Journal club on Ribo-seq

2/10/2015

Han Fang

Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling

Nicholas T. Ingolia,* Sina Ghaemmaghami,† John R. S. Newman, Jonathan S. Weissman

Published Online February 12 2009
Science 10 April 2009:
Vol. 324 no. 5924 pp. 218–223
DOI: 10.1126/science.1168978

Mammalian microRNAs predominantly act to decrease target mRNA levels

Huili Guo[1,2], Nicholas T. Ingolia[3,4], Jonathan S. Weissman[3,4] & David P. Bartel[1,2]

Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes

Nicholas T. Ingolia,[1,3,*] Liana F. Lareau,[2] and Jonathan S. Weissman[1]

Ribosome profiling: new views of translation, from single codons to genome scale

Nicholas T. Ingolia

PubMed "ribosome profiling"

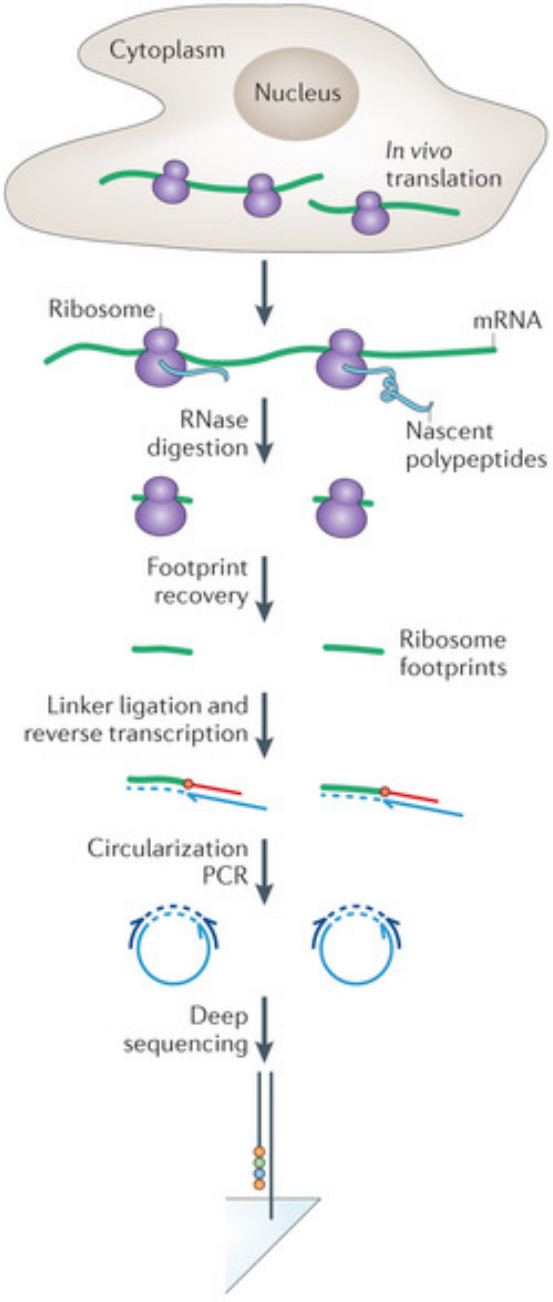
RSS Save search Advanced

Display Settings: Summary, 20 per page, Sorted by Recently Added

Results: 1 to 20 of 121 << First <

From 2014/01/01

Results: 1 to 20 of 66



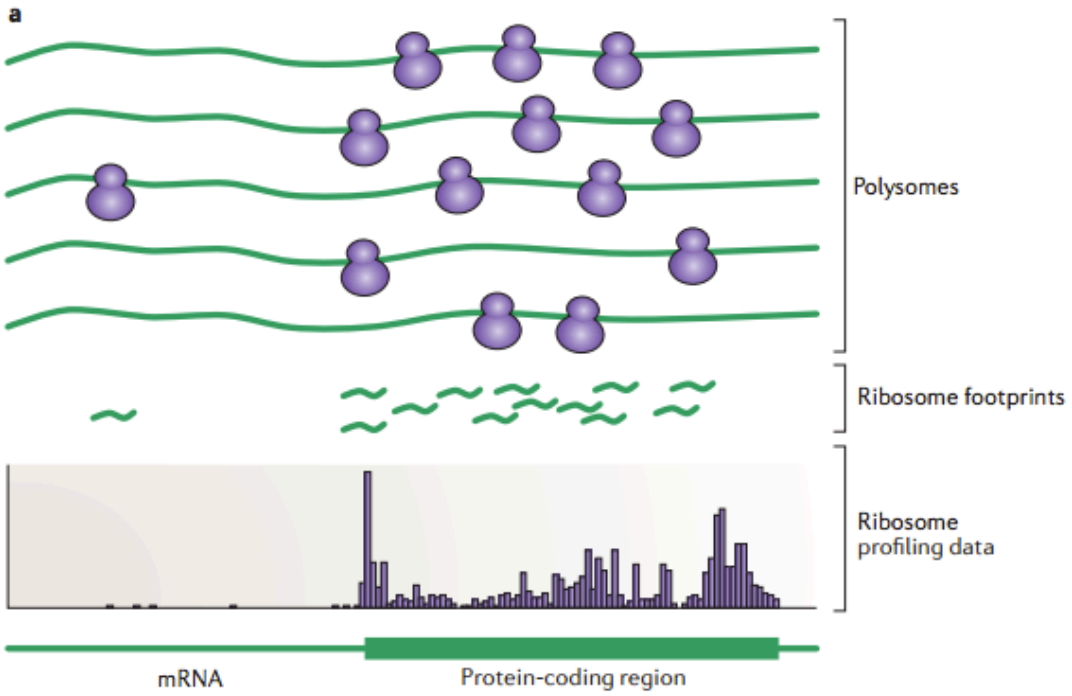
Limitations

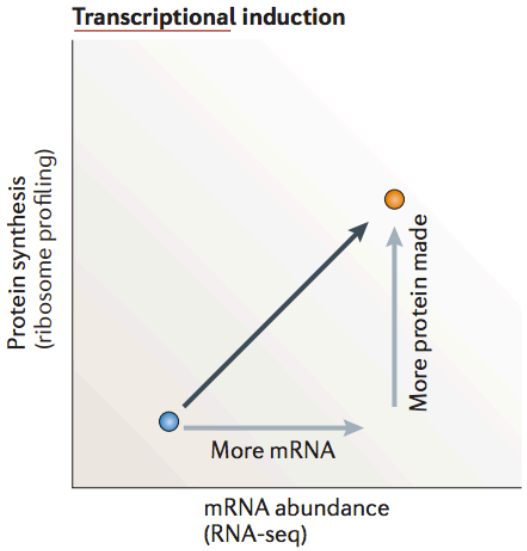
- Depends on mapping.
- Ribosome-protected fragments are fairly short (~30bp); (repetitive sequences, alternative transcripts, longer or PE seq unavailable.
- Existing metrics/statistics, e.g. FPKM by Cufflinks, and log2 ratio by edgeR is by definition not suitable for Ribo-seq abundance representation.
- Both mRNA & riboseq focus on the current rate of protein production and not on the total abundance of a protein.

