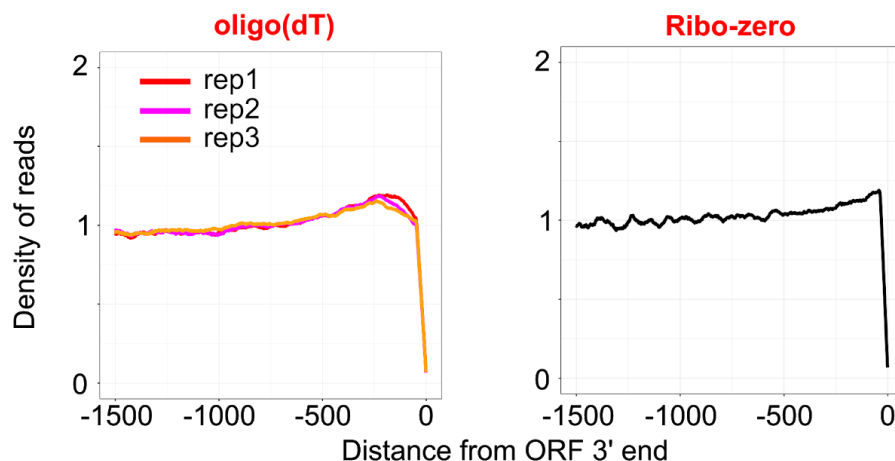# Supplementary Figures

Figure S1. Relationship between the number of RPFs, the mRNA levels and ORF lengths inferred from the mRNA-seq and Ribo-seq data from (4). Counterintuitively, the mRNA level alone explains more of the variance in RPF numbers than the combination of mRNA level and ORF length.


Figure S2. Relative coverage of transcripts by reads obtained from sequencing RNAs with two distinct protocols: oligo(dT) vs. Ribo-zero-based purification, as a function of the distance from the stop codon (ORF end). The coverage of each position in individual ORFs by reads was normalized by the average coverage over the entire ORF, then ORFs were anchored at the 3 end, and, for each position relative to ORF end, the mean relative coverage was computed, based on all ORFs that extended at least that far upstream from the stop. Values are shown relative to the mean coverage per position along the ORF.
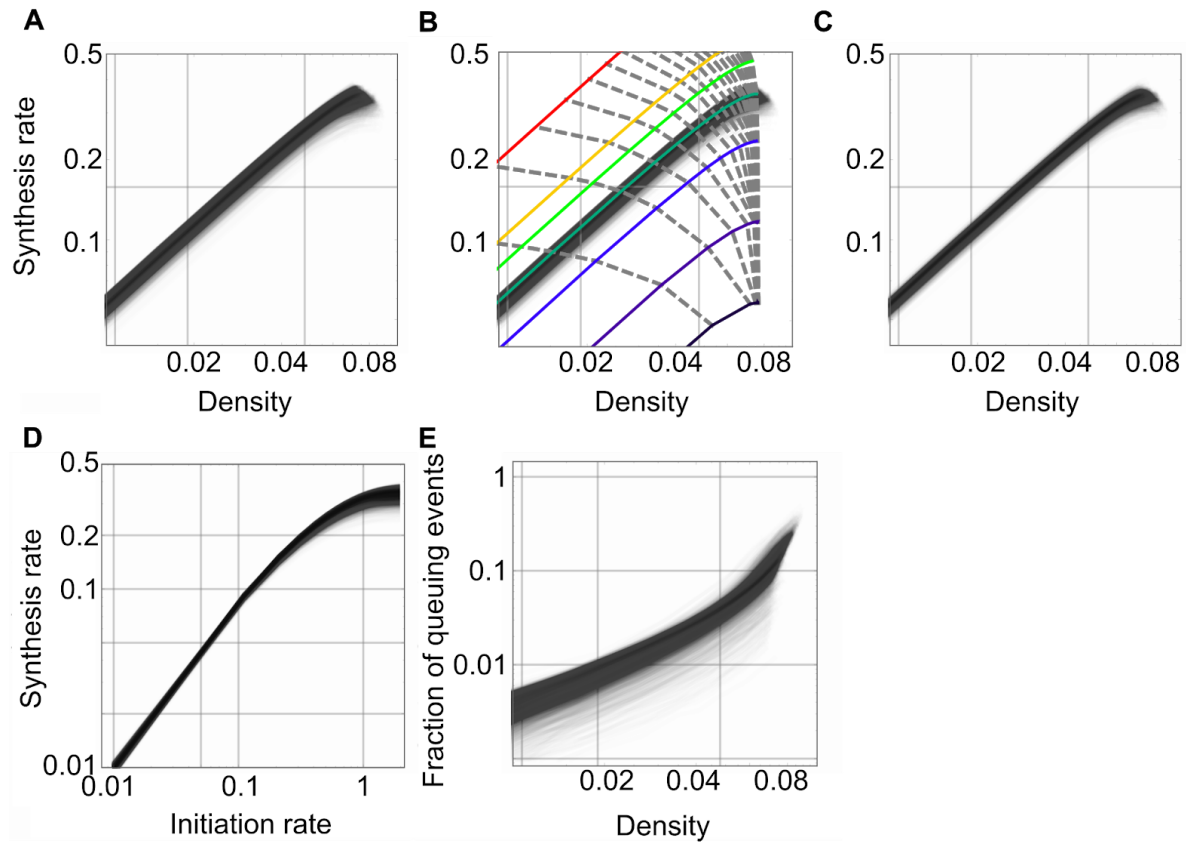
