Supplemental Information

Contents

1 Supplemental Experimental Procedures

1.1 Simulation model

We use a whole-cell model of protein translation to simulate the dynamics of protein production (Shah et al., 2013). Briefly, the model assumes a fixed total number of ribosomes and tRNAs, and it describes how these entities initiate and elongate a fixed supply of mRNAs.

We define a genome with $n = 4862$ genes, each with a prescribed coding sequence, and fixed mRNA abundance A_i. Gene i encodes an mRNA of length L_i codons and has a corresponding probability of translation initiation, denoted p_i, which is described below.

Each codon of type j is decoded by one of 41 iso-accepting tRNA species, denoted $\phi(j)$, which has a fixed total abundance $T^t_{\phi(j)}$ in the cell. Each molecule of tRNA species $\phi(j)$ is either free in the cell, or bound, along with a ribosome, to a codon of type j on an mRNA in the cell. Thus, the total number of tRNAs of type $\phi(j)$ can be decomposed into those that are currently bound and those that are currently free: $T^t_{\phi(j)} = T^b_{\phi(j)} + T^f_{\phi(j)}$. Similarly, the total number of ribosomes, R^t, can be decomposed into bound and free: $R^t = R^b + R^f$. Moreover, the number of bound ribosomes always equals the total number of bound tRNAs of all species: $R^b = \sum_{k=1}^{41} T^b_k$.

Initiation and elongation events in the cell occur at rates that are determined by the current state of system (the number of free ribosomes, and the locations of all bound ribosomes) and by the underlying physical parameters of the cell. The underlying physical parameters are simply the volume of the cell, and the characteristic lengths and diffusion constants of ribosomes and tRNA molecules. The time between subsequent events are exponentially distributed, and Monte Carlo simulations proceed by incrementing time according to exponential deviates and re-computing rates of subsequent events (Gillespie, 1977).

1.1.1 Diffusion of ribosomes and tRNAs

In a cell of fixed volume, the average time required for any given molecule to move to one position, known as the characteristic time τ, is given by

$$\tau = \frac{\lambda^2}{6D} \tag{S1}$$

where D is the diffusion coefficient of the molecule and λ is its characteristic length. The characteristic times of tRNAs and ribosomes are $\tau_t = 4.45 \times 10^{-7}$ s and $\tau_r = 5 \times 10^{-4}$ s, respectively (Shah et al., 2013).

1.1.2 Translation initiation rates

Let ρ_i be the initiation rate at an mRNA of gene i. The rate ρ_i is set to zero if any of the first 10 codons of the mRNA is currently bound by a ribosome. Otherwise, the rate is

$$\rho_i = p_i \frac{R^f}{\tau_r N_r}. \tag{S2}$$

The term $\frac{R^f}{\tau_r N_r}$ in this equation denotes the rate at which free ribosomes (R^f) diffuse to a given mRNA molecule. And the term p_i denotes the probability with which a ribosome will actually initiate translation of an mRNA molecule, once it has diffused to its 5' end. The parameters p_i allow us to account for sequence-specific variation in initiation rates among genes.

1.1.3 Translation elongation rates

Consider a ribosome bound at codon of type j at position k on an mRNA. Its elongation rate is set to zero if any of the following $k + 10$ codons of the mRNA are currently occupied by another ribosome. Otherwise, the elongation rate depends on the number of free cognate tRNAs for that codon $T^f_{\phi(j)}$ and the wobble parameter associated with the tRNA-codon pair w_j. If there is a perfect match between the tRNA and the codon, then $w_j = 1$. Else $w_{ry/yr} = 0.64$ if the mismatch is due to a purine-pyrimidine wobble or $w_{rr/yy} = 0.61$ if the mismatch is due to purine-purine or pyrimidine-pyrimidine wobble (Curran & Yarus, 1989; Lim & Curran, 2001). The elongation rate

is thus given by

$$\frac{T^f_{\phi(j)} w_j}{\tau_t N_t} \tag{S3}$$

In addition, the time spent by a ribosome in selecting the cognate tRNA depends on the relative abundances of various competing tRNAs as well as organism specific kinetic rates associated with ribosomal proofreading. Using the method described in (Shah et al., 2013) we estimate the average time s spent by the ribosome in kinetic proofreading to select the correct tRNA. As a result, accounting for tRNA competition, the actual elongation rate of a codon is

$$\frac{T^f_{\phi(j)} w_j s}{\tau_t N_t} \tag{S4}$$

1.1.4 Translation termination

We assume that translation termination is an instantaneous event that occurs immediately after elongation of the last codon at position L. Upon termination the pool of free ribosomes and free tRNAs corresponding to the codon j' at position $L-1$ each increases by 1 ($R^f \rightarrow R^f + 1$; $T^f_{\phi(j)'} \rightarrow T^f_{\phi(j)'} + 1$) .

1.1.5 Estimating initiation probabilities using ribosome-profiling data

We use analytical approximation of the whole-cell simulation model (Shah et al., 2013) described above to estimate the gene-specific probability of translation initiation once a free ribosome reaches the 5' end of an mRNA. The gene-specific initiation probability p_i is given by Eqn. 27 in (Shah et al., 2013) as follows:

$$p_i \approx \frac{R^b_i x}{A_i L_i \left(\sum_{j=1}^{61} \frac{u_{j,i}}{w_j T^t_{\phi(j)}} \right)} \tag{S5}$$

where

$$x = \frac{\tau_r N_r s}{0.15 R^t \tau_t N_t}$$

The term x depends on the bio-physical parameters of tRNAs, ribosomes and volume of an yeast cell, whose values are described above. The term $\frac{R_i^b}{A_i L_i}$ describes the density of ribosome on an mRNA of gene i and is equivalent to the translation efficiency (TE) described as:

$$\text{TE} = \frac{\text{RPF RPKM}}{\text{mRNA RPKM}} \tag{S6}$$

Thus we estimate the gene-specific initiation probability from ribosome profiling data by plugging in the experimentally determined TE ratios in Eqn. S5.

1.2 Elongation arrest by cycloheximide

In ribosome profiling experiments, ribosomes are often stabilized by addition of chemicals that arrest elongation to prevent run-offs during sample preparation. Cycloheximide (CHX) is usually the preferred elongation inhibitor (Ingolia et al., 2009; Zinshteyn & Gilbert, 2013; Brar et al., 2012; Gerashchenko et al., 2012; Artieri & Fraser, 2014b; McManus et al., 2014). However, it is unclear whether addition of CHX biases estimates of ribosome densities on mRNAs and hence subsequently affects inferences of protein translation dynamics based on these estimates. To explore how addition of CHX affects ribosome densities, we simulate protein translation in a cell by modeling the action of CHX.

CHX arrests ribosomes on mRNAs by binding with a tRNA in the E-site of the ribosome (Schneider-Poetsch et al., 2010). However, the E-site of a ribosome is almost always empty except for a short period immediately following an elongation and translocation event (Chen et al., 2011). As a result, upon addition of CHX to the cell, a recently elongated ribosome becomes a potential target for CHX. We model the action of CHX by assuming that whenever a ribosome elongates a codon, there is a constant probability with which CHX binds and arrests it. Assuming that the binding of CHX is a reversible process, we model CHX dissociation with a constant rate per bound CHX.

We begin by first simulating protein translation in a normal cell yeast till it reaches equilibrium (1500 sec). After this, whenever a ribosome at codon positions k of an mRNA elongates to $k+1$, the ribosome is arrested at $k+1$ by CHX with constant probability p_{chx-on}. CHX dissociates from a

bound ribosome at with a constant rate $r_{chx-off}$. We vary the probability of CHX-binding p_{chx-on} and its dissociation rate $r_{chx-off}$ to understand its effects on ribosome densities and dynamics of protein translation. We find that when CHX acts rapidly and has a low dissociation rate ($p_{chx-on} = 0.2$, $r_{chx-off} = 0.01$), we see peak excess relative ribosome-footprint density (e'_j Eqn S10) following the first ten codons is $\sim$ 400% , which is to comparable to ramps observed in ribosome profiling experiments with CHX pre-treatment (Figure S1) but significantly higher than the ramp observed without CHX in both simulations ($\sim$ 20%) and in the current study ($\sim$ 60%). This suggests that a large ramp of ribosome densities in the 5' region is likely an artifact of the action of CHX (Gerashchenko & Gladyshev, 2014).

1.3 Mapping ribosome profiling reads

The *S. cerevisiae* reference genome sequence and transcript models were downloaded from Ensembl at `ftp://ftp.ensembl.org/pub/release-74/fasta/saccharomyces_cerevisiae/dna/` and `ftp://ftp.ensembl.org/pub/release-74/gtf/saccharomyces_cerevisiae/`.

Data was processed using a framework written in Python. Reads were trimmed from the right of adapter sequences according to the specific library preparations used to generate each data set: reads from poly-adenlyated libraries were trimmed of all trailing As, and reads from libraries prepared with a pre-adynlyated linker (either 'CTGTAGGCACCATCAAT' or 'TCGTATGCCGTCTTCTGCTTG') were trimmed to the first position from the left at which the next 10 bases in the read were within a hamming distance of 1 from the first 10 bases of the linker sequence or to where a suffix of the read exactly matched a prefix of the linker sequence. For our data, 8 nt of randomized barcode sequence was trimmed from the left of each read and appended to the read's name. Reads originating from ribosomal RNAs were pre-filtered by mapping to an index of yeast rRNA sequences with bowtie2 version 2.2.1. Filtered reads were then mapped to the yeast genome and spliced transcript models using tophat2 v2.0.9. Reads mapping to any tRNA or other noncoding RNA genes were discarded. For each annotated coding sequence, counts of the number of uniquely mapped reads on the sense strand whose 5'-most mapped base occupied every position from 50 nt upstream of the start codon to 50 nt downstream of the stop codon were calculated. To

calculate codon occupancies, only trimmed reads of length 28, 29 and 30 (for which the identity of the codon occupying the A-site of the ribosome could be most reliably inferred) were used. Reads of length 28 and 29 were assigned to the codon at position +14, 15, or 16 from the start of the read, and reads of length 30 were assigned to the codon at position +15, 16, or 17. To calculate read densities, reads of all lengths were included.

Data sources: (GSE* indicates GEO accession number)

Flash-freeze: GSE75897

Ingolia: GSE13750 (Ingolia et al., 2009)

Zinshteyn: GSE45366 (Zinshteyn & Gilbert, 2013)

Gerashchenko: personal communication with Maxim Gerashchenko (Gerashchenko et al., 2012)

Artieri: GSE50049 (Artieri & Fraser, 2014b)

Mcmanus: GSE52119 (McManus et al., 2014)

Guydosh: GSE52968 (Guydosh & Green, 2014)

Lareau: GSE58321 (Lareau et al., 2014)

Gardin: SRP044053 (Gardin et al., 2014)

Pop: GSE63789 (Pop et al., 2014)

Jan: GSE61012 (Jan et al., 2014)

Williams: GSE61011 (Williams et al., 2014)

Nedialkova: GSE67387 (Nedialkova & Leidel, 2015)

1.4 Estimating tRNA abundances from RNA-seq

To estimate tRNA abundances, RNA-seq reads from the Ribo-Zero-treated sample were mapped to annotated *S. cerevisiae* tRNA loci (downloaded from Ensembl) using `Bowtie`, allowing up to one mismatch and sampling alignments for multiply mapping reads (options `-n 1 -l 25 -e 100 -M 1 --best --strata`). The total number of reads corresponding to each tRNA anticodon were tallied, and wobble parameters were used to estimate cognate tRNA abundances for individual codons.