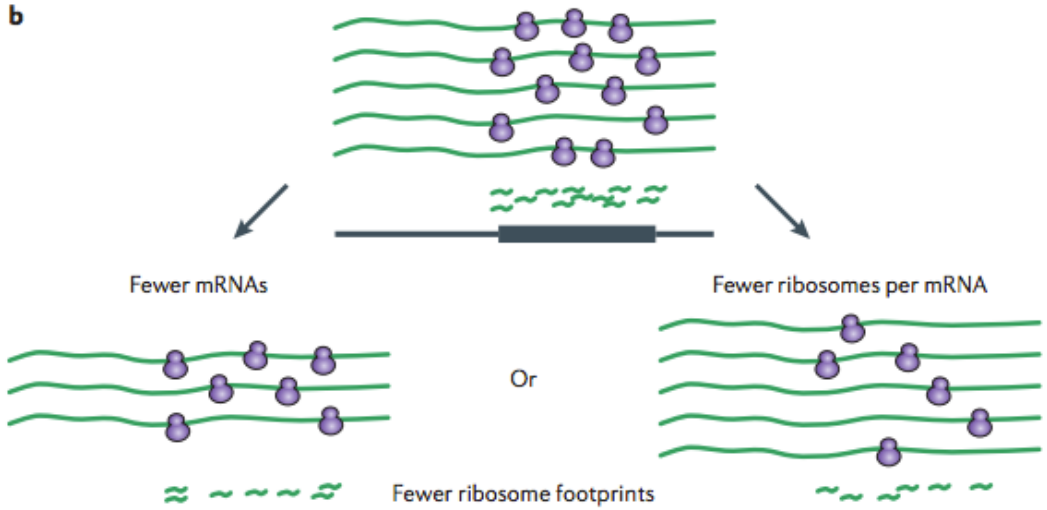
- Discard some information that is found in polysome profiles due to the footprints, i.e. foot prints rather than entire transcripts.
- Nuclease digestion degrades the 5´ and 3´ UTRs of transcripts, which may contain regulatory information.
- Obscures the presence of distinct mRNA subpopulations

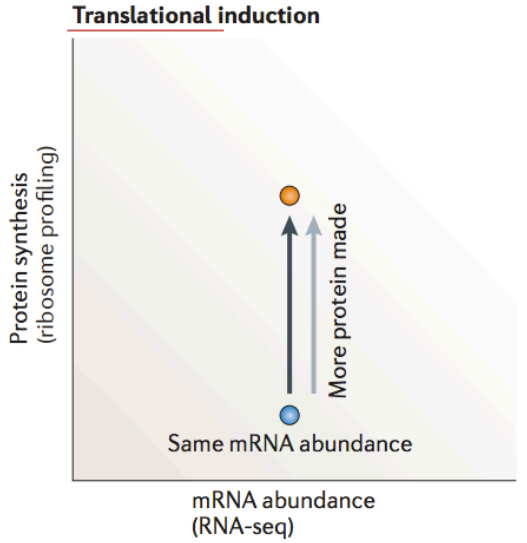
Ingolia, Nat Rev Genet (2014)

Analysis of ribosome occupancy data









Ingolia, Nat Rev Genet (2014)

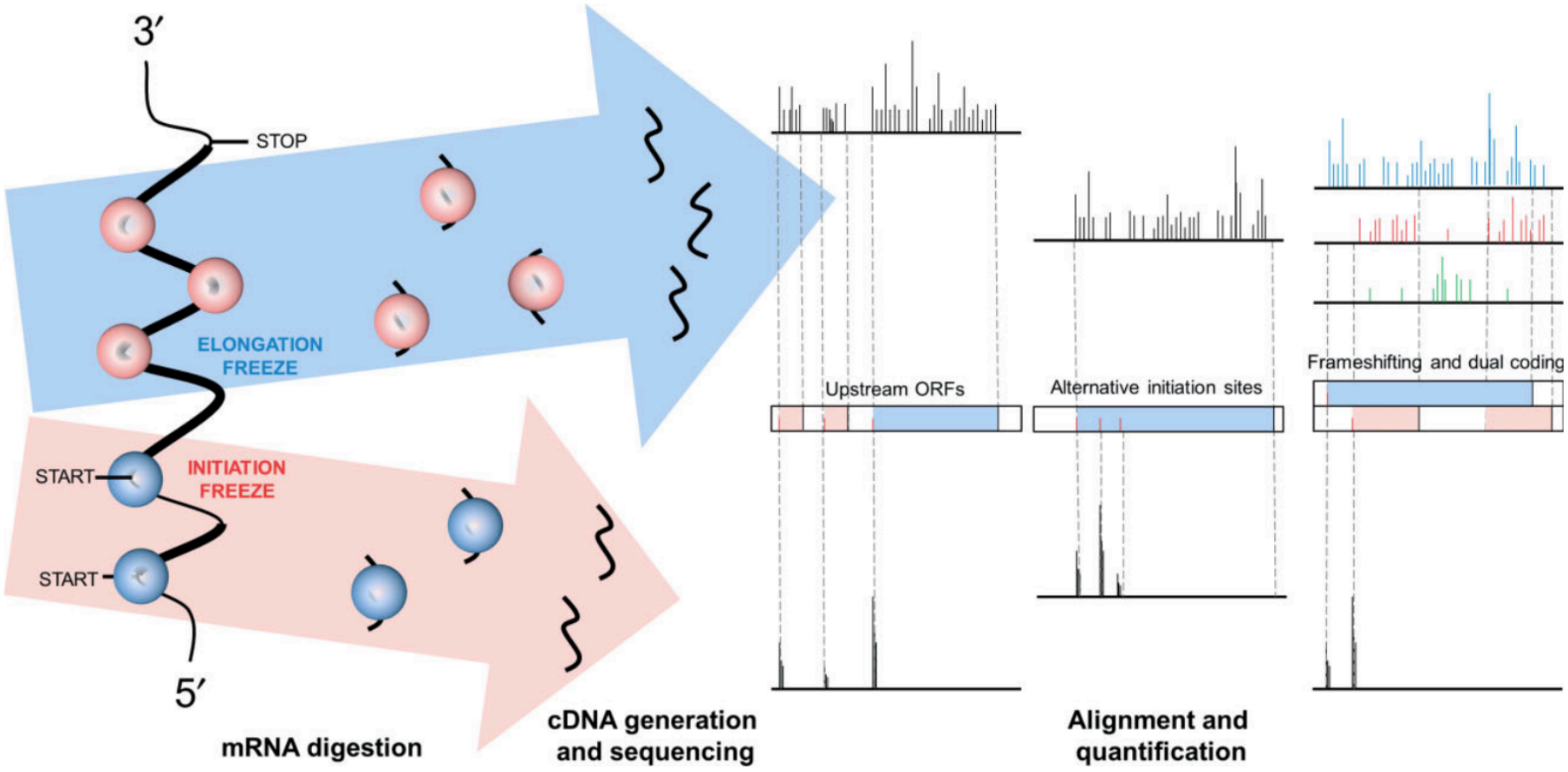


Fig 3 Michel and Baranov (2013)