Figure S3. Inhomogeneous TASEP simulations with codon-specific speeds computed from Ribo-seq-based codon densities. **(A)** Ribosome density vs protein synthesis rate per mRNA. **(B)** Similar to (A) but with elongation and initiation isoclines superimposed. **(C)** Ribosome density vs protein synthesis rate normalized by average codon speed. Part of the variability in (A) can be explained by the average elongation rate, leading to less variability compared to (A). **(D)** In the model, the relationship between translation initiation and protein synthesis rate per mRNA deviates from linearity only when the initiation rate is very high and ribosomes queue on the transcript, as shown in (E). **(E)** Fraction of queuing events as a function of the ribosome density on the transcript.
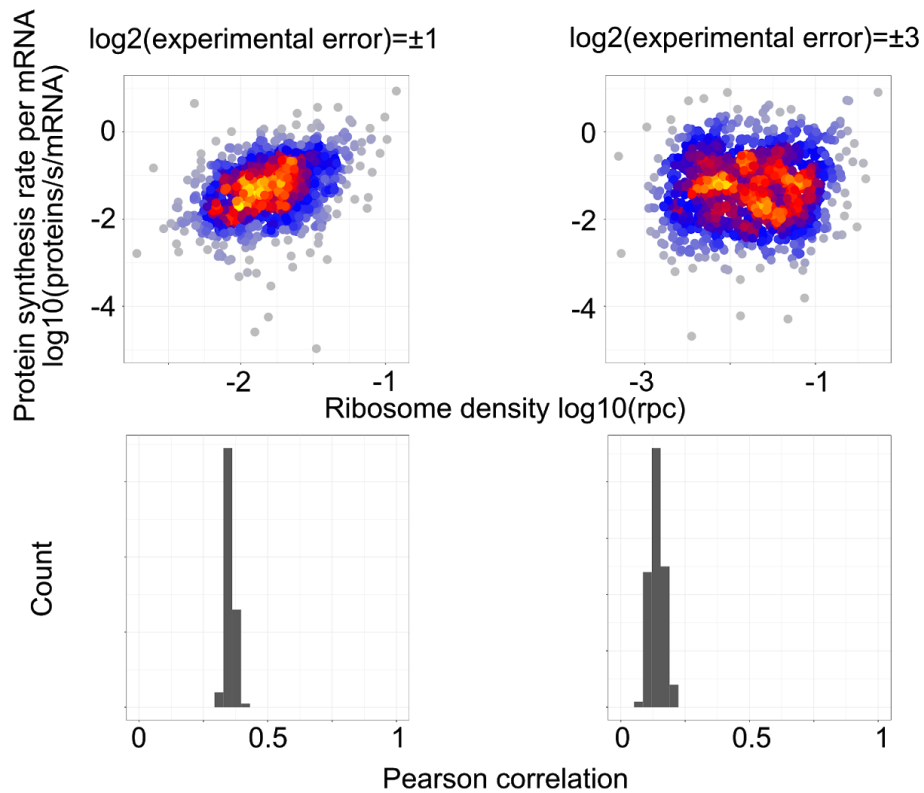
Figure S4. Effect of 'experimental error' on the density vs synthesis scatter. An uncorrelated and uniformly distributed log2(error) has been added to both ribosome densities and protein synthesis rates. On the top panels, density vs. synthesis plots are shown for two different ranges of log2(error) values, [-1,1] (left) and [-3,3] (right). Bottom panels show the respective distributions of synthesis-density correlations over 100 samplings.