1.5 Metagene analyses of ribosome and mRNA densities

To understand how ribosome and mRNA densities vary along the length of a transcript, we estimate position-specific ribosome densities of individual genes into a composite metagene. Let $x_{i,j}$ be the number of mapped RPF reads to position j of gene i based on its A-site. In order to avoid biases induced due genes with low coverage of reads, we restricted our analyses to genes with at least 128 mapped total mapped reads. To account for differences in initiation rates between different genes, we calculate the normalized ribosome density $z_{i,j}$ at codon position j by normalizing the mapped reads at that codon by the mean number of mapped reads in that gene.

$$z_{i,j} = \frac{x_{i,j}}{(\sum_{j=1}^{L_i} x_{i,j})/L_i} \tag{S7}$$

$$L_i = \text{Length of gene } i \text{ in codons}$$

We calculate the excess ribosome densities e_j at a particular position j by averaging the normalized ribosome density $z_{i,j}$ across all genes whose length is at least j ($L_i <= j$).

$$e_j = \frac{\sum_{i=1}^{N} z_{i,j}}{\sum_{i=1}^{N} \delta_i(j)} \tag{S8}$$

$$\delta_i(j) = \begin{cases} 1 & \text{if } L_i \geq j \\ 0 & \text{if } L_i < j \end{cases} \tag{S9}$$

$$N = \text{Number of genes}$$

The amount of excess ribosome densities e_j at the 5' ends of genes vary with each dataset. As a result, the excess ribosome densities asymptote at different values depending on the dataset, making it harder to compare the peaks of ribosome densities across datasets. To account for these differences, we estimated relative excess densities e'_j by normalizing excess ribosome-densities e_j with excess ribosome densities in the region spanning 450 and 500 codons. This region was chosen based on the observation that excess ribosome densities in all ribosome profiling dataset reach an

asymptote around 450-500 codons. The relative excess ribosome densities e'_j were calculated as follows:

$$e'_j = \frac{e_j}{(\sum_{j=450}^{500} e_j)/51} \tag{S10}$$

We report peak excess ribosome densities as the maximum of e'_j in a region spanning 10 to 500 codons. We ignore the first few codons as they are highly variable and result in sharp peaks due to continued initiation events.

We calculate excess mRNA coverage at each nucleotide position j similar to excess RPF reads described above. Let $y_{i,j}$ be the number of times mRNA reads overlapped position j of gene i. The normalized mRNA coverage at $g_{i,j}$ at nucleotide position j is obtained by normalizing the coverage at that nucleotide by the mean coverage of all nucleotides in that ORF.

$$g_{i,j} = \frac{y_{i,j}}{(\sum_{j=1}^{L'_i} y_{i,j})/L'_i} \tag{S11}$$

$$L'_i = \text{Length of gene } i \text{ in nucleotides}$$

We calculate the excess mRNA coverage h_j at a particular position j by averaging the normalized mRNA coverage $g_{i,j}$ across all genes whose length is at least j ($L'_i <= j$).

$$h_j = \frac{\sum_{i=1}^{N} g_{i,j}}{\sum_{i=1}^{N} \delta_i(j)} \tag{S12}$$

$$\delta_i(j) = \begin{cases} 1 & \text{if } L'_i \geq j \\ 0 & \text{if } L'_i < j \end{cases} \tag{S13}$$

The relative excess mRNA coverage h'_j based on 1350-1500 nucleotides were calculated as follows:

$$h'_j = \frac{h_j}{(\sum_{j=1350}^{1500} h_j)/151} \tag{S14}$$

To estimate the degree of bias in mRNA measurements in the 3' ends of genes, we scale excess

mRNA coverage at a nucleotide j by 1350-1500 nucleotides from the stop codon as follows:

$$h''_j = \frac{h_j}{(\sum_{j=L'_m-1350}^{L'_m-1500} h_j)/151} \tag{S15}$$

$$L'_m = \text{Max length of genes in nucleotides}$$

1.6 Comparing 5' and 3' codon-specific ribosome densities

To estimate codon-specific ribosome densities in the 5' and 3' ends of genes, we begin by first calculating normalized ribosome densities within a gene $z_{i,j}$ (Eqn. S7) for all genes with at least 250 codons. Normalizing ribosome-densities within a gene removes the effect of differences in initiation rates among genes when comparing normalized reads across many genes. The average ribosome density of all ribosome reads at codon type k in the 5' ($v_k^{5'}$) and 3' ($v_k^{3'}$) ends across the genome is then given by

$$v_k^{5'} = \frac{\sum_{i=1}^{N} \sum_{j=1 \forall c(i,j)=k}^{200} z_{i,j}}{\sum_{i=1}^{N} n_{i,k}^{5'}} \tag{S16}$$

$$v_k^{3'} = \frac{\sum_{i=1}^{N} \sum_{j=201 \forall c(i,j)=k}^{L_i} z_{i,j}}{\sum_{i=1}^{N} n_{i,k}^{3'}} \tag{S17}$$

$$n_{i,k}^{5'} = \text{Number of codons of type } k \text{ in gene } i \text{ in codons } 1 \leq j \leq 200$$

$$n_{i,k}^{3'} = \text{Number of codons of type } k \text{ in gene } i \text{ in codons } 201 \leq j \leq L_i$$

$$c(i,j) = \text{Identity of codon at position } j \text{ of gene } i$$

1.7 Estimating codon-specific elongation times

Most theoretical studies of protein translation assume a negative relationship between the elongation time of a codon and its tRNA abundance. To test this, we estimate codon-specific elongation times from ribosome densities as follows:

In order to avoid the confounding effects of 5' ribosomal ramp on our estimates of codon-specific ribosome densities, we restrict our analyses to genes with at least 250 codons and only consider

RPF reads mapped from codon position 200 onwards. Let $x_{i,j}$ be the number of mapped RPF reads to position j of gene i. The normalized ribosome density $z'_{i,j}$ at codon position $j >= 200$ is given by normalizing the mapped reads at that codon by the mean number of mapped reads from codon position 200 to L_i.

$$z'_{i,j} = \begin{cases} \frac{x_{i,j}}{(\sum_{j=200}^{L_i} x_{i,j})/(L_i - 199)} & \text{if } j \geq 200 \\ \text{NA} & \text{if } j < 200 \end{cases} \tag{S18}$$

Normalizing ribosome-densities within a gene removes the effect of differences in initiation rates among genes when comparing normalized reads across many genes. The average ribosome density of all ribosome reads at codon type k (v_k) across the genome is then given by

$$v_k = \frac{\sum_{i=1}^{N} \sum_{j=200 \forall c(i,j)=k}^{L_i} z'_{i,j}}{\sum_{i=1}^{N} n_{i,k}} \tag{S19}$$

$$n_{i,k} = \text{Number of codons of type } k \text{ in gene } i \text{ in codons } 200 \leq j \leq L_i$$

$$c(i,j) = \text{Identity of codon at position } j \text{ of gene } i$$

The expected elongation time of a codon is then directly proportional to average codon-specific ribosome density, v_k because codons with longer elongation times have higher average ribosome densities.

1.8 Estimating protein synthesis rates

We estimate protein synthesis rates of individual genes using the densities of ribosomes on their mRNAs. However, the ribosome densities per mRNA as estimated by taking ratios of RPF RPKM and mRNA RPKM are likely biased. The main source of this bias is the presence of a 5' ramp of RPF reads that varies with position along a transcript.

1.8.1 Ramp correction factor

In order to obtain unbiased estimates of gene-specific ribosome-densities, we apply a position-dependent correction factor to RPF reads that accounts for the ramp. However, the observed ramp of ribosome densities is partly a result of codon ordering within genes in addition to experimental artifacts. Our position-dependent correction factor accounts for the ribosomal ramp due to experimental artifacts by explicitly taking into account the contribution of codon usage dependent ramp.

The total excess ribosome density at a position j across all gene is given by e_j (see above, Eqn. S8). To calculate the expected excess ribosome density at a position j due to codon ordering within a gene, we first calculate average ribosome density for codon k, v_k as described above (Eqn. S19). The expected excess ribosome density d_j at position j due to patterns of codon usage is given as follows:

The relative codon-usage expected ribosome density $q_{i,j}$ at position j of gene i is given by normalizing the expected ribosome density $v_{c(i,j)}$ at that codon by the mean expected ribosome density for that gene.

$$q_{i,j} = \frac{v_{c(i,j)}}{(\sum_{j=1}^{L_i} v_{c(i,j)})/L_i} \tag{S20}$$

$$L_i = \text{Length of gene } i \text{ in codons}$$

$$c(i,j) = \text{Identity of codon at position } j \text{ of gene } i$$

The average codon-usage expected ribosome density at position j (d_j) across the genome is then given by

$$d_j = \frac{\sum_{i=1}^{N} q_{i,j}}{\sum_{i=1}^{N} \delta_i(j)} \tag{S21}$$

$$\delta_i(j) = \begin{cases} 1 & \text{if } L_i \geq j \\ 0 & \text{if } L_i < j \end{cases} \tag{S22}$$

The ramp correction factor f_j at position j along any gene is then defined as the ratio of observed ramp e_j over the expected ramp d_j.

$$f_j = \frac{e_j}{d_j} \tag{S23}$$

1.8.2 Unbiased estimate of ribosome density per mRNA

Let $x_{i,j}$ be the number of mapped RPF reads and $y_{i,j}$ be the number of mapped mRNA reads to position j of gene i. The unbiased estimate of ribosome density r_i per mRNA for gene i is defined as

$$\text{Mean corrected RPF reads} = \frac{\sum_{j=1}^{L_i} \frac{x_{i,j}}{f_j}}{L_i} \tag{S24}$$

$$\text{Mean mRNA reads} = \frac{\sum_{j=1}^{L_i} y_{i,j}}{L_i} \tag{S25}$$

$$\begin{aligned} r_i &= \frac{\text{Mean corrected RPF reads}}{\text{Mean mRNA reads}} \\ &= \frac{\sum_{j=1}^{L_i} \frac{x_{i,j}}{f_j}}{\sum_{j=1}^{L_i} y_{i,j}} \end{aligned} \tag{S26}$$

where f_j is the ramp correction factor (see above, Eqn. S23).

1.8.3 Initiation efficiency

We estimate initiation efficiency of a genes using the analytic approximations for the initiation probability p_i based on steady-state behavior of a whole cell simulation described above (Eqn. S5).

$$p_i \approx \frac{R_i^b x}{A_i L_i \left(\sum_{j=1}^{61} \frac{u_{j,i}}{w_j T_{\phi(j)}^t} \right)}$$

where

$$x = \frac{\tau_r N_r s}{0.15 R^t \tau_t N_t}$$

see Methods (1. Simulation model section) above for details on parameter notations and values. $\frac{R_i^b}{A_i L_i}$ describes the ribosome density per mRNA on gene i. Here we substitute $\frac{R_i^b}{A_i L_i}$ with unbiased estimates of ribosome density per mRNA r_i (Eqn. S26) in Eqn. S5. Moreover, estimates of codon-specific ribosome densities v_k (Eqn. S19) reflect average elongation times of codons – codons with longer elongation times have higher ribosome densities. Therefore, we substitute expected elongation time of a codon given by $\frac{1}{w_j T_{\phi(j)}^t}$ in Eqn. S5 with estimates of codon-specific ribosome densities v_k. As a result, our initiation efficiency p_i^E for gene i is estimated solely from ribosome profiling data and is defined as

$$p_i^E \approx \frac{r_i x G}{\sum_{j=1}^{61} u_{j,i} v_k} \tag{S27}$$

where G is the global scaling parameter, which scales r_i such that the total number of ribosomes within a cell are $200,000$, total number of mRNAs are $60,000$ and total number of tRNAs are $3,300,000$ based on empirical estimates (Shah et al., 2013).