Published online: December 23, 2014

Article

SOURCE DATA | TRANSPARENT PROCESS | OPEN ACCESS

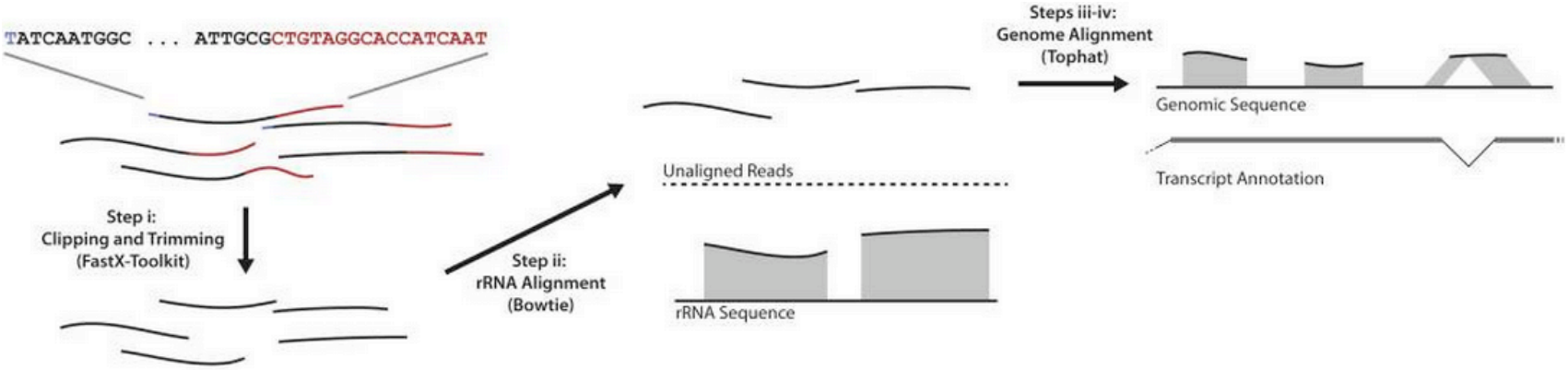
molecular systems biology

Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation

Cristina Pop[1,*], Silvi Rouskin[2], Nicholas T Ingolia[3], Lu Han[4], Eric M Phizicky[4], Jonathan S Weissman[2] & Daphne Koller[1]

1a. Data pre-processing

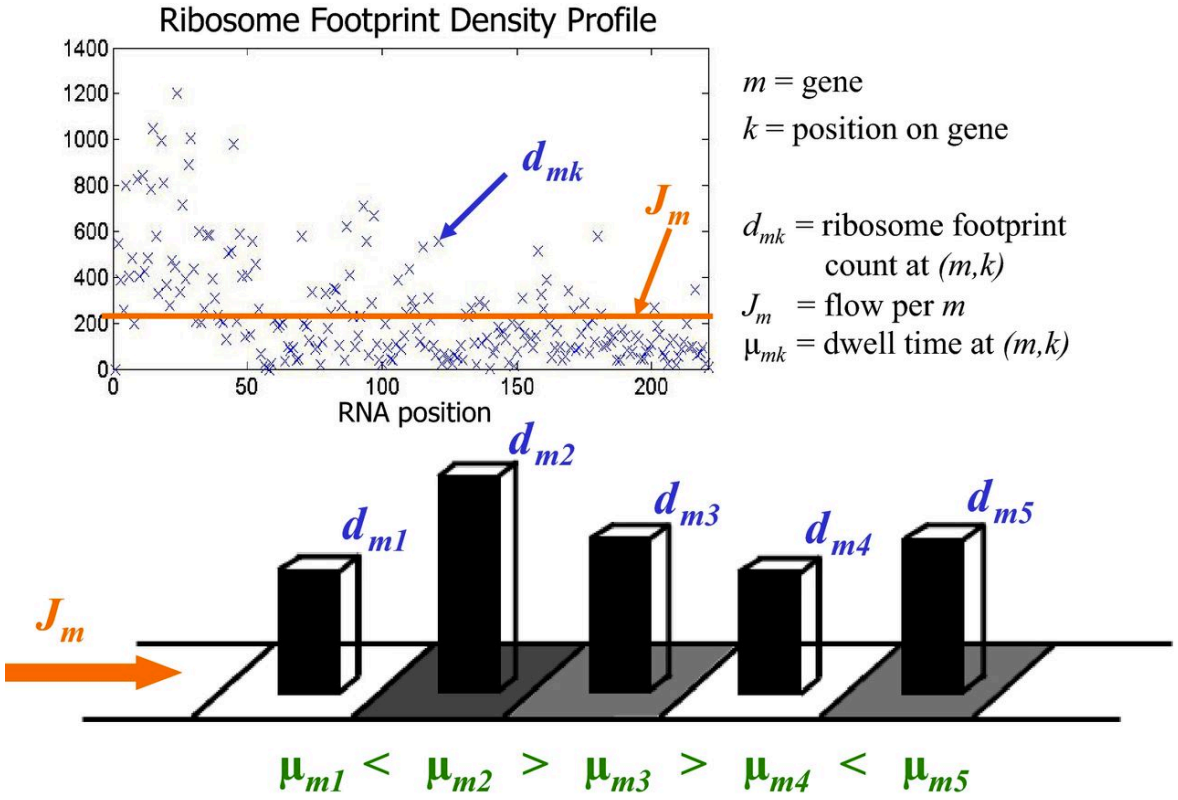
- RPF were aligned against S288C assembly R63.
- kept reads: uniquely mapped & no more than two mismatches & lengths between 28 and 31.



- Genes were ignored if: did not have an AUG start codon, had internal stop codons, had < 50% of positions on the coding sequence with at least one mapped mRNA count, or if all the footprint counts were 0 over the gene.
- codon usage bias is measure by the tRNA adaptation index (tAI)

Ingolia et al., (2012) & Pop (2014)

1b. Queuing model for elongation process



count at position = flow * dwell at position

$$d_{mk} = J_m \mu_{mk}$$

Figure 1

- use a single μ_m^c for every copy of codon c on message m.
- add a pseudo-count of 1 to all FP counts and use the logarithm of normalized counts in the Poisson term
- during model training, ignore the first 100 codons

$$\max_{\mu_m^c, \mu^c} \log \prod_{m,k'} \mu_m^{c(d'_{mk}/J_m)} \exp\left(-\mu_m^c\right) - C\left[\sum_{m,c} w_m^c \left(\log \mu_m^c - \log \mu^c\right)^2\right]$$

log likelihood term assuming Poisson dist.

L2 soft constraint term

1c. Choice of C did not affect results —correlations for codon bias measures, protein abundance, and outliers

$$\max_{\mu_m^c,\mu^c} \log \prod_{m,k'} \mu_m^{c(d'_{mk}/J_m)} \exp\left(-\mu_m^c\right) - C\left[\sum_{m,c} w_m^c \left(\log \mu_m^c - \log \mu^c\right)^2\right]$$

Result		const = 1	const = 10	const = 1000	const = 10000	const = 100000	No μ_m^c
μ^c (const=100)	r	1.000	1.000	1.000	1.000	1.000	1.000
	p	1E-202	5E-206	9E-150	1E-105	6E-95	1E-96
μ_m^c (const=100)	r	1.000	1.000	1.000	0.983	0.838	NA
	p	0	0	0	0	0	NA
J_m (const=100)	r	1.000	1.000	1.000	1.000	0.999	0.994
	p	0	0	0	0	0	0
tAI	r	0.210	0.210	0.210	0.213	0.217	0.211
	p	0.104	0.104	0.104	0.100	0.094	0.103
tRNA abund (Cy5)	r	0.144	0.144	0.140	0.140	0.140	0.133
	p	0.380	0.4380	0.393	0.393	0.393	0.420
tRNA abund (Cy3)	r	0.144	0.144	0.140	0.140	0.140	0.133
	p	0.417	0.417	0.429	0.429	0.429	0.456
PA (Newman et al)	r	0.7885	0.7885	0.7886	0.7889	0.7882	0.7782
PA (de Godoy et al)	r	0.6802	0.6802	0.6802	0.6802	0.6786	0.6710