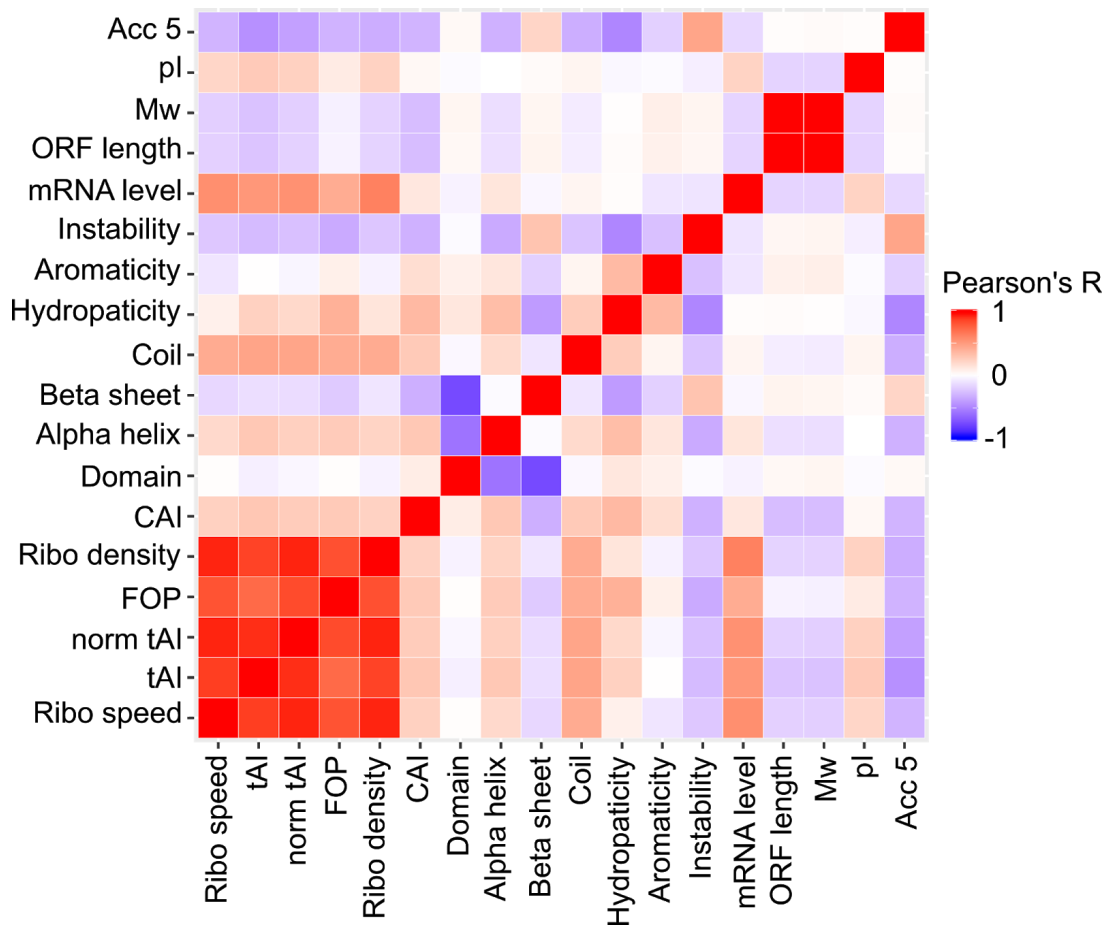
Figure S5. Heatmap of Pearson correlation coefficients of pairs of features of transcripts/proteins investigated in this study.