1.8.4 Protein synthesis rates

Protein synthesis rate S_i of a gene i within a cell depend on total number of mRNAs for that gene (A_i) and initiation rate per mRNA (ρ_i Eqn. S2) and the number of free ribosomes. Here we modify Eqn. S2) by substituting initiation probabilities (p_i) with estimates of initiation efficiencies (p_i^E)

based on profiling data (Eqn. S27).

$$S_i = p_i^E A_i \frac{R^f}{\tau_r N_r} \tag{S28}$$

In estimating S_i, we assume that 15% of the total 200,000 ribosomes are free and the rest are bound to mRNAs, such that $R^f = 3 \times 10^4$. Furthermore, we scale mRNA abundances as measured by mRNA RPKM such that the total number of mRNAs are 60,000 ($\sum_i A_i = 6 \times 10^4$).

2 Supplemental Text

2.1 Co-translational folding and inter-domain linkers

During protein translation, a growing polypeptide chain begins to fold as soon as it emerges from the ribosome - a process known as co-translational folding. Several studies have suggested that pausing of ribosome at specific instances is necessary for the nascent polypeptide to take native-like folds (Kimchi-Sarfaty et al., 2007; Pechmann & Frydman, 2013). If ribosomal pausing significantly affects co-translation folding, then we expect a higher density of ribosomes in regions between protein domains. To test this, we downloaded domain assignments for individual genes in the *S. cerevisiae* genome from SGD – `http://downloads.yeastgenome.org/curation/calculated_protein_info/domains/domains.tab` based on InterProScan `http://www.ebi.ac.uk/interpro` on Jan 14, 2014. Domain assignments were based on InterProScan classifications (Jones et al., 2014) obtained from the Superfamily database (Wilson et al., 2009) and Pfam database.

2.2 Widespread pauses after polybasic but not at proline residues

Aside from the abundance of cognate tRNA, the ribosome-footprint density at a particular codon in an ORF might also be influenced by interactions between the emerging nascent polypeptide and the ribosome. Indeed, extensive polybasic tracts can stall translation, presumably because of electrostatic interactions between the nascent polypeptide and the negatively charged exit tunnel (Lu & Deutsch, 2008; Tuller et al., 2011; Brandman et al., 2012; Charneski & Hurst, 2013). However, the relationship between polybasic tracts and ribosome densities has been the subject of some debate (Charneski & Hurst, 2013; Artieri & Fraser, 2014a). To help resolve this issue, we estimated excess ribosome-footprint densities ($z_{i,j}$, Eqn. S7) in windows of 10 amino acids with varying number of positively charged residues. We observed a sharp peak of ribosome density at the end of highly positively charged regions (defined as six arginine or lysine residues within a 10 amino-acid window), which would position the basic residues within the exit tunnel (Figure S3). The pause amplitude steadily decreased as the number of basic residues within the window decreased, but pausing was still apparent with as few as three basic residues within the 10-amino-acid window.

Our results on pausing resembled the findings of other recent studies (Lu & Deutsch, 2008; Tuller et al., 2011; Brandman et al., 2012; Charneski & Hurst, 2013). However, this pausing signal was absent in every ribosome-profiling dataset with CHX pre-treatment (Figure S3), whereas it was present in six of eight datasets without CHX pre-treatment, which provides further evidence that drug treatment obscures the locations of natural elongation pauses. When simulating ribosome densities surrounding polybasic stretches in our whole-cell model, we found a relative depletion of ribosomes in polybasic stretches (Figure S3N), which indicates that elongation stalls within polybasic stretches were not caused by biased codon usage and that the observed excess ribosome-footprint density likely underestimated the direct effect of polybasic stretches on elongation rates.

Previous studies have also reported an excess of elongating ribosomes at sites where a proline amino acid is positioned in the P site of the ribosome (Artieri & Fraser, 2014a; Martens et al., 2015). This increased density is attributed to the difficulty of forming a peptide bond between the incoming amino acid in the A site and proline in the P site. When we looked for excess ribosome density at P-site proline codons, we also observed a strong accumulation of ribosomes in datasets that used CHX pre-treatment (Figure S2). However, this stall at P-site prolines was not observed in either our flash-freeze dataset or most other published flash-freeze datasets (Figure S2). Thus, most (if not all) of the excess ribosome density previously observed at P-site proline codons can be attributed to the effects of CHX pre-treatment.

2.3 Contribution of translational control to proportional synthesis

Although narrower than reported in earlier ribosome-profiling studies, the IE distribution that we observed was still large enough to enable the cell to tune synthesis rates via translational control. One scenario in which this might be important is in the proportional synthesis of the subunits for multi-protein complexes. A recent genome-wide study of protein-synthesis rates in E. coli and S. cerevisiae concluded that components of multi-subunit complexes are usually synthesized in precise proportion to their stoichiometry (Li et al., 2014). In E. coli, the subunits of multi-protein complexes are usually encoded on the same polycistronic mRNA and thus can be synthesized in different proportions only if they have different translation-initiation rates. In eukaryotes, however,

the subunits of protein complexes are encoded on separate mRNAs, which enables proportional synthesis to be achieved through control of mRNA abundance (via transcription rate and mRNA half-life). Nonetheless, translational control might still compensate for differences in mRNA abundance and thereby achieve more precise stoichiometry of synthesis rates. To explore this possibility, we examined the synthesis rates, mRNA abundances, and IEs of the subunits of stably associated complexes previously shown to undergo proportional synthesis (Li et al., 2014). mRNAs encoding subunits of heterodimeric complexes had roughly similar abundances (within 2 fold), indicating that most of their proportional synthesis is achieved through coordinated mRNA levels (Figure S6D). The same was true for mRNAs encoding multi-protein complexes, after accounting for subunit stoichiometry (within ≈2 fold, Figure S6E), as well for mRNAs encoding heterodimeric complexes containing alternative paralogous subunits (within 1.4 fold, Figure S6F). These observations were consistent with the narrow range of IEs in yeast; with limited translational control, proportional synthesis requires roughly proportional mRNA levels. However, in 12 out of 18 cases subunit stoichiometry was more accurately reflected by synthesis rates than by mRNA abundances (Figures S6E-G), as quantified by the coefficients of variation (Figure S6G). For example, the subunits of the heterodimeric FACT complex were translated at equal levels even though Spt16-encoding mRNA was 59% more abundant than the Pob3-encoding mRNA. Similarly, in the mitochondrial *alpha*-ketoglutarate dehydrogenase complex, higher expression of the Kdg2 subunit relative to the Kdg1 subunit (which have 2:1 stoichiometry in the complex) was achieved entirely at the level of translation. We also found that mRNAs encoding proportionally synthesized subunits of heterodimeric complexes tended to have similar IEs ($R^2 = 0.72$, Figure S6D), suggesting that such mRNAs might be coregulated at the level of translation. Collectively, these results suggest a tendency for translational control to compensate for small differences in mRNA levels to help achieve more proportional synthesis.

2.4 Regression model

Initiation efficiency (p^E) of a gene depends on several features of a coding sequence. In order to identify a set of features that explain the most variation in p^E, we use the multiple regression

framework in R (R Core Team, 2012). We regressed the p^E of a gene against its length, mRNA abundance (RPKM), 5' UTR length and its GC content, 5' cap folding energy, and number of upstream ATGs in 5' UTR (uATG). We restricted the analyses to 2549 genes with experimentally verified 5' UTR lengths and number of upstream ATGs (Arribere & Gilbert, 2013). To estimate 5' cap folding energy we used sequences of length 70 nts from the 5' end of the mRNA transcript as sequences of these lengths showed the highest correlation with TE ratios (Figure S7). We calculated the folding energies using RNAfold algorithm from Vienna RNA package (Hofacker et al., 1994) at 37 °C. The values of p^E, mRNA RPKM and protein length were log-transformed in the regression model.

To identify which features explain the highest amount of variation in initiation efficiencies, we used Akaike's Information Criteria (AIC) for model selection. We performed both step-up and step-down model selection using the `stepAIC` function in `MASS` package in R. We find that the multiple regression model that best explains the variation in p^E even after penalizing for model complexity includes all the 6 variables considered (Table S6). This model explains $\sim 58\%$ of variation in p^E across all the genes considered. We find that initiation efficiencies scale positively with predicted folding energies and mRNA abundances. This indicates that genes with weaker 5' cap structure and RNA structure around the start site have higher rates of initiation. In contrast, genes with a higher number of uATGs and longer 5' UTRs and coding sequence length have lower p^E. Moreover, genes that have higher mRNA abundances also tend to have higher p^E and hence higher ribosome densities (TE) on them. This suggests that genes under selection for higher protein abundances are selected both at the level of transcription – leading to higher mRNA abundances and at the translation level – leading to higher ribosome densities.

Supplemental References

Arribere, J. A. & Gilbert, W. V. (2013). Roles for transcript leaders in translation and mRNA decay revealed by transcript leader sequencing. *Genome Res*, 23(6), 977–987.

Artieri, C. G. & Fraser, H. B. (2014a). Accounting for biases in riboprofiling data indicates a major role for proline in stalling translation. *Genome Res*, 24(12), 2011–2021.

Artieri, C. G. & Fraser, H. B. (2014b). Evolution at two levels of gene expression in yeast. *Genome Res*, 24(3), 411–421.

Brandman, O., Stewart-Ornstein, J., Wong, D., Larson, A., Williams, C. C., Li, G.-W., Zhou, S., King, D., Shen, P. S., Weibezahn, J., Dunn, J. G., Rouskin, S., Inada, T., Frost, A., & Weissman, J. S. (2012). A ribosome-bound quality control complex triggers degradation of nascent peptides and signals translation stress. *Cell*, 151(5), 1042–1054.

Brar, G. A., Yassour, M., Friedman, N., Regev, A., Ingolia, N. T., & Weissman, J. S. (2012). High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science*, 335(6068), 552–557.

Charneski, C. A. & Hurst, L. D. (2013). Positively charged residues are the major determinants of ribosomal velocity. *PLoS Biol*, 11(3), e1001508.

Chen, C., Stevens, B., Kaur, J., Smilansky, Z., Cooperman, B. S., & Goldman, Y. E. (2011). Allosteric vs. spontaneous exit-site (E-site) tRNA dissociation early in protein synthesis. *Proc Natl Acad Sci USA*, 108(41), 16980–16985.

Curran, J. F. & Yarus, M. (1989). Rates of aminoacyl-tRNA selection at 29 sense codons in vivo. *J Mol Biol*, 209(1), 65–77.

Gardin, J., Yeasmin, R., Yurovsky, A., Cai, Y., Skiena, S., & Futcher, B. (2014). Measurement of average decoding rates of the 61 sense codons in vivo. *eLife*, 3.

Gerashchenko, M. V. & Gladyshev, V. N. (2014). Translation inhibitors cause abnormalities in ribosome profiling experiments. *Nucl Acids Res*, 42(17), e134–e134.

Gerashchenko, M. V., Lobanov, A. V., & Gladyshev, V. N. (2012). Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress. *Proc Natl Acad Sci USA*, 109(43), 17394–17399.

Gillespie, D. (1977). Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry*, 81(25), 2340–2361.

Guydosh, N. R. & Green, R. (2014). Dom34 rescues ribosomes in 3' untranslated regions. *Cell*, 156(5), 950–962.

Hofacker, I. L., Fontana, W., & Stadler, P. (1994). Fast folding and comparison of RNA secondary structures. *Monatshefte für Chemie*, 125, 167–188.

Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S., & Weissman, J. S. (2009). Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science*, 324(5924), 218–223.