Table S5

1d. Optimization

$$\max_{\mu_m^c,\mu^c} \log \prod_{m,k'} \mu_m^{c(d'_{mk}/J_m)} \exp\left(-\mu_m^c\right) - C\left[\sum_{m,c} w_m^c \left(\log \mu_m^c - \log \mu^c\right)^2\right]$$

1) $Fix\ J_m\ to\ D_m = \sum_{m \in k} d_{mk}/L_m$

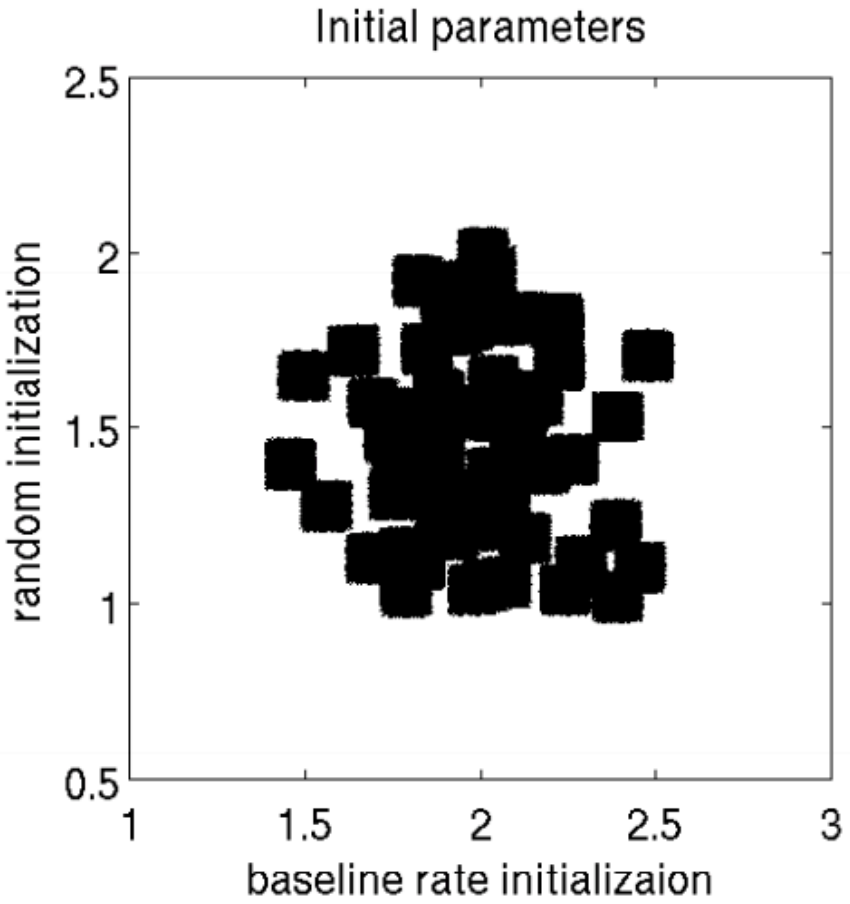
2) Initialize μ_m^c & μ_m *by averaging over counts*(m,k) *for codon*$(m,k) = c$

3) Estimate parameters by iterating through codons c and learning μ_m^c & μ_m

4) Stop codons are excluded

5) Compute $\mathrm{J}_m = \sum_{k \in m} \frac{d_{mk}}{\mu_m} / L_m = \sum_{k \in m} \frac{d_{mk}}{\mu_m^{c=codon(m,k)}} / L_m$