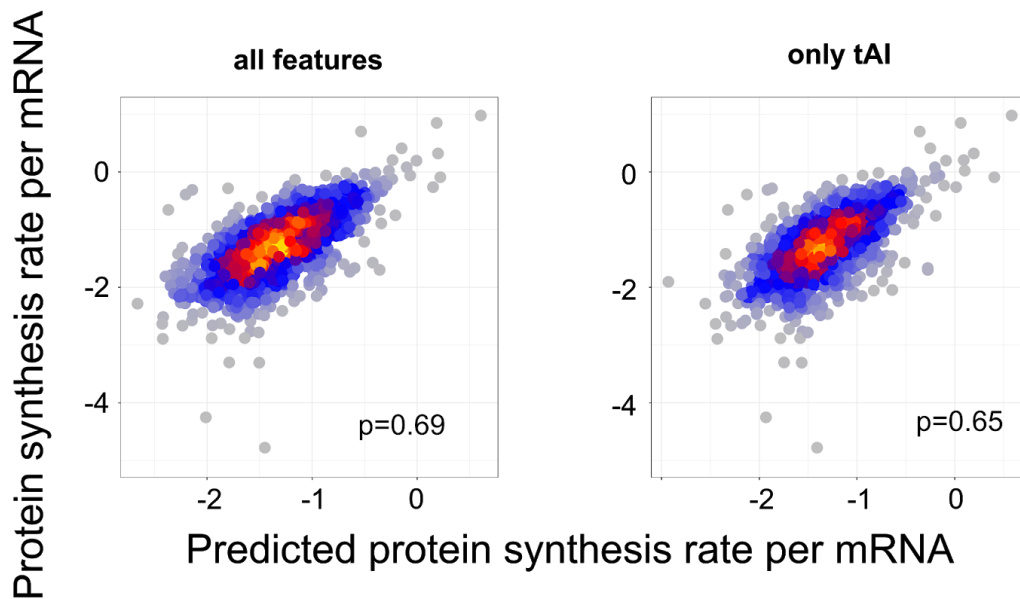

Figure S6. Linear model prediction of protein synthesis rates based on all features in Supplementary Figure 5 (left panel) and all features except for normalized tAI, CAI, FOP (right panel).
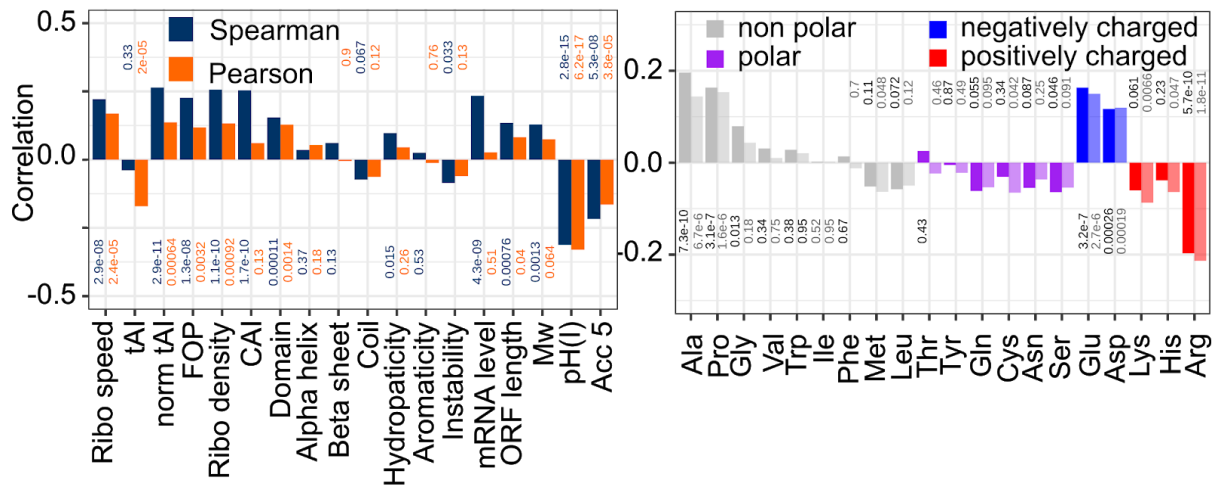
Figure S7. Correlation coefficients between the synthesis-density ratio and features related to codon speed, biochemical properties of the encoded protein and RNA accessibility (left) or amino acid frequencies in the encoded proteins (right), based on Ribo-seq and RNA-seq data from (4). Spearman (blue in left and dark shade in right panels) and Pearson (orange in left and light shade in right panels) correlation coefficients are indicated on the y-axis, corresponding p-values are written on the plots. Ribo-seq and Ribo-zero data sets from (4) were used to obtain the ribosome densities, the pSILAC data generated in this study was used to compute protein synthesis rates.
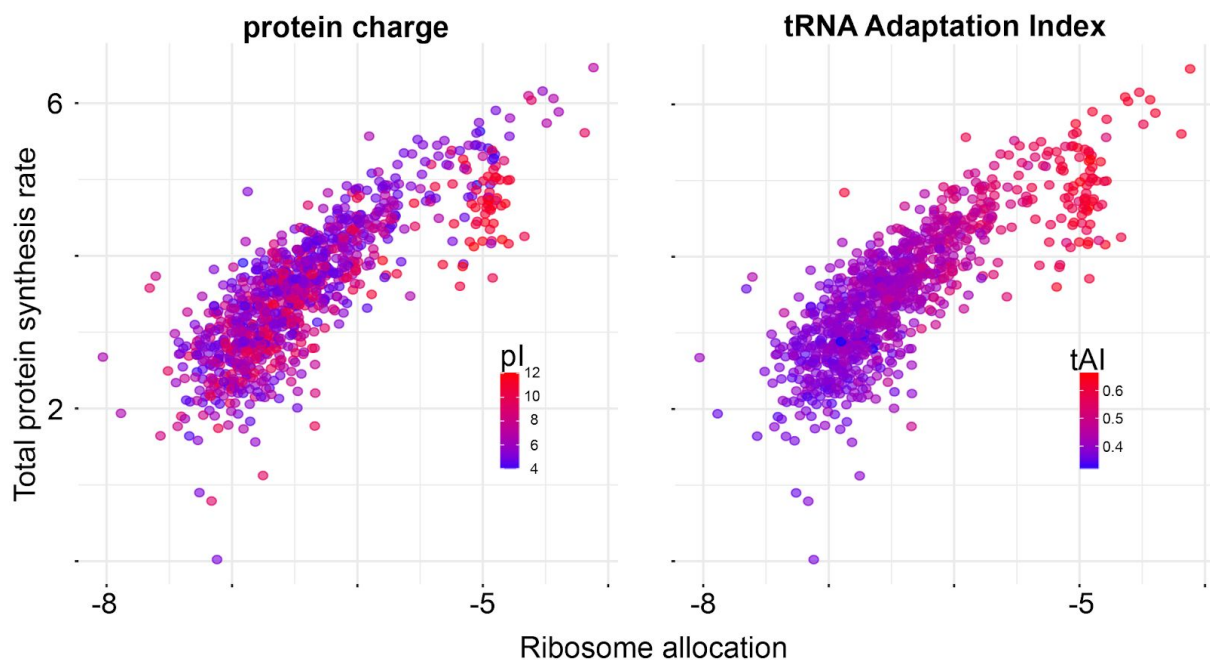

Figure S8. Total protein synthesis rate in function of the number of ribosomes allocated to translation of each mRNA species. Each point represents a transcript, the color indicating isoelectric point of the encoding protein (left) or the tRNA adaptation index (right).