Jan, C. H., Williams, C. C., & Weissman, J. S. (2014). Principles of ER cotranslational translocation revealed by proximity-specific ribosome profiling. *Science*, 346(6210), 1257521.

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S. Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30(9), 1236–1240.

Kimchi-Sarfaty, C., Oh, J. M., Kim, I.-W., Sauna, Z. E., Calcagno, A. M., Ambudkar, S. V., & Gottesman, M. M. (2007). A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science*, 315(5811), 525–528.

Lareau, L. F., Hite, D. H., Hogan, G. J., & Brown, P. O. (2014). Distinct stages of the translation elongation cycle revealed by sequencing ribosome-protected mRNA fragments. *eLife*, 3, e01257.

Li, G.-W., Burkhardt, D., Gross, C., & Weissman, J. S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell*, 157(3), 624–635.

Lim, V. I. & Curran, J. F. (2001). Analysis of codon:anticodon interactions within the ribosome provides new insights into codon reading and the genetic code structure. *RNA*, 7(7), 942–957.

Lu, J. & Deutsch, C. (2008). Electrostatics in the ribosomal tunnel modulate chain elongation rates. *J Mol Biol*, 384(1), 73–86.

Martens, A. T., Taylor, J., & Hilser, V. J. (2015). Ribosome A and P sites revealed by length analysis of ribosome profiling data. *Nucl Acids Res*, 43(7), 3680–3687.

McManus, C. J., May, G. E., Spealman, P., & Shteyman, A. (2014). Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. *Genome Res*, 24(3), 422–430.

Nedialkova, D. D. & Leidel, S. A. (2015). Optimization of Codon Translation Rates via tRNA Modifications Maintains Proteome Integrity. *Cell*, 161(7), 1606–1618.

Pechmann, S. & Frydman, J. (2013). Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nat Struct Mol Biol*, 20(2), 237–243.

Pop, C., Rouskin, S., Ingolia, N. T., Han, L., Phizicky, E. M., Weissman, J. S., & Koller, D. (2014). Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation. *Mol Syst Biol*, 10(12), 770–770.

R Core Team (2012). R: A language and environment for statistical computing. *R foundation for Statistical Computing*.

Schneider-Poetsch, T., Ju, J., Eyler, D. E., Dang, Y., Bhat, S., Merrick, W. C., Green, R., Shen, B., & Liu, J. O. (2010). Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin. *Nat Chem Biol*, 6(3), 209–217.

Shah, P., Ding, Y., Niemczyk, M., Kudla, G., & Plotkin, J. B. (2013). Rate-limiting steps in yeast protein translation. *Cell*, 153(7), 1589–1601.

Tuller, T., Veksler-Lublinsky, I., Gazit, N., Kupiec, M., Ruppin, E., & Ziv-Ukelson, M. (2011). Composite effects of gene determinants on the translation speed and density of ribosomes. *Genome Biol*, 12(11), R110.

Williams, C. C., Jan, C. H., & Weissman, J. S. (2014). Targeting and plasticity of mitochondrial proteins revealed by proximity-specific ribosome profiling. *Science*, 346(6210), 748–751.

Wilson, D., Pethica, R., Zhou, Y., Talbot, C., Vogel, C., Madera, M., Chothia, C., & Gough, J. (2009). SUPERFAMILY–sophisticated comparative genomics, data mining, visualization and phylogeny. *Nucl Acids Res*, 37(Database issue), D380–6.

Zinshteyn, B. & Gilbert, W. V. (2013). Loss of a conserved tRNA anticodon modification perturbs cellular signaling. *PLoS Genet*, 9(8), e1003675.

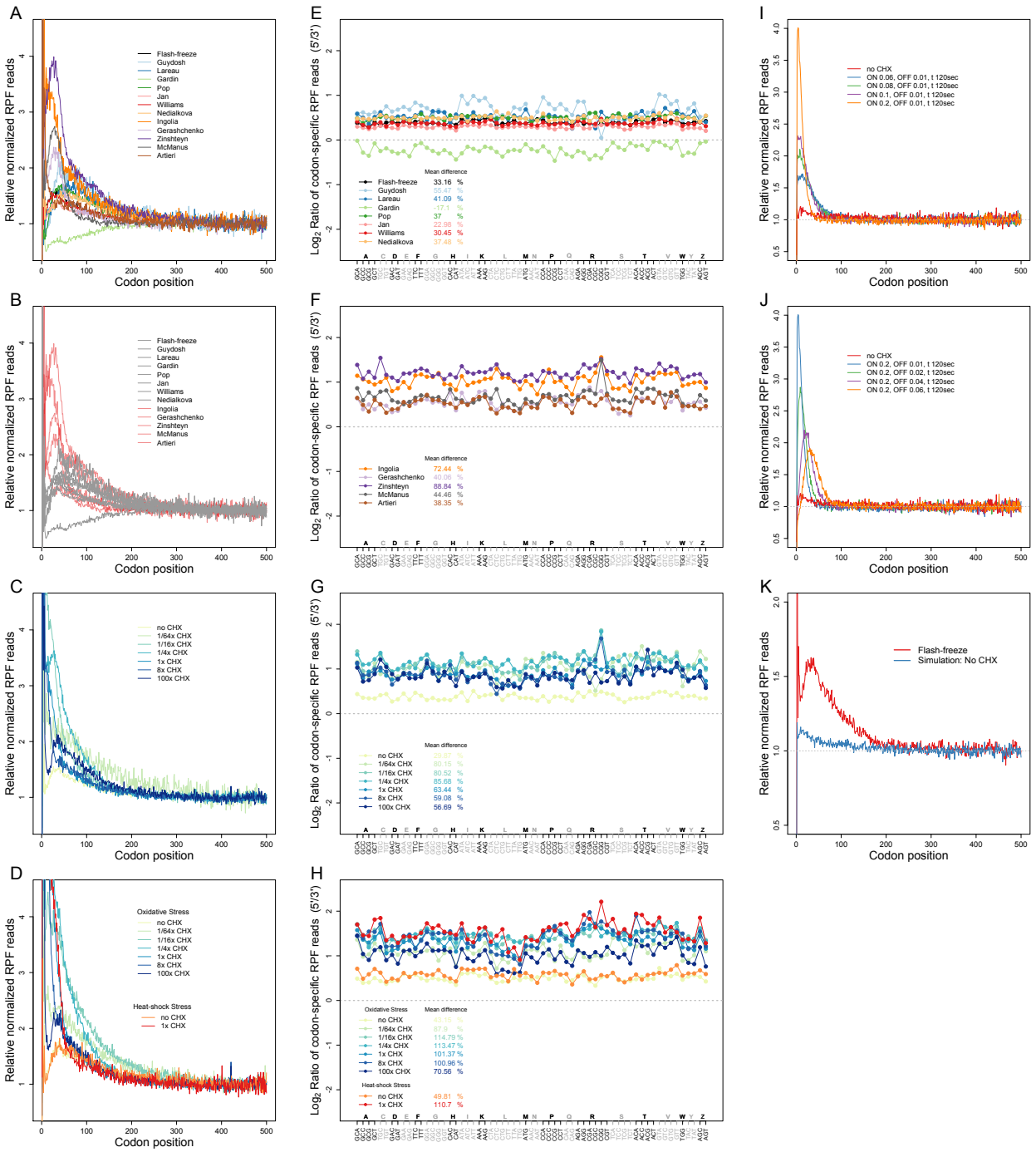
A
Relative normalized RPF reads
Codon position
Flash-freeze
Guydosh
Lareau
Gardin
Pop
Jan
Williams
Nedialkova
Ingolia
Gerashchenko
Zinshteyn
McManus
Artieri
E
Log2 Ratio of codon-specific RPF reads (5'/3')
Mean difference
Flash-freeze 33.16 %
Guydosh 25.49 %
Lareau 41.09 %
Gardin -17.1 %
Pop 37 %
Jan 22.98 %
Williams 30.45 %
Nedialkova 37.48 %
I
Relative normalized RPF reads
Codon position
no CHX
ON 0.06, OFF 0.01, t 120sec
ON 0.08, OFF 0.01, t 120sec
ON 0.1, OFF 0.01, t 120sec
ON 0.2, OFF 0.01, t 120sec
B
Relative normalized RPF reads
Codon position
Flash-freeze
Guydosh
Lareau
Gardin
Pop
Jan
Williams
Nedialkova
Ingolia
Gerashchenko
Zinshteyn
McManus
Artieri
F
Log2 Ratio of codon-specific RPF reads (5'/3')
Mean difference
Ingolia 72.44 %
Gerashchenko 40.06 %
Zinshteyn 88.84 %
McManus 44.46 %
Artieri 38.35 %
J
Relative normalized RPF reads
Codon position
no CHX
ON 0.2, OFF 0.01, t 120sec
ON 0.2, OFF 0.02, t 120sec
ON 0.2, OFF 0.04, t 120sec
ON 0.2, OFF 0.06, t 120sec
C
Relative normalized RPF reads
Codon position
no CHX
1/64x CHX
1/16x CHX
1/4x CHX
1x CHX
8x CHX
100x CHX
G
Log2 Ratio of codon-specific RPF reads (5'/3')
Mean difference
no CHX 29.87 %
1/64x CHX 80.15 %
1/16x CHX 80.52 %
1/4x CHX 85.68 %
1x CHX 63.44 %
8x CHX 59.08 %
100x CHX 56.69 %
K
Relative normalized RPF reads
Codon position
Flash-freeze
Simulation: No CHX
D
Relative normalized RPF reads
Codon position
Oxidative Stress
no CHX
1/64x CHX
1/16x CHX
1/4x CHX
1x CHX
8x CHX
100x CHX
Heat-shock Stress
no CHX
1x CHX
H
Log2 Ratio of codon-specific RPF reads (5'/3')
Oxidative Stress
Mean difference
no CHX 43.15 %
1/64x CHX 97.9 %
1/16x CHX 114.79 %
1/4x CHX 113.47 %
1x CHX 101.37 %
8x CHX 100.96 %
100x CHX 70.56 %
Heat-shock Stress
no CHX 49.81 %
1x CHX 110.7 %

Figure S1 *(preceding page)*: Analyses of the 5' ramp of ribosomes, Related to Figure 1. (**A-D**) Metagene analyses of RPF density, performed as in Figure 1D, comparing results of the current study (flash-freeze) to those of published studies. Published datasets in (**A**) are labeled by the first author's name: Guydosh (Guydosh & Green, 2014), Lareau (Lareau et al., 2014), Gardin (Gardin et al., 2014), Pop (Pop et al., 2014), Jan (Jan et al., 2014), Williams (Williams et al., 2014), Nedialkova (Nedialkova & Leidel, 2015), Ingolia (Ingolia et al., 2009), Gerashchenko (Gerashchenko et al., 2012), Zinshteyn (Zinshteyn & Gilbert, 2013), McManus (McManus et al., 2014), Artieri (Artieri & Fraser, 2014). (**B**) Same panel as in (**A**) with datasets using CHX-pretreatment colored in red and datasets without CHX-pretreatment colored in gray. (**C-D**) Datasets from Gerashchenko (Gerashchenko & Gladyshev, 2014) with varying levels of CHX-pretreatment under both unstressed (**C**) and stressed (**D**) conditions. (**E-H**) Comparison of codon-specific RPFs within and beyond the 5' ramp in datasets in (**A-D**). (**I-K**) Metagene analyses of RPF density based on the whole-cell simulation model (Shah *et al.* 2013), performed as in Figure 1D. Simulations with CHX pre-treatment were performed as described in Section 2 of the Supporting File. We simulated protein synthesis under no-CHX, and either (**I**) four different CHX arrest probabilities (p_{chx-on}) and a fixed CHX dissociation rate ($r_{chx-off}$), or (**J**) a fixed CHX arrest probability (p_{chx-on}) and four different CHX dissociation rates ($r_{chx-off}$). Increasing p_{chx-on} led to higher ramps, whereas increasing $r_{chx-off}$ led to lower ramps as well as a shift of the peak ribosome density towards the 3' end. (**K**) Metagene analysis of RPF density observed in the flash-freeze experiment compared with the results of the whole-cell simulation model. Otherwise, as in (**I**).