1e. Optimization is not sensitive to initiation



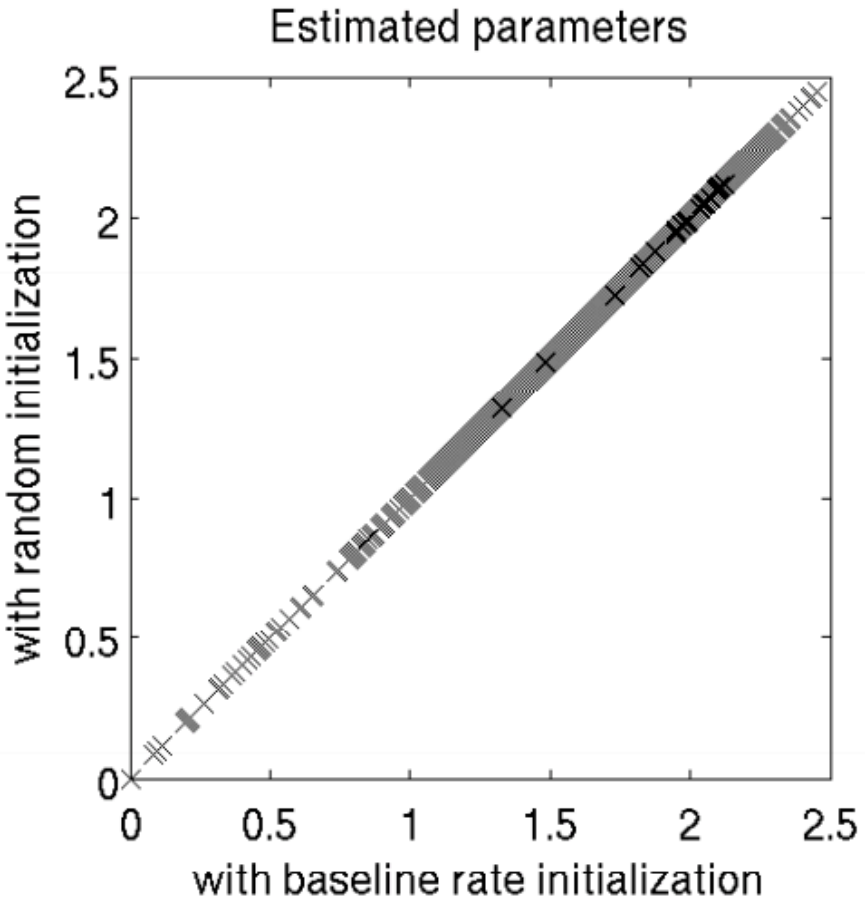


Figure S11

1f. Small improvement over baseline average counts

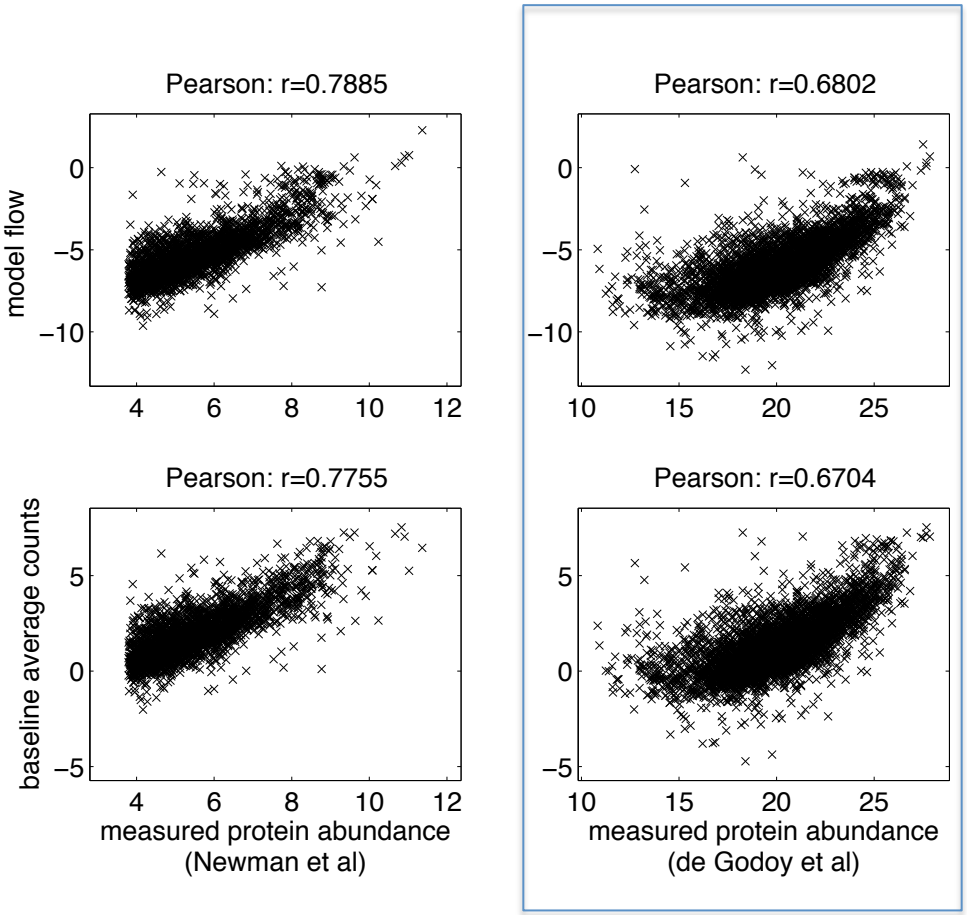


Figure S1, Pop et al (2014)

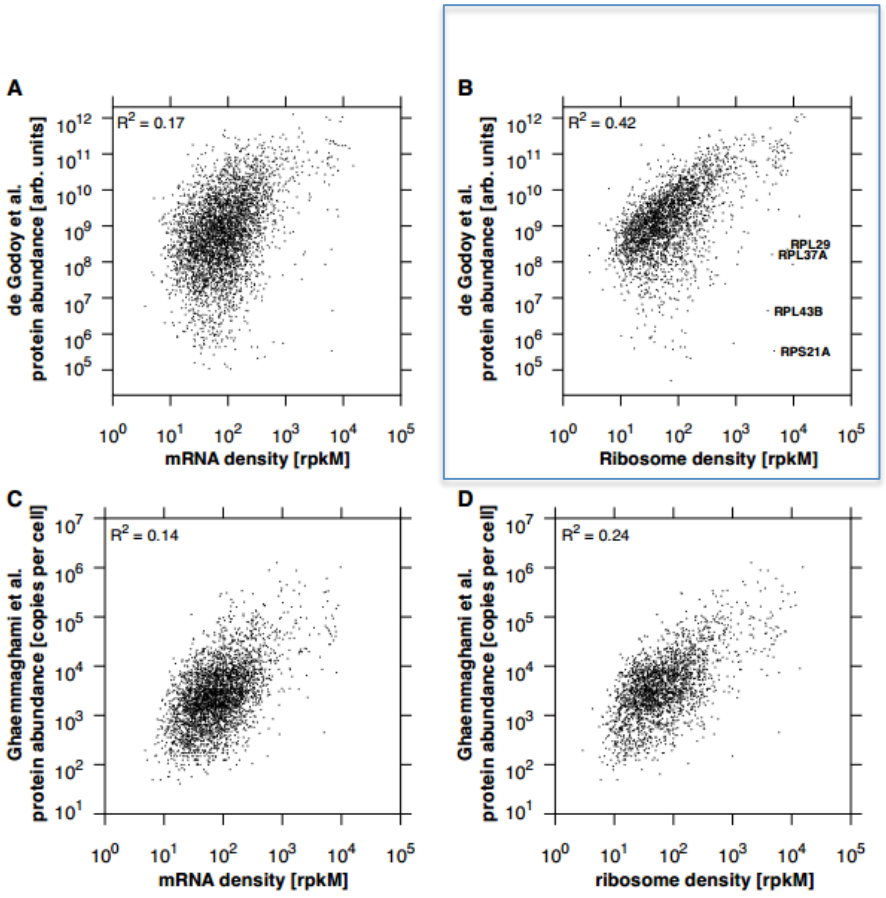


Figure S9, Ingolia et al (2009)

2a. Codon translation is not affected by tRNA abundance or body sequence

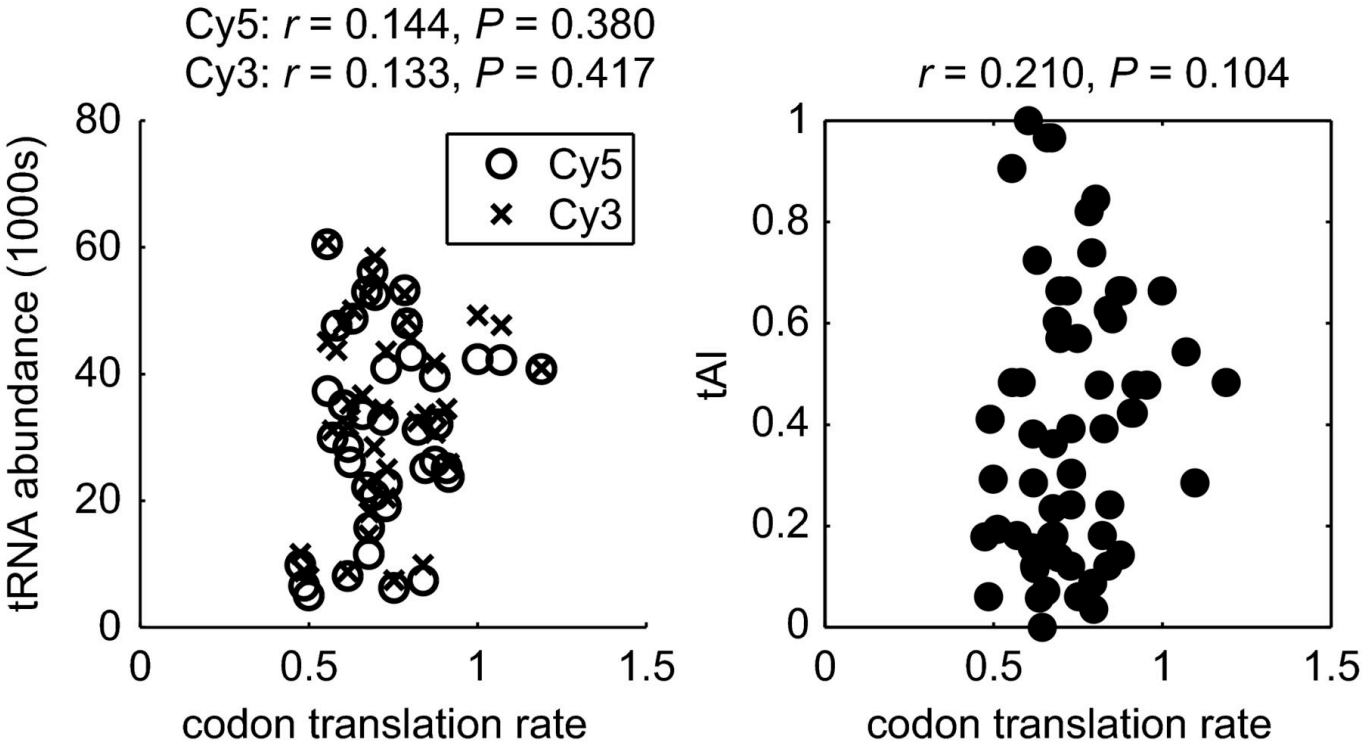


Figure 2

The same insignificant correlation exists in the raw footprint data (r = 0.112, P = 0.392)