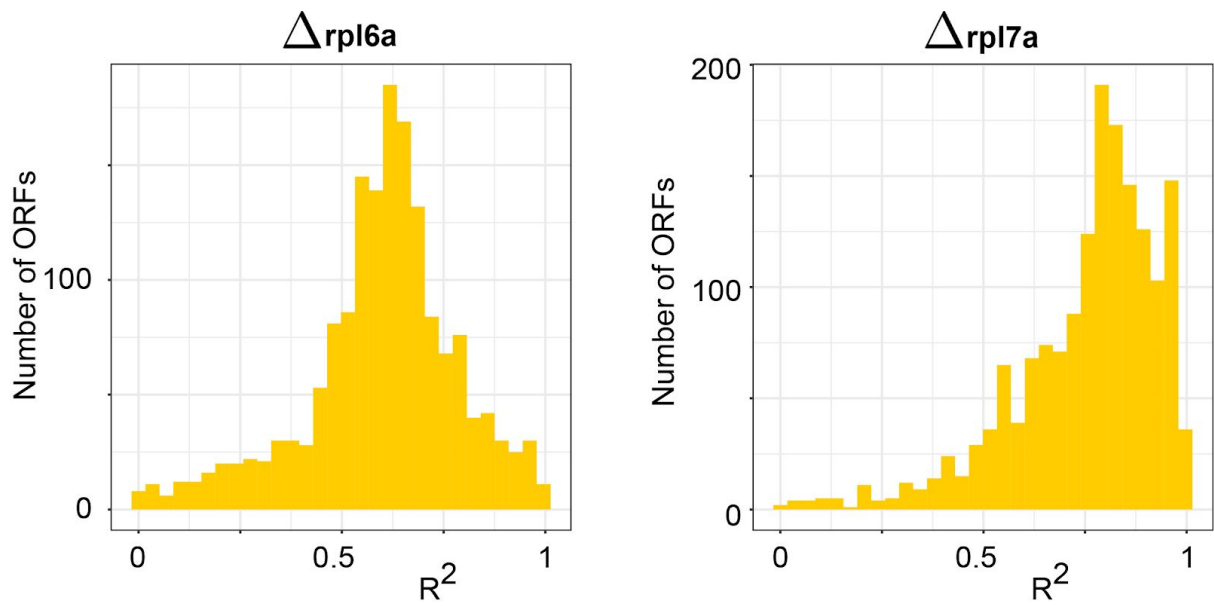
Figure S9. R^2 values for the linear fit to peptide accumulation data from *Δrpl6a* and *Δrpl7a* strains.
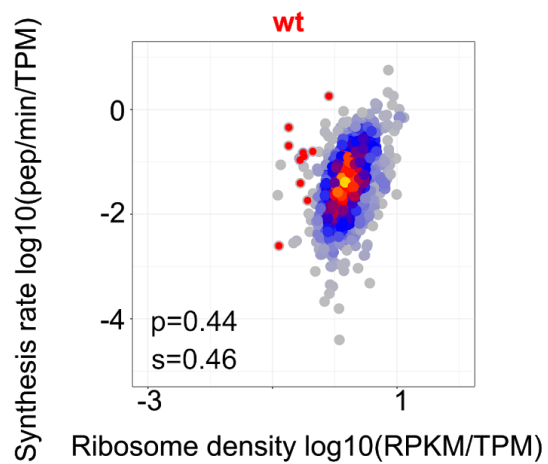

Figure S10. Synthesis-density plot for the wt strain indicating those genes for which the ribosome densities were lower than 0.01 RPKM/TPM in the *Δrpl7a* strain (in red, see also Figure 7).