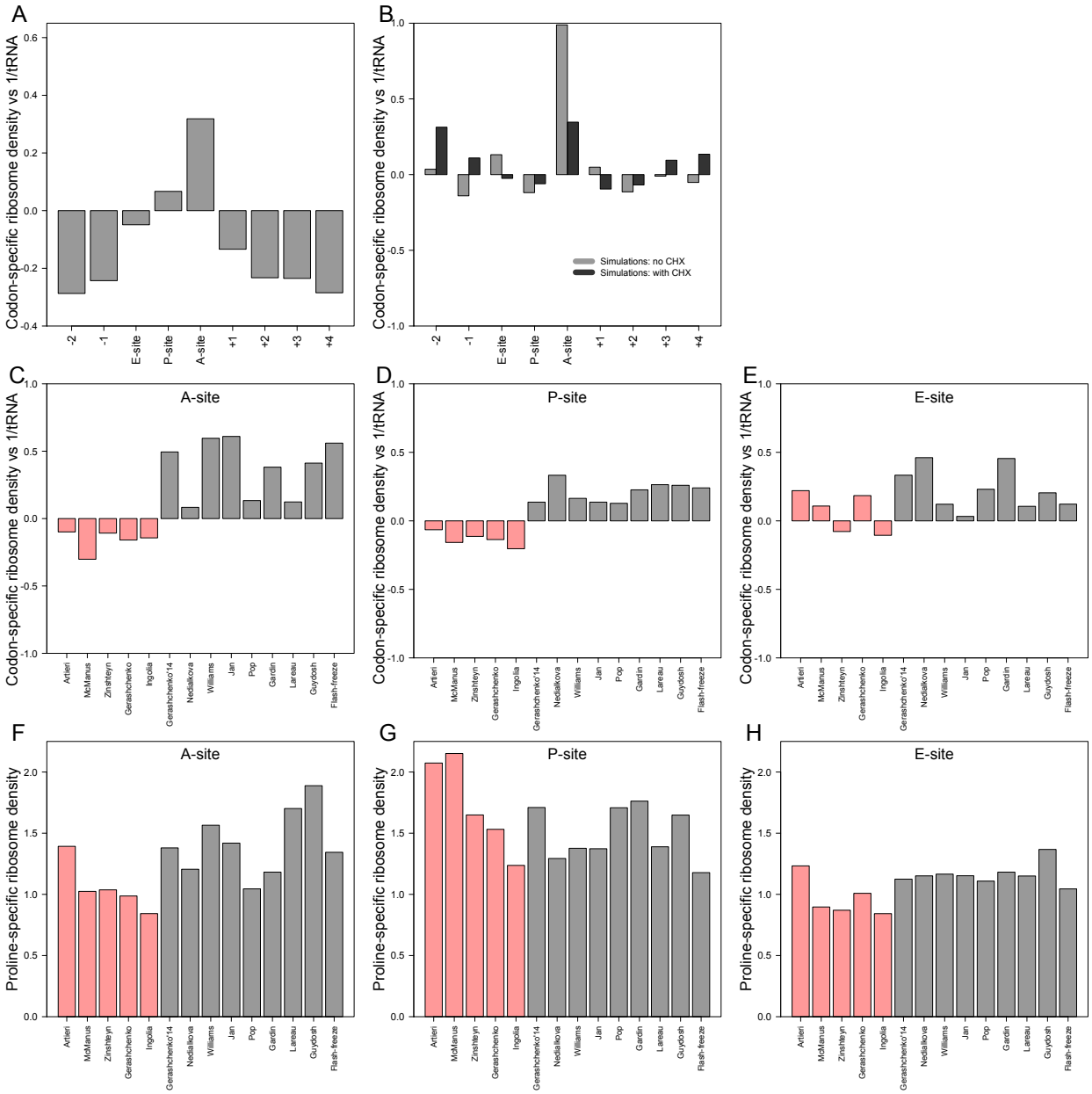

Figure S2: Relationship between cognate tRNA abundances and codon-specific ribosome densities, Related to Figure 2. (**A**) Correlation between codon-specific excess ribosome densities from the flash-freeze dataset and cognate tRNA abundances estimated using RNA-seq (Table S1). Otherwise, as in Figure 2B. (**B**) Correlation between codon-specific excess ribosome densities and cognate tRNA abundances in simulations with (black) and without (gray) CHX pretreatment. Otherwise, as in Figure 2B. (**C-E**) Correlation between codon-specific excess ribosome densities at the A-, P- and E-sites and cognate tRNA abundances, respectively. (**F-H**). Proline-specific ribosome densities in 13 ribosome-profiling experiments. RPF reads were assigned to the A-, P-, and E-site positions based on the distance from the 5' ends of fragments, to calculate proline-specific excess ribosome densities. Published datasets are labeled by the first author's name: Guydosh (Guydosh & Green, 2014), Lareau (Lareau et al., 2014), Gardin (Gardin et al., 2014), Pop (Pop et al., 2014), Jan (Jan et al., 2014), Williams (Williams et al., 2014), Nedialkova (Nedialkova & Leidel, 2015), Gerashchenko 2014 (Gerashchenko & Gladyshev, 2014), Ingolia (Ingolia et al., 2009), Gerashchenko (Gerashchenko et al., 2012), Zinshteyn (Zinshteyn & Gilbert, 2013), McManus (McManus et al., 2014), Artieri (Artieri & Fraser, 2014). Datasets using CHX-pretreatment colored in red and datasets without CHX-pretreatment colored in gray. Otherwise, as in Figure 2B.

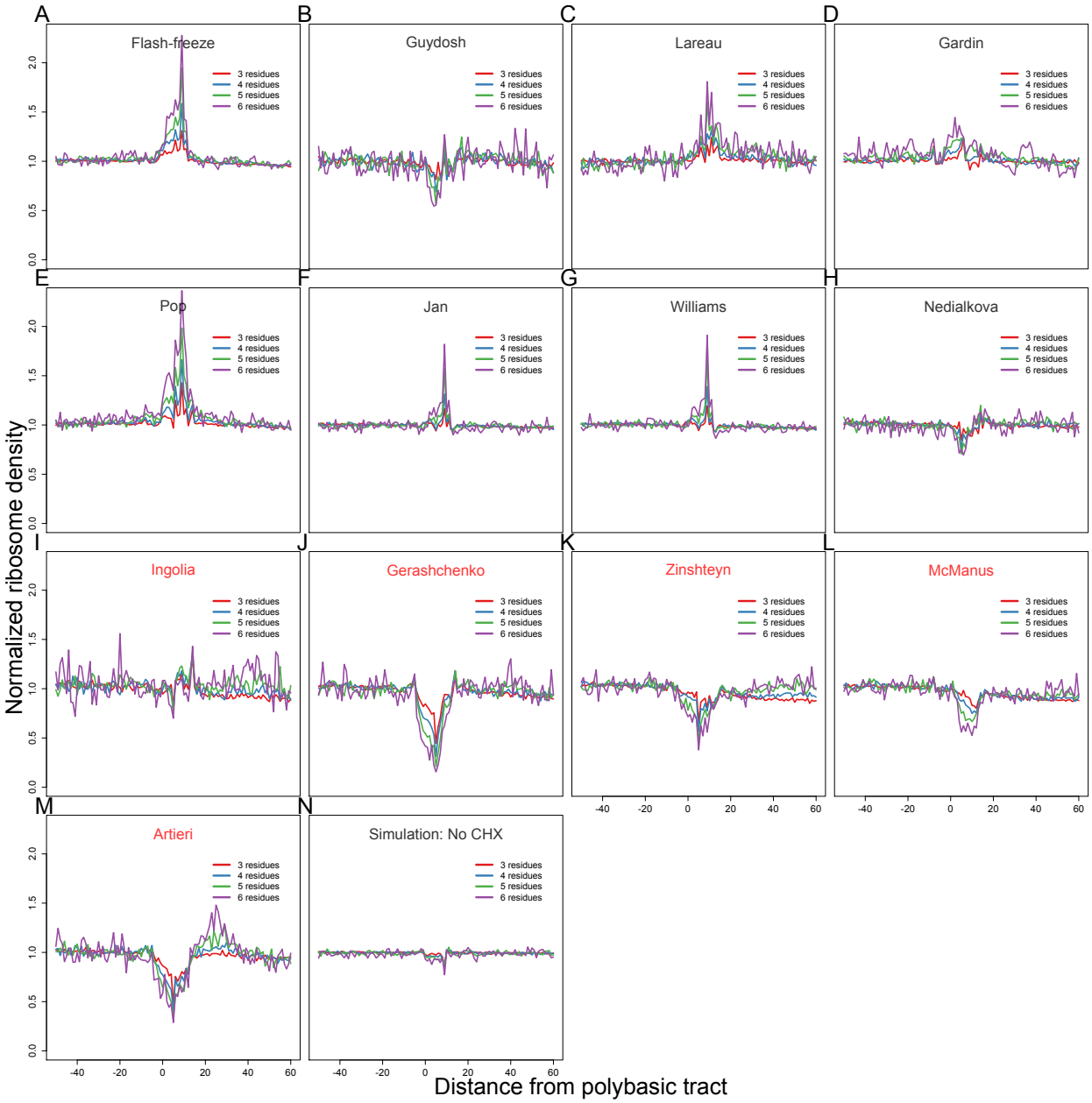
A
Flash-freeze
B
Guydosh
C
Lareau
D
Gardin
E
Pop
F
Jan
G
Williams
H
Nedialkova
I
Ingolia
J
Gerashchenko
K
Zinshteyn
L
McManus
M
Artieri
N
Simulation: No CHX
3 residues
4 residues
5 residues
6 residues
Normalized ribosome density
Distance from polybasic tract

Figure S3 *(preceding page)*: Analyses of ribosome stalling near polybasic tracts, Related to Figure 2. (**A-N**) Metagene analysis of normalized ribosome density surrounding polybasic tracts in our flash-freeze, twelve other published ribosome-profiling experiments and whole-cell simulations. Regions within ORFs that contained the indicated number of basic residues (arginine and lysine) within a stretch of 10 amino acids were aligned by the start of the region. Plotted are the normalized ribosome densities ($z_{i,j}$, Eqn. S7) observed at each codon position. Published datasets are labeled by the first author's name: Guydosh (Guydosh & Green, 2014), Lareau (Lareau et al., 2014), Gardin (Gardin et al., 2014), Pop (Pop et al., 2014), Jan (Jan et al., 2014), Williams (Williams et al., 2014), Nedialkova (Nedialkova & Leidel, 2015), Ingolia (Ingolia et al., 2009), Gerashchenko (Gerashchenko et al., 2012), Zinshteyn (Zinshteyn & Gilbert, 2013), McManus (McManus et al., 2014), Artieri (Artieri & Fraser, 2014). Datasets using CHX-pretreatment names in red and datasets without CHX-pretreatment named in black.