Left: Insignificant Spearman correlation between estimated codon translation rates and tRNA abundance from microarray measurements using either fluorophore **Cy3 or Cy5** on 39 codons with measured levels.

Right: The correlation to tAI) is also not significant. (tAI: a measure of codon bias based on tRNA gene copy number relative to the overall collection of isoacceptor tRNAs)

2b. measure the effect of tRNA abundance on codon translation rate

- Created three mutant yeast species to test whether

(i) tRNA **overexpression** speeds up translation: **AGG-OE (13)**

(ii) the tRNA **body itself** causes the tRNA-dependent rate effect observed in other studies: **AGG-QC (similar)**

(iii) **depletion** of tRNA slows down ribosomes: **ACA-K (0.3)**

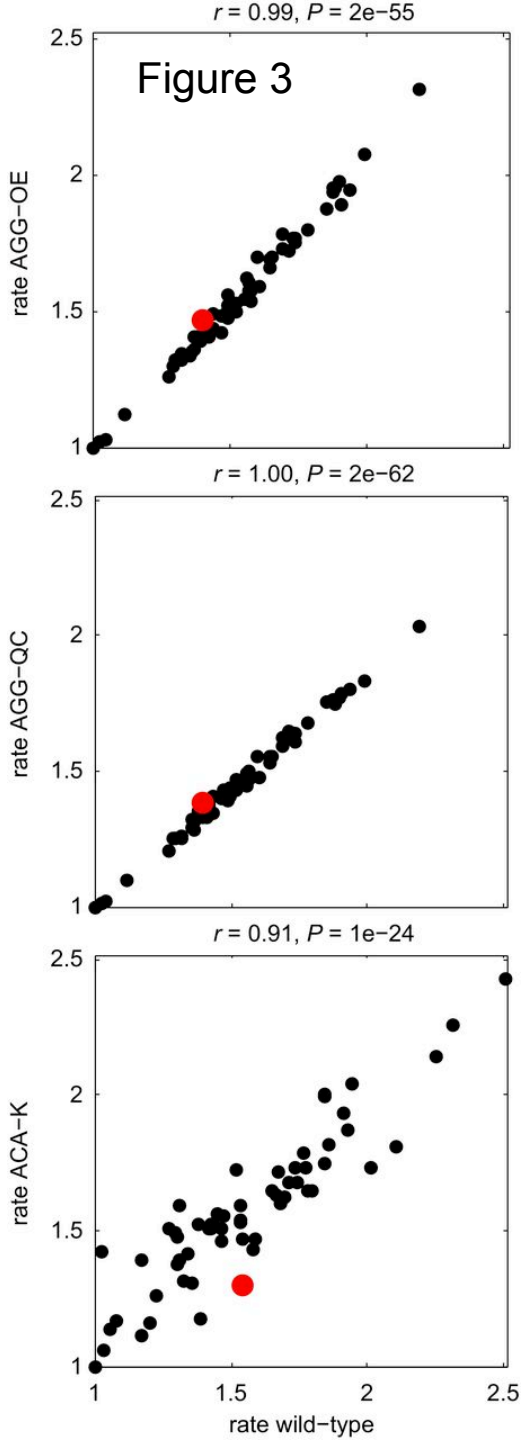
- Generated ribosome profiling data and ran model on these mutants to test whether

i. AGG codons are translated **faster** in AGG-OE and AGG-QC

(AGG is a rare codon)

i. ACA codons are translated **slower** in ACA-K.

(ACA is a heavily used codon)



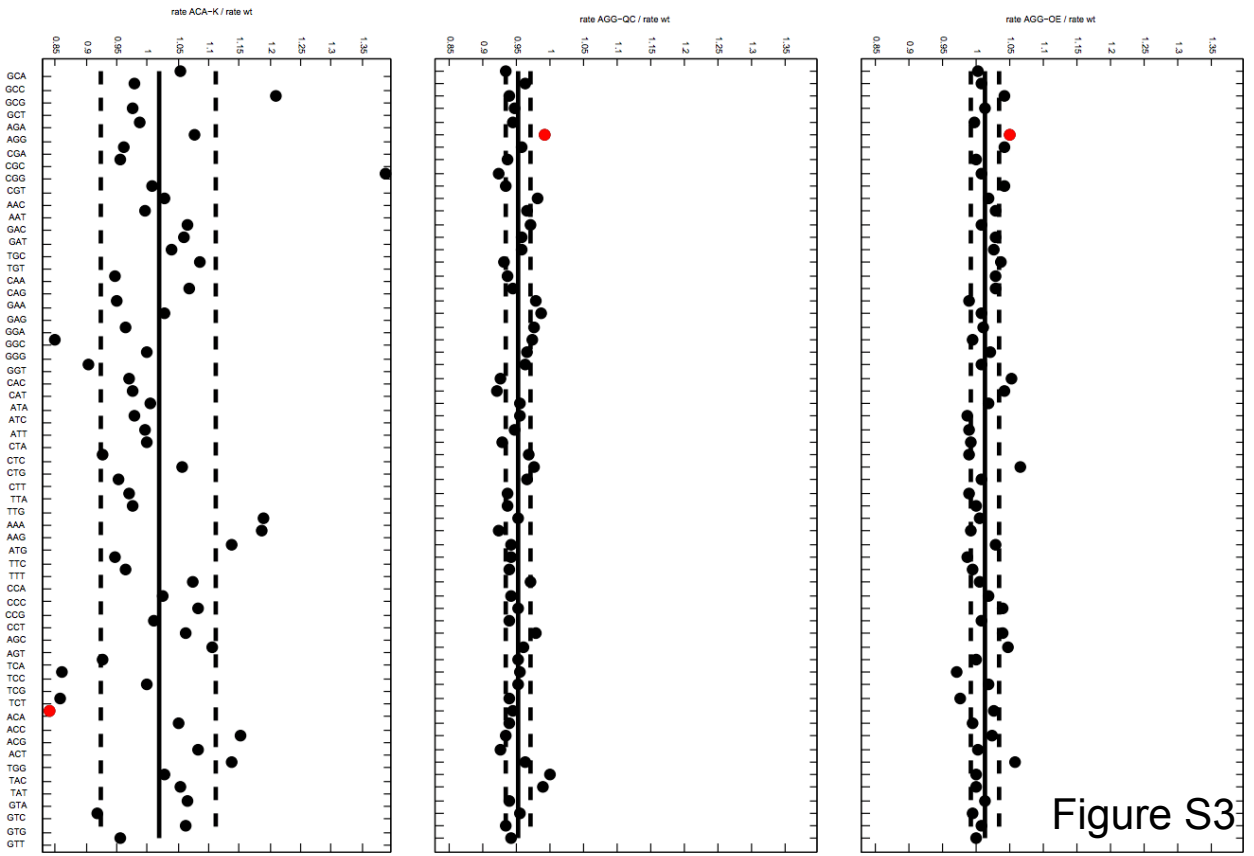
2c. no significant change in the elongation rates of the affected codon in any of the three mutants compared to wild-type

Several-fold changes in tRNA abundance do not affect ribosome dwell time.