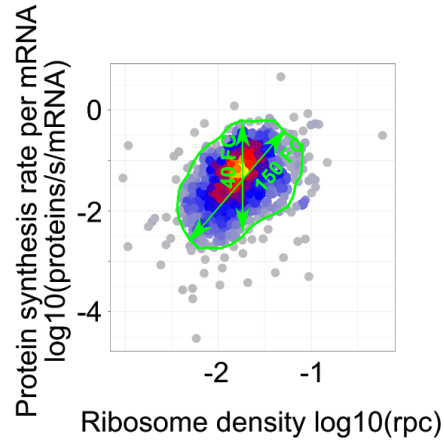
Figure S11. Absolute synthesis-density plot based on ribosome footprints and RNA-seq (Ribo-zero) data from (4) (see also Figure 3).
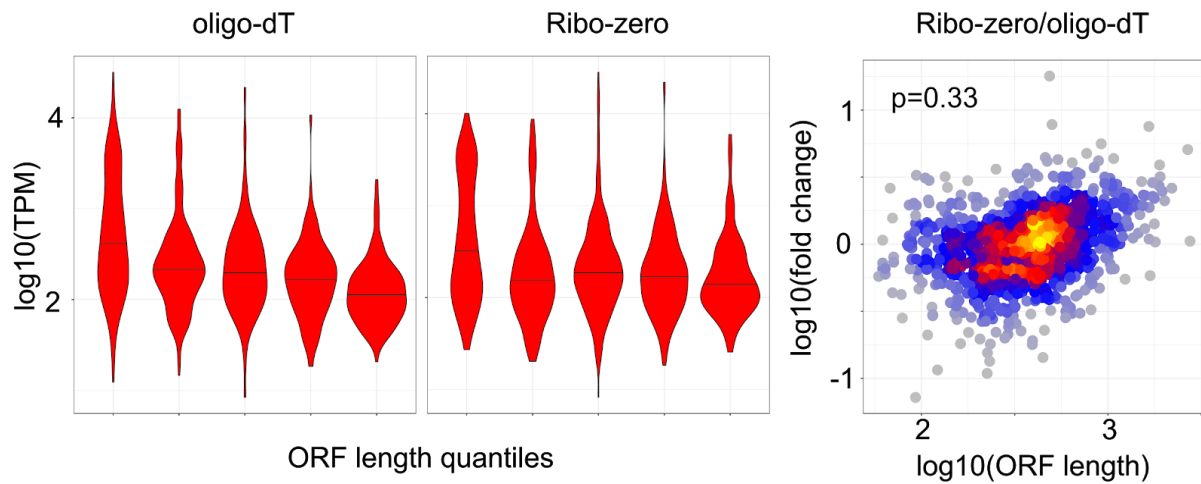
Figure S12. ORF-length-dependent bias in transcript capture rate with Ribo-zero vs. oligo(dT) selection. The plots show the distribution of inferred expression levels of ORFs in bins of ORF length, each bin containing 20% of the transcripts. Bins are sorted from shortest (left) to longest (right) ORF. The log10 ratio of expression levels estimated with Ribo-zero relative to oligo(dT), as a function of ORF length. The bias is apparent for long ORFs, whose expression estimates are systematically higher when the Ribo-zero protocol is used than when oligo(dT) is used.
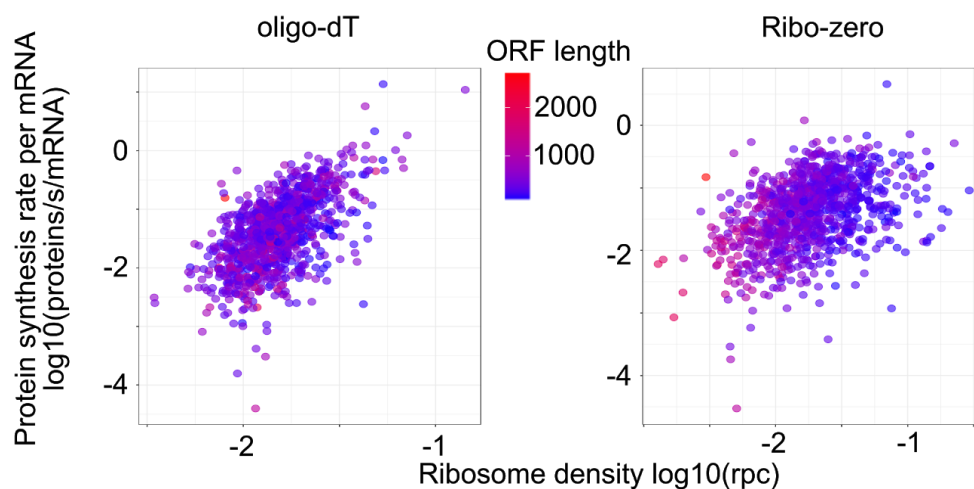
Figure S13. Influence of ORF length on the protein synthesis rate - ribosome density relationship. The length of ORFs (in nucleotides) is indicated on the color scale. ORF length is positively correlated with the coordinate of the ORF along the first principal component of the synthesis - density data. Pearson correlation coefficients are -0.09 (p-value = 0.0059) in the oligo(dT)-based analysis and -0.32 (p-value < 2e-16) in the analysis based on the Ribo-zero protocol.