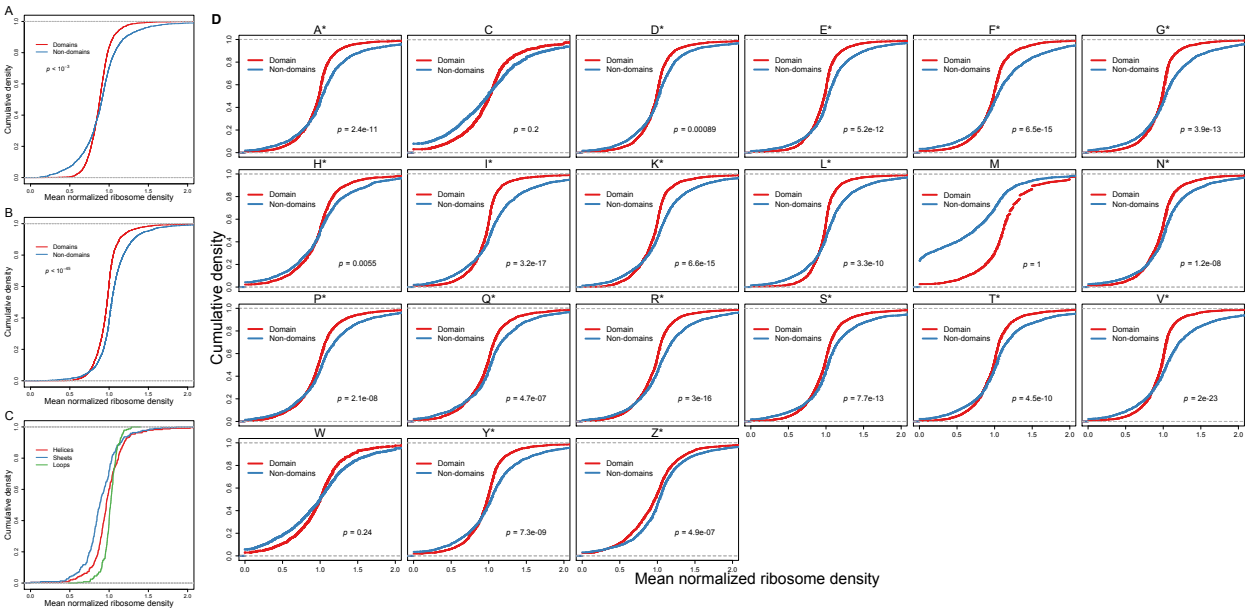

Figure S4: Relationship between elongation dynamics and either domain architecture or protein secondary structure. Related to Figure 3. (**A**) Cumulative distributions of normalized ribosome densities within and outside of protein-folding domains, considering ORFs with at least 250 codons but ignoring the first 200 codons in each ORF. Otherwise, as in Figure 3A. (**B**) Cumulative distributions of normalized ribosome densities within and outside of protein-folding domains. Domain assignments were based on InterProScan classifications (Jones *et al.* 2014) using the Pfam database (Bateman *et al.* 2002). Otherwise, as in Figure 3A. (**C**) Cumulative distributions of normalized ribosome densities within the indicated classes of protein secondary structures (helices, sheets and loops). Secondary structure assignments for proteins were obtained from (Pechmann and Frydman 2013). Otherwise, as in Figure 3A. **D**) Relationship between elongation dynamics and domain architecture for individual amino acids. Cumulative distributions of normalized ribosome densities within and outside of protein-folding domains, analyzed for each amino acid. Mean normalized ribosome densities ($z_{i,j}$, Eqn. S7) for codons within the domain-encoding and non-domain-encoding regions were individually calculated for each ORF. Panels are labeled with the single-letter code for each amino acid. Serine codons are partitioned into two sets with 4 and 2 codons (S and Z, respectively). Statistically significant ($p < 0.05$) slowing outside of protein-folding domains is indicated (asterisk). Otherwise, as in Figure 3A.

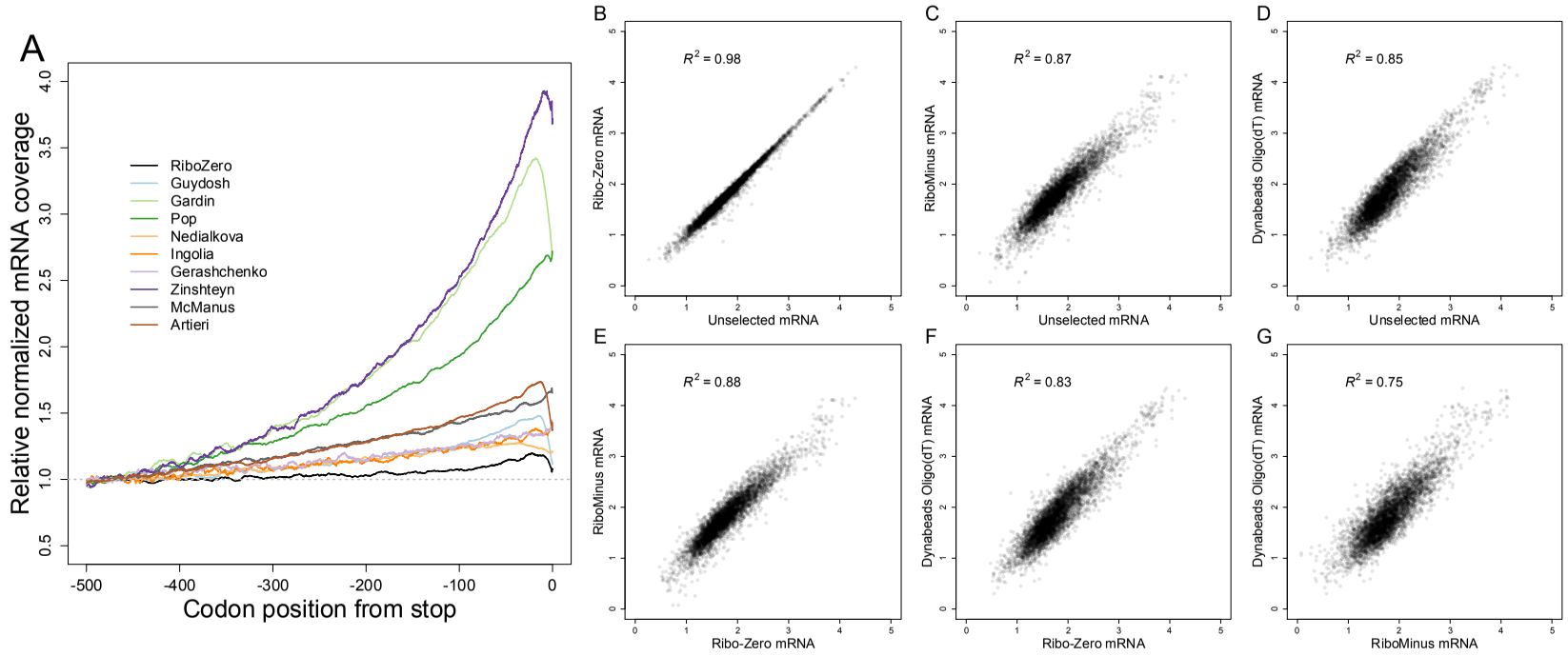

Figure S5: 3' bias observed in RNA-seq datasets and effect of mRNA enrichment methods on mRNA abundance measurements, Related to Figure 4. (**A**) Metagene analysis of RNA-seq read coverage in ten ribosome-profiling experiments. Otherwise, as in Figure 4B. (**B-G**) Pairwise comparisons of mRNA abundances (Log_{10} RPKM) following mRNA enrichment by the indicated methods. Otherwise, as in Figure 4A. For comparison, Figure 4A is repeated as panel A.

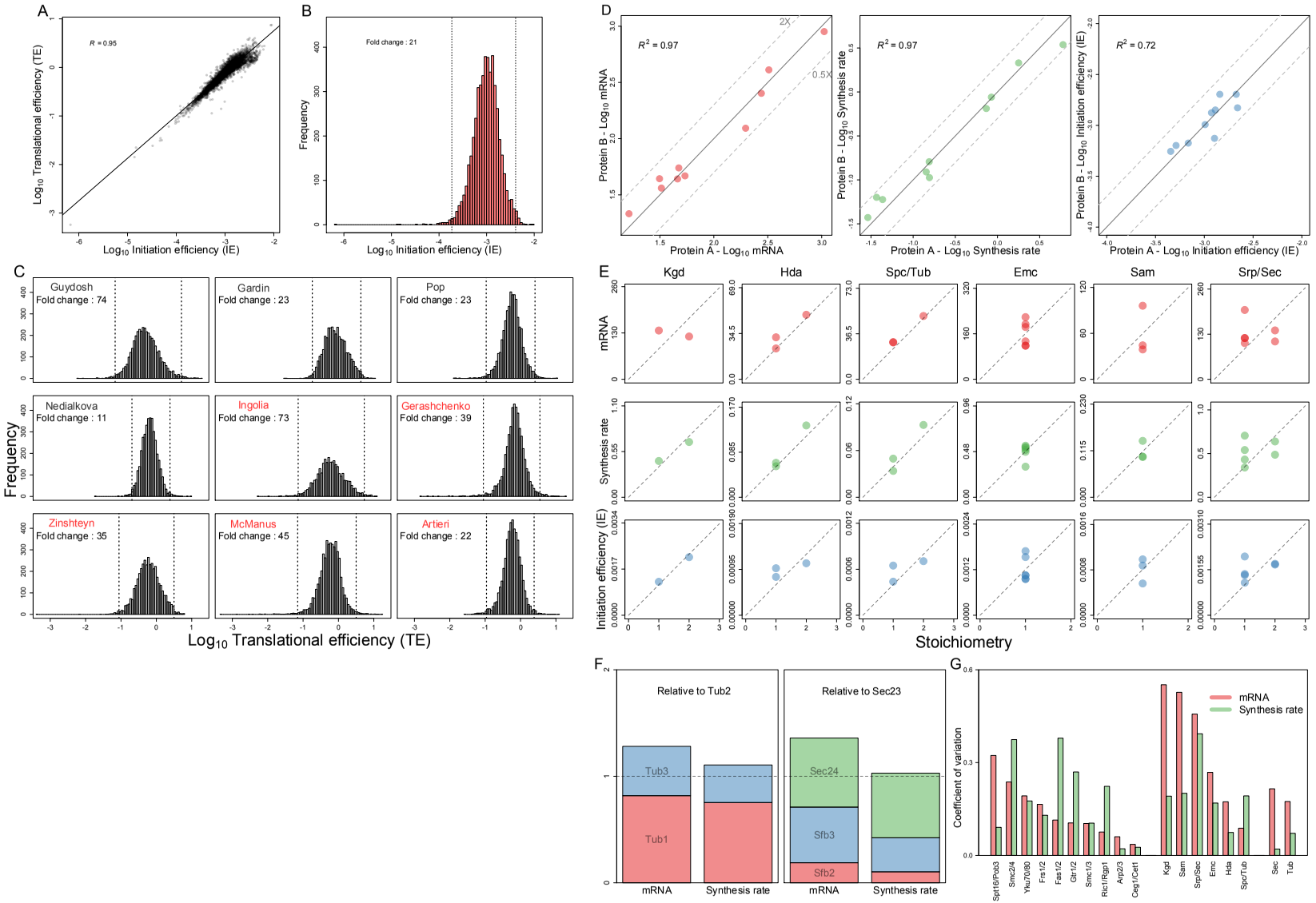
A
R = 0.95
Log10 Translational efficiency (TE)
Log10 Initiation efficiency (IE)
B
Fold change : 21
Frequency
Log10 Initiation efficiency (IE)
C
Guydosh
Fold change : 74
Gardin
Fold change : 23
Pop
Fold change : 23
Nedialkova
Fold change : 11
Ingolia
Fold change : 73
Gerashchenko
Fold change : 39
Zinshteyn
Fold change : 35
McManus
Fold change : 45
Artieri
Fold change : 22
Frequency
Log10 Translational efficiency (TE)
D
R2 = 0.97
2X
0.5X
Protein B - Log10 mRNA
Protein A - Log10 mRNA
R2 = 0.97
Protein B - Log10 Synthesis rate
Protein A - Log10 Synthesis rate
R2 = 0.72
Protein B - Log10 Initiation efficiency (IE)
Protein A - Log10 Initiation efficiency (IE)
E
Kgd
Hda
Spc/Tub
Emc
Sam
Srp/Sec
mRNA
Synthesis rate
Initiation efficiency (IE)
Stoichiometry
F
Relative to Tub2
Tub3
Tub1
Relative to Sec23
Sec24
Sfb3
Sfb2
mRNA
Synthesis rate
mRNA
Synthesis rate
G
mRNA
Synthesis rate
Coefficient of variation
Spt16/Pob3
Smc2/4
Yhu70/80
Frs1/2
Fas1/2
Gtr1/2
Smc1/3
Ric1/Rgp1
Arp2/3
Ceg1/Cet1
Kgd
Sam
Srp/Sec
Emc
Hda
Spc/Tub
Sec
Tub