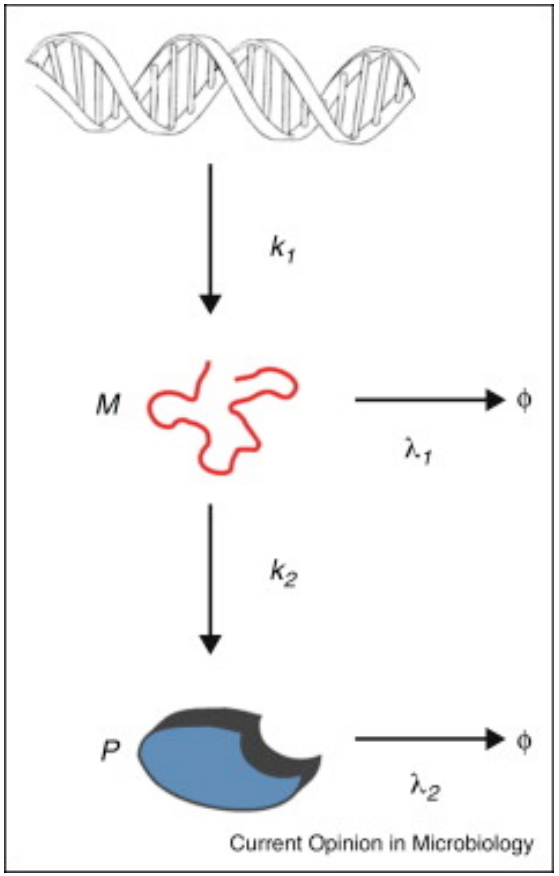
Rate ACA-K/ rate wt

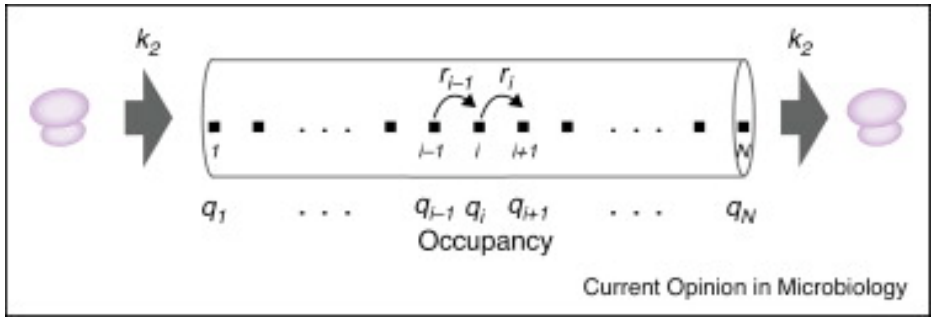
Rate AGG-QC/ rate wt

Rate AGG-OE/ rate wt



3a. Translation efficiency (TE)





q_i ~ the probability that the ith codon is occupied by a ribosome.
R_i ~ the ribosome translocation rate constant from i to i + 1.

$$\frac{d}{dt}q_i = r_{i-1}q_{i-1} - r_i q_i$$

$$\begin{cases} \frac{d}{dt}q_1 = k_2 - r_1 q_1 \\ \frac{d}{dt}q_N = r_{N-1}q_{N-1} - k_2 \end{cases}$$

$$r_1 q_1 = \cdots = r_{i-1}q_{i-1} = r_i q_i = \cdots = k_2 \quad \text{->} \quad q_i = \frac{k_2}{r_i}$$

$$\frac{d}{dt}M = k_1 - \lambda_1 M$$

$$\frac{d}{dt}P = k_2 M - \lambda_2 P$$

$$\langle q \rangle = k_2 \left\langle \frac{1}{r} \right\rangle = k_2 \langle \tau \rangle$$

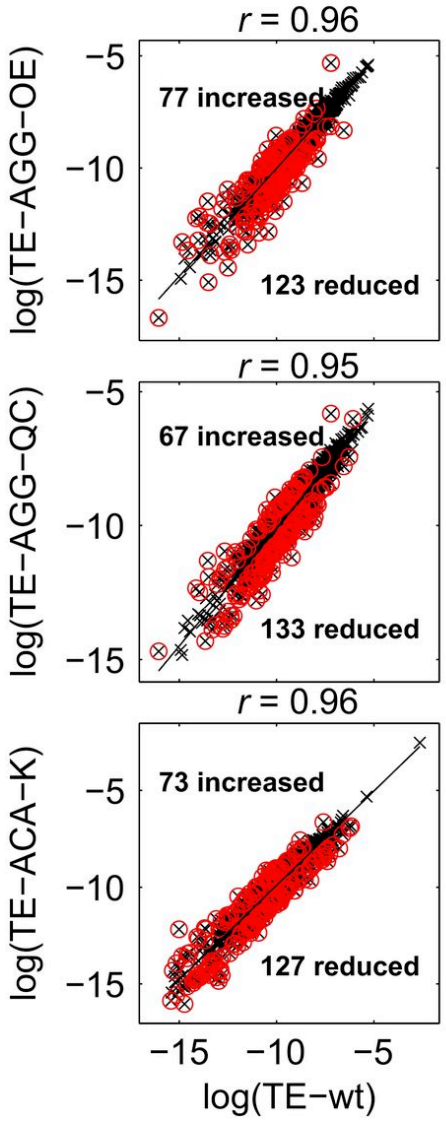
<τ> is the average translocation time per codon.

TE can be defined as the rate of protein production per mRNA, i.e TE= K_2.

$$k_2 = \frac{1}{\langle \tau \rangle}\frac{\langle q \rangle M}{M} = \frac{1}{\langle \tau \rangle}\frac{\text{ribsome density}}{\text{RNA level}}$$

Li 2015

3b. Translation efficiency is mildly affected by tRNA knockdown but not by overexpression