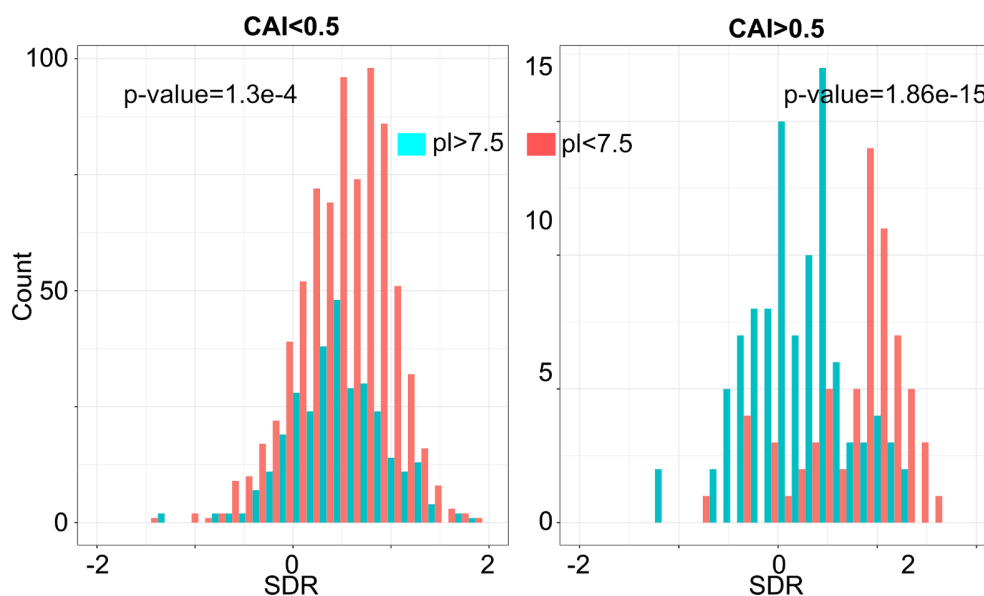
Figure S14. Comparison of SDR distributions for proteins with pI < 7.5 (red) and >7.5 (cyan), for two classes of transcripts: those with low (< 0.5, n=1093) CAI (left panel) and those with high (> 0.5, n=172) CAI (right panel).

Supplementary Datasets

Dataset S1. Ribosome densities and synthesis rates with RPF from (4) together with oligo(dT) RNA-seq from (17) for wild type condition.

Dataset S2. Ribosome densities and synthesis rates with RPF and Ribo-zero from (4) for wild type condition.

Dataset S3. Estimated features from the linear models.

Dataset S4. Ribosome densities (based on data from (17)) and synthesis rates for wild type condition.

Dataset S5. Ribosome densities (based on data from (17)) and synthesis rates for *Δrpl6a* mutants.

Dataset S6. Ribosome densities (based on data from (17)) and synthesis rates for *Δrpl7a* mutants.

Dataset S7. Outliers, with log10 decrease in ribosome density < -2 in *Δrpl7a* mutants.

Dataset S8. Gini indexes for ORFs calculated from (50).