Figure S6 *(preceding page)*: Analyses of IEs and TEs in log-phase yeast cells, Related to Figure 5. (**A**) Relationship between TE values (TE, Eqn. S6) and IE values (p^E, Eqn. S27). (**B**) Distribution of IE values. Otherwise, as in Figure 5A. (**C**) Distributions of TE measurements from nine published ribosome-profiling experiments. Ribosome-profiling datasets are labeled by the first author's name: Guydosh (Guydosh & Green, 2014), Gardin (Gardin et al., 2014), Pop (Pop et al., 2014), Nedialkova (Nedialkova & Leidel, 2015), Ingolia (Ingolia et al., 2009), Gerashchenko (Gerashchenko et al., 2012), Zinshteyn (Zinshteyn & Gilbert, 2013), McManus (McManus et al., 2014), Artieri (Artieri & Fraser, 2014). Datasets using CHX-pretreatment names in red and datasets without CHX-pretreatment named in gray. Otherwise, as in Figure 5A. (**D**) mRNA abundance and translational control contribute to proportional synthesis, Analysis of complexes with two equimolar subunits. For each heterodimeric complex, the mRNA abundances (left), estimated synthesis rates (middle), and IEs (right) of the individual subunits are plotted (with the subunit whose systematic name is first alphabetically on the x axis). All of the heterodimeric complexes characterized in Li et al. (2014) are shown, with the exception of the Smc2/4 complex, which substantially deviates from proportional synthesis in both our dataset and the published dataset. Dashed lines indicate 2-fold differences. (**E**) Analysis of multi-protein complexes. For each complex, the mRNA abundances, synthesis rates, and IEs of its subunits are plotted as a function of subunit stoichiometry. Dashed line passes through the origin and mean of the data-points. (**F**) Analyses of complexes containing paralogous subunits. For each complex, data for alternative subunits are plotted in the same column relative to the data for the constitutive subunit (Tub2 for the $\alpha\beta$-tubulin complex, Sec23 for the COPII vesicle coat). (**G**) Comparison of the differences in mRNA abundances with those of synthesis rates for each of the complexes in (**D**) and (**E**). Data were normalized to subunit stoichiometry, and coefficients of variation (CVs) were calculated using all subunits within each complex. CVs were calculated similarly for the Sec23- and Tub2-containing complexes, except that data for paralogous subunits (Tub1 and Tub3; Sec24, Sfb2, and Sfb3) were first summed.

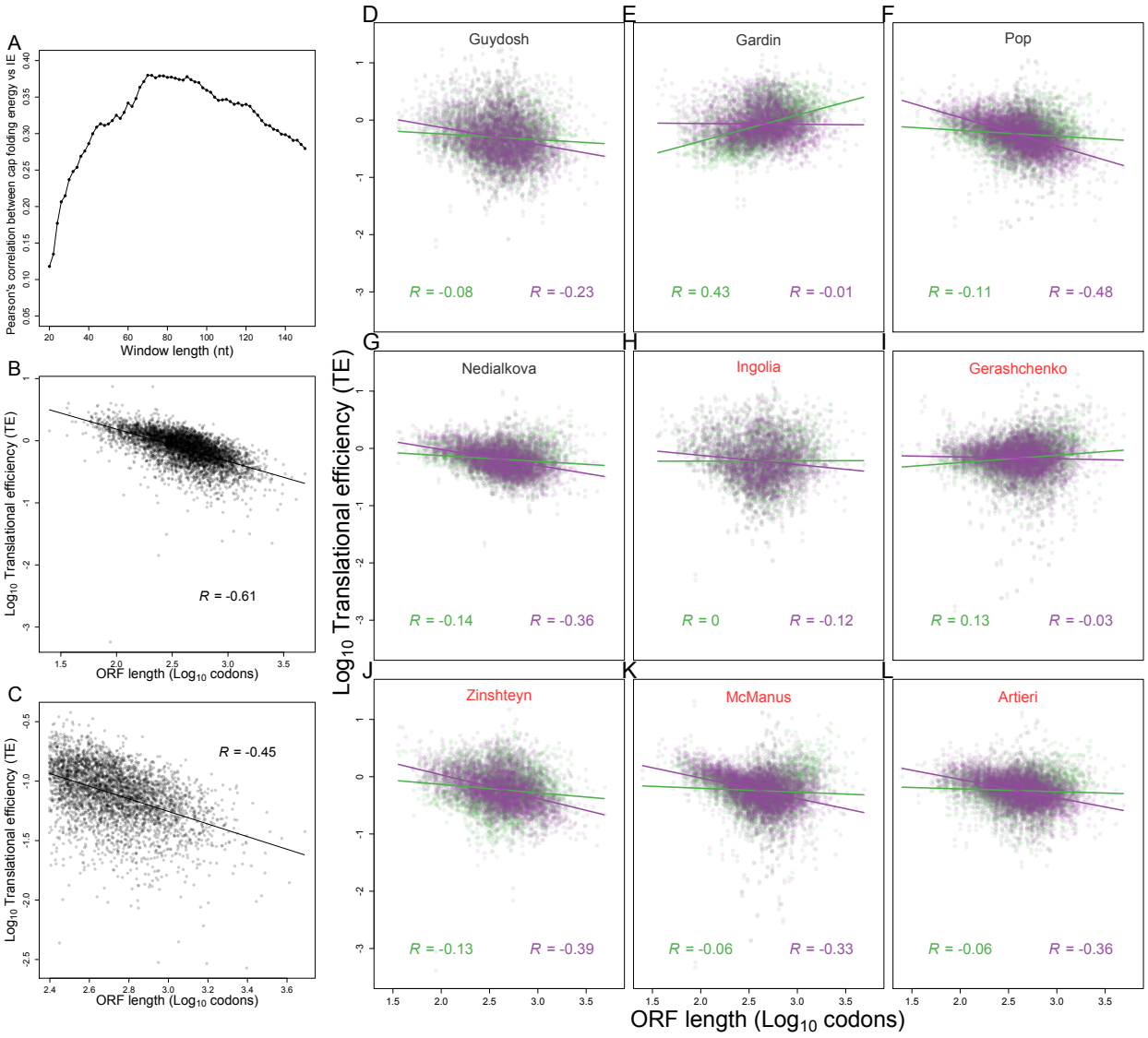

Figure S7: Analyses of the relationship between gene length and TE, Related to Figure 6. (**A**) Influence of window length on the correlation between predicted RNA structure and IE. Correlations between predicted RNA secondary structures near the 5' cap (5' cap FE) and IE (p^E, Eqn. S27), as a function of the length of the folded sequence. RNAfold was used to estimate RNA folding energies for windows of the indicated lengths downstream of the 5'-cap. (**B**) Relationship between TE and ORF length observed when analyzing the flash-freeze dataset. The best linear least-squares fit to the data is shown (line), with the Pearson R. (**C**) Relationship between TE and ORF length observed after excluding data from within the 5' ramp. Only genes with 250 codons were considered, and RPF and RNA-seq reads mapping to the first 200 codons of each ORF were excluded. Otherwise, as in (**B**). (**D-L**) Relationship between TE and ORF length and the role of 3' bias observed in RNA-seq reads in published ribosome-profiling datasets. For each ORF, mRNA RPKM calculated based on reads mapping to the entire ORF is shown as green dots. mRNA RPKM when calculated based only on reads mapping to the last 200 codons of an ORF are shown as purple dots. Ribosome-profiling datasets are labeled by the first author's name: Guydosh (Guydosh & Green, 2014), Gardin (Gardin et al., 2014), Pop (Pop et al., 2014), Nedialkova (Nedialkova & Leidel, 2015), Ingolia (Ingolia et al., 2009), Gerashchenko (Gerashchenko et al., 2012), Zinshteyn (Zinshteyn & Gilbert, 2013), McManus (McManus et al., 2014), Artieri (Artieri & Fraser, 2014). Datasets using CHX-pretreatment names in red and datasets without CHX-pretreatment named in gray.

Table S1 - Correlations between estimates of tRNA abundances and codon-specific ribosome-densities, Related to Figure 2

Pearson's correlation	*(tRNA GCN * Wobble)*	*tRNA abundance (Microarray * Wobble)*	*tAI*	*RiboDensity at A-site*
*tRNA abundance (RNA-seq * Wobble)*	0.748	0.608	0.713	0.268
*(tRNA GCN * Wobble)*		0.778	0.947	0.512
*tRNA abundance (Microarray * Wobble)*			0.713	0.459
tAI				0.509

Spearman's correlation	*(tRNA GCN * Wobble)*	*tRNA abundance (Microarray * Wobble)*	*tAI*	*RiboDensity at A-site*
*tRNA abundance (RNA-seq * Wobble)*	0.777	0.605	0.752	0.319
*(tRNA GCN * Wobble)*		0.785	0.917	0.560
*tRNA abundance (Microarray * Wobble)*			0.711	0.463
tAI				0.532

Table S2 - Codon-specific estimates of ribosome-densities and tRNA abundances, Related to Figure 2.