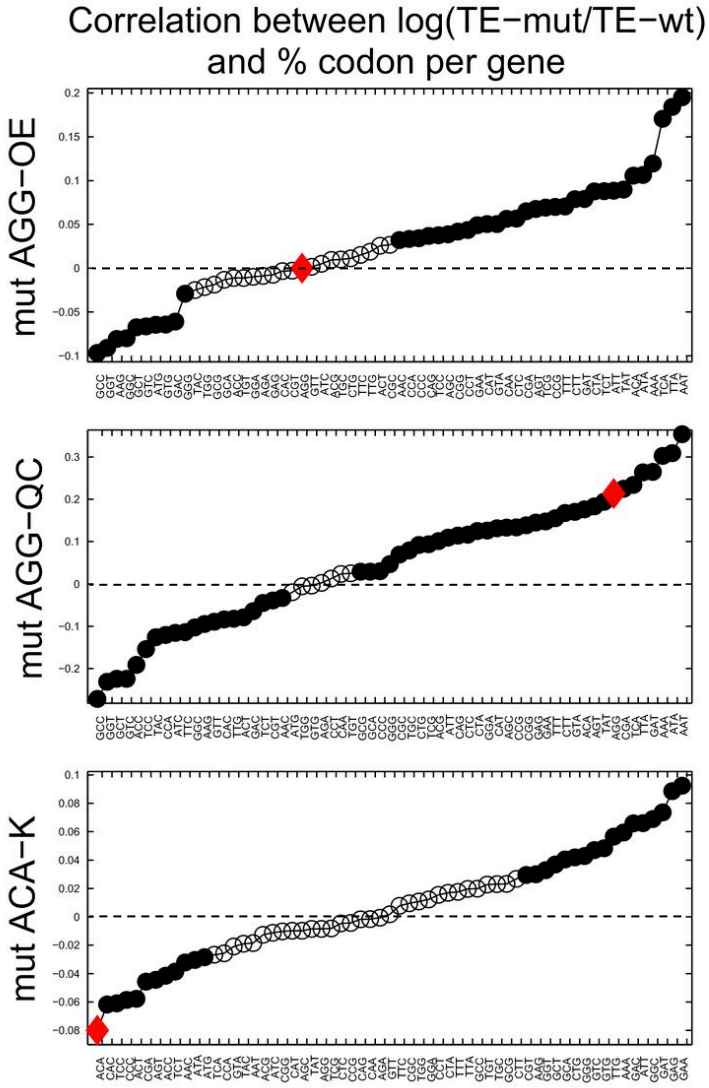


Figure 4

4a. Factors for elongation efficiency

- Key Q: What signals do affect elongation efficiency and translation efficiency?

i. **slow outliers**: at each position k along a message m as positions where ribosomes are stalled more than expected.

ii. **fast outliers**: stalled less than expected

$outlier\ strength\ \Delta_{mk}: the\ deviation\ from\ expected\ dwell\ time,$

$\Delta_{mk} = \frac{d_{mk} - E[d_{mk}]}{s_{mk}},$

where s_{mk} *is a sd representing the variance in that count due to gene abundance and codon.*

Divide genes into 32 quantiles (abundance) and compute var per codon.

Thus,

$$\begin{cases} slow\ outliers: \Delta_{mk} > T \\ fast\ outliers: \Delta_{mk} < T \\ non\ outliers: \Delta_{mk} \in (-1, 1) \end{cases}$$

4b. Model for translation efficiency (TE)

Learn through the whole training set

Learn through cross validation

$$\min_{w} \sum_{m} \left(TE_m - w^T f_m\right)^2 + \lambda_1 \sum_{p} |w_p| + \lambda_2 \sum_{p} w_p^2$$

Fit w to the TE of a set of gene{m}

Enforce sparsity & shrinkage

Create a null model where w is learned from TEs randomly permutated among genes.
Final w are the average over all training/testing combinations.

The features used are minimal to maximize the # of genes that have these characterized:

1) tAI of gene (codon usage bias);
2) computationally predicted energy of 5' UTR, 3' UTR, mRNA,
3) window around the start codon with highest correlation with TE;
4) length of coding sequence;
5) mRNA abundance;
6) identity of bases overlapping the Kozac site.

4c. All codons show negative correlation between outlier strength and proximity to gene start

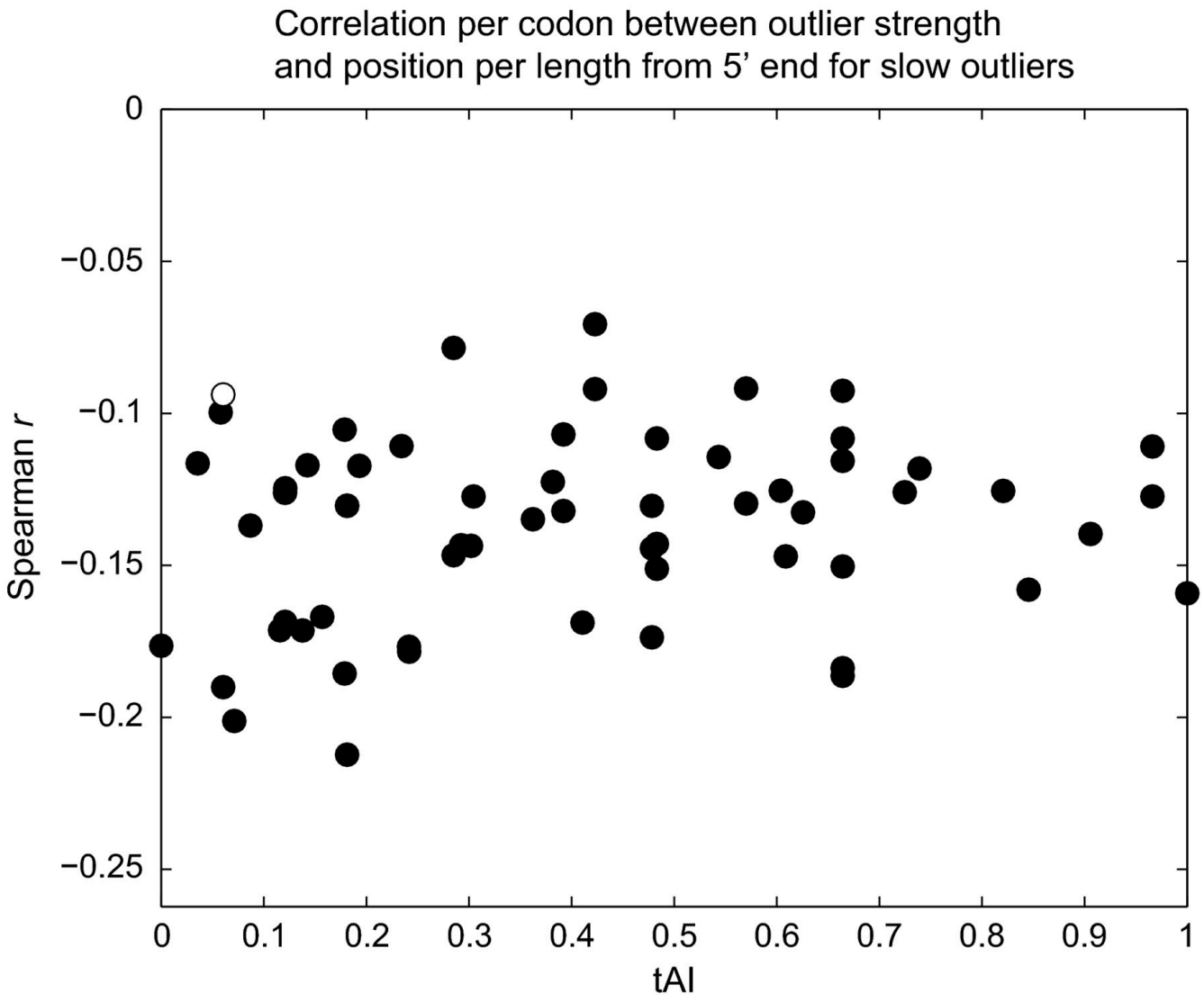


Figure 5

5a. Structural features and sequence motif around the start codon