AA	Codon	tRNA (anticodon)	tRNA gene copy number	Wobble	(tRNA GCN * Wobble)	tAI	tRNA abundance (RNA-seq)	tRNA abundance (RNA-seq * Wobble)	tRNA abundance (Microarray)	tRNA abundance (Microarray * Wobble)	RiboDensity at A-site
A	GCA	TGC	5	1	5	0.30795	384682	384682	26087	26087	1.4224238
A	GCC	AGC	11	0.64	7.04	0.487685		136141.44		41927.04	0.6982879
A	GCG	TGC	5	0.64	3.2	0.098522		246196.48		16695.68	1.7913012
A	GCT	AGC	11	1	11	0.67734	212721	212721	65511	65511	0.7439787
C	TGC	GCA	4	1	4	0.246305	106137	106137	84453	84453	1.3124677
C	TGT	GCA	4	0.64	2.56	0.108128		67927.68		54049.92	0.8129126
D	GAC	GTC	16	1	16	0.985222	1789916	1789916	105539	105539	0.9749501
D	GAT	GTC	16	0.64	10.24	0.432512		1145546.24		67544.96	0.9771609
E	GAA	TTC	14	1	14	0.862069	1618235	1618235	121285	121285	0.9358888
E	GAG	CTC	2	1	2	0.399015	393379	393379	NA	8663.214286	1.3490997
F	TTC	GAA	10	1	10	0.615764	104391	104391	81287	81287	0.6956495
F	TTT	GAA	10	0.64	6.4	0.27032		66810.24		52023.68	0.686277
G	GGA	TCC	3	1	3	0.184729	287325	287325	61092	61092	1.5260655
G	GGC	GCC	16	1	16	0.985222	826342	826342	70286	70286	1.4416823
G	GGG	CCC	2	1	2	0.182266	129338	129338	21404	21404	1.7256386
G	GGT	GCC	16	0.64	10.24	0.432512		528858.88		44983.04	1.0667916
H	CAC	GTG	7	1	7	0.431034	727396	727396	81702	81702	0.9154641
H	CAT	GTG	7	0.64	4.48	0.189224		465533.44		52289.28	0.9922004
I	ATA	TAT	2	1	2	0.123233	54352	54352	39556	39556	1.394014
I	ATC	AAT	13	0.64	8.32	0.576355		586391.04		67769.6	0.7039792
I	ATT	AAT	13	1	13	0.800493	916236	916236	105890	105890	0.6284445
K	AAA	TTT	7	1	7	0.431034	222273	222273	82386	82386	0.978243
K	AAG	CTT	14	1	14	1	397111	397111	83036	83036	1.0943089
L	CTA	TAG	3	1	3	0.184729	94608	94608	63675	63675	0.9165169
L	CTC	GAG	1	1	1	0.061576	63503	63503	13613	13613	1.1024675
L	CTG	TAG	3	0.64	1.92	0.059113		60549.12		40752	1.4565058
L	CTT	GAG	1	0.64	0.64	0.027032		40641.92		8712.32	0.9176823
L	TTA	TAA	7	1	7	0.431034	271592	271592	49562	49562	0.7272969
L	TTG	CAA	10	1	10	0.753695	709768	709768	96605	96605	0.7904184
M	ATG	CAT	10	1	10	0.615764	266917	266917	67121	67121	0.8415256
N	AAC	GTT	10	1	10	0.615764	378101	378101	110849	110849	0.6898425
N	AAT	GTT	10	0.64	6.4	0.27032		241984.64		70943.36	0.7078108
P	CCA	TGG	10	1	10	0.615776	286531	286531	112091	112091	1.0842912
P	CCC	AGG	2	0.64	1.28	0.08867		36940.8		11043.84	1.3106768
P	CCG	TGG	10	0.64	6.4	0.197044		183379.84		71738.24	3.1205299
P	CCT	AGG	2	1	2	0.123153	57720	57720	17256	17256	1.0239082
Q	CAA	TTG	9	1	9	0.554187	436225	436225	89917	89917	0.8659004
Q	CAG	CTG	1	1	1	0.238916	89093	89093	NA	9990.777778	1.6325536
R	AGA	TCT	11	1	11	0.67734	683001	683001	98864	98864	1.1308364
R	AGG	CCT	1	1	1	0.278325	106890	106890	12911	12911	1.939357
R	CGA	ACG	6	0.61	3.66	0.000037		261137.95		36380.4	3.6442612
R	CGC	ACG	6	0.64	3.84	0.26601		273980.8		38169.6	1.6633606
R	CGG	CCG	1	1	1	0.061576	60792	60792	15330	15330	4.0202765
R	CGT	ACG	6	1	6	0.369458	428095	428095	59640	59640	1.0459032
S	TCA	TGA	3	1	3	0.184797	86913	86913	58804	58804	1.0015308
S	TCC	AGA	11	0.64	7.04	0.487685		507436.8		58654.72	0.7072965
S	TCG	CGA	1	1	1	0.12069	34793	34793	49376	49376	1.3549748
S	TCT	AGA	11	1	11	0.67734	792870	792870	91648	91648	0.6624122
T	ACA	TGT	4	1	4	0.246373	105862	105862	47598	47598	1.3252482
T	ACC	AGT	11	0.64	7.04	0.487685		244374.4		36396.16	0.6976534
T	ACG	CGT	1	1	1	0.140394	113104	113104	16861	16861	1.7572598
T	ACT	AGT	11	1	11	0.67734	381835	381835	56869	56869	0.692477
V	GTA	TAC	2	1	2	0.123239	145923	145923	33845	33845	0.9784524
V	GTC	AAC	14	0.64	8.96	0.62069		515950.72		56675.84	0.6841453
V	GTG	CAC	2	1	2	0.162562	118162	118162	44651	44651	1.2778867
V	GTT	AAC	14	1	14	0.862069	806173	806173	88556	88556	0.6336201
W	TGG	CCA	6	1	6	0.369458	298500	298500	60236	60236	1.7898831
Y	TAC	GTA	8	1	8	0.492611	343867	343867	91388	91388	0.9352348
Y	TAT	GTA	8	0.64	5.12	0.216256		220074.88		58488.32	0.8397308
Z	AGC	GCT	2	1	2	0.123153	249948	249948	61334	61334	1.1973314
Z	AGT	GCT	2	0.64	1.28	0.054064		159966.72		39253.76	0.9498069

Table S3 – Analysis of 3' bias in RNA-seq datasets, Related to Figure 4.

Pearson correlation between mRNA abundances (Log_{10} RPKM)	**Arava**	**Holstege**	**Causton**	**Sun**
Ingolia	0.645	0.743	0.626	0.74
Gerashchenko	0.7	0.828	0.747	0.862
Zinshteyn	0.652	0.773	0.645	0.75
McManus	0.667	0.79	0.733	0.797
Artieri	0.715	0.852	0.788	0.848
Guydosh	0.66	0.777	0.666	0.791
Gardin	0.673	0.828	0.728	0.846
Pop	0.727	0.845	0.774	0.894
Nedialkova	0.722	0.878	0.8	0.884
Unselected	0.753	0.875	0.819	0.883
Ribo-Zero	0.761	0.869	0.812	0.882
Ribo-Minus	0.722	0.839	0.76	0.84
Dynabeads	0.719	0.869	0.795	0.897
Pearson correlation between ratio of mRNA abundances and ORF length (Log_{10})	Relative to Arava	Relative to Holstege	Relative to Causton	Relative to Sun
Ingolia	-0.653	-0.341	-0.331	-0.344
Gerashchenko	-0.607	-0.45	-0.406	-0.468
Zinshteyn	-0.619	-0.353	-0.325	-0.344
McManus	-0.551	-0.308	-0.306	-0.277
Artieri	-0.547	-0.403	-0.38	-0.367
Guydosh	-0.65	-0.436	-0.392	-0.43
Gardin	-0.638	-0.496	-0.44	-0.496
Pop	-0.496	-0.305	-0.298	-0.328
Nedialkova	-0.521	-0.272	-0.261	-0.237
Unselected	-0.112	0.206	0.072	0.217
Ribo-Zero	-0.107	0.209	0.077	0.226
Ribo-Minus	-0.061	0.24	0.105	0.244
Dynabeads	-0.51	-0.411	-0.376	-0.444

Table S4 - Excess mRNA coverage in 3’ relative to 5’ based on 50 codons (150 nt), Related to Figure 4.

	Metagene 5’ excess	Metagene 3’ excess	Metagene excess 3’/5’	Gene-specific median ratio 3’/5’	Gene-specific median excess 3’/5’
Unselected	0.891	1.096	23.00%	1.223	22.30%
RiboZero	0.879	1.125	28.00%	1.284	28.40%
RiboMinus	0.893	1.367	53.10%	1.532	53.20%
Dynabeads	0.549	2.104	283.20%	3.757	275.70%
Guydosh	0.782	1.284	64.20%	1.563	56.30%
Gardin	0.535	2.014	276.70%	3.586	258.60%
Pop	0.562	1.789	218.20%	2.918	191.80%
Nedialkova	0.753	1.17	55.50%	1.493	49.30%
Ingolia	0.812	1.234	52.00%	1.416	41.60%
Gerashchenko	0.786	1.206	53.50%	1.46	46.00%
Zinshteyn	0.587	2.137	263.70%	3.284	228.40%
McManus	0.709	1.368	93.00%	1.699	69.90%
Artieri	0.726	1.397	92.30%	1.791	79.10%

Table S5 – Correlations between RPF, mRNA and TE across profiling datasets, Related to Figure 5.

RPF	Guydosh	Lareau	Gardin	Pop	Jan	Williams	Nedialkova	Ingolia	Gerashchenko	Zinshteyn	McManus	Artieri
Flash-freeze	0.98	0.962	0.96	0.986	0.971	0.949	0.986	0.982	0.972	0.971	0.973	0.972
Guydosh		0.966	0.942	0.983	0.957	0.934	0.98	0.979	0.961	0.973	0.958	0.971
Lareau			0.937	0.975	0.938	0.886	0.971	0.975	0.941	0.951	0.929	0.95
Gardin				0.96	0.946	0.908	0.964	0.94	0.95	0.923	0.954	0.957
Pop					0.967	0.935	0.987	0.986	0.973	0.961	0.962	0.97
Jan						0.978	0.968	0.957	0.952	0.954	0.964	0.955
Williams							0.934	0.926	0.926	0.943	0.946	0.935
Nedialkova								0.978	0.975	0.961	0.967	0.974
Ingolia									0.965	0.98	0.961	0.962
Gerashchenko										0.945	0.966	0.952
Zinshteyn											0.966	0.96
McManus												0.967

mRNA	Guydosh	Gardin	Pop	Nedialkova	Ingolia	Gerashchenko	Zinshteyn	McManus	Artieri
Flash-freeze	0.77	0.818	0.888	0.909	0.686	0.844	0.749	0.816	0.883
Guydosh		0.942	0.921	0.899	0.971	0.908	0.952	0.888	0.886
Gardin			0.939	0.902	0.923	0.926	0.933	0.857	0.901
Pop				0.937	0.882	0.955	0.878	0.9	0.934
Nedialkova					0.853	0.941	0.866	0.928	0.943
Ingolia						0.884	0.936	0.849	0.842
Gerashchenko							0.869	0.898	0.922
Zinshteyn								0.858	0.865
McManus									0.93

TE	Guydosh	Gardin	Pop	Nedialkova	Ingolia	Gerashchenko	Zinshteyn	McManus	Artieri
Flash-freeze	0.542	-0.025	0.655	0.594	0.479	0.498	0.55	0.553	0.561
Guydosh		0.626	0.854	0.797	0.924	0.674	0.873	0.702	0.806
Gardin			0.553	0.448	0.701	0.632	0.607	0.43	0.574
Pop				0.795	0.838	0.81	0.802	0.804	0.854
Nedialkova					0.818	0.685	0.744	0.761	0.727
Ingolia						0.729	0.907	0.777	0.797
Gerashchenko							0.633	0.774	0.757
Zinshteyn								0.673	0.731
McManus									0.784

Table S6 – Model selection of determinants of initiation efficiencies, Related to Figure 7.

Model - Initiation Efficiency	R^2
log(pE) ~ (5' cap folding energy)	0.144
log(pE) ~ log(Length)	0.2295
log(pE) ~ log(mRNA)	0.3925
log(pE) ~ log(Length) + (5' cap folding energy)	0.3262
log(pE) ~ log(mRNA) + (5' cap folding energy)	0.448
log(pE) ~ log(Length) + log(mRNA) +(5' cap folding energy)	0.5521
log(pE) ~ log(Length) + log(mRNA) + (5' cap folding energy) + (5' UTR length)	0.5604
log(pE) ~ log(Length) + log(mRNA) + (5' cap folding energy) + (5' UTR length) + (# uATGs)	0.5658
log(pE) ~ log(Length) + log(mRNA) + (5' cap folding energy) + (5' UTR length) + (# uATGs) + (UTR GC)	0.5675
log(pE) ~ log(Length) + log(mRNA) + (5' cap folding energy) + (5' UTR length) + (# uATGs presence/absence)	0.5784
log(pE) ~ log(Length) + log(mRNA) + (5' cap folding energy) + (5' UTR length) + (# uATGs presence/absence) + (UTR GC)	0.5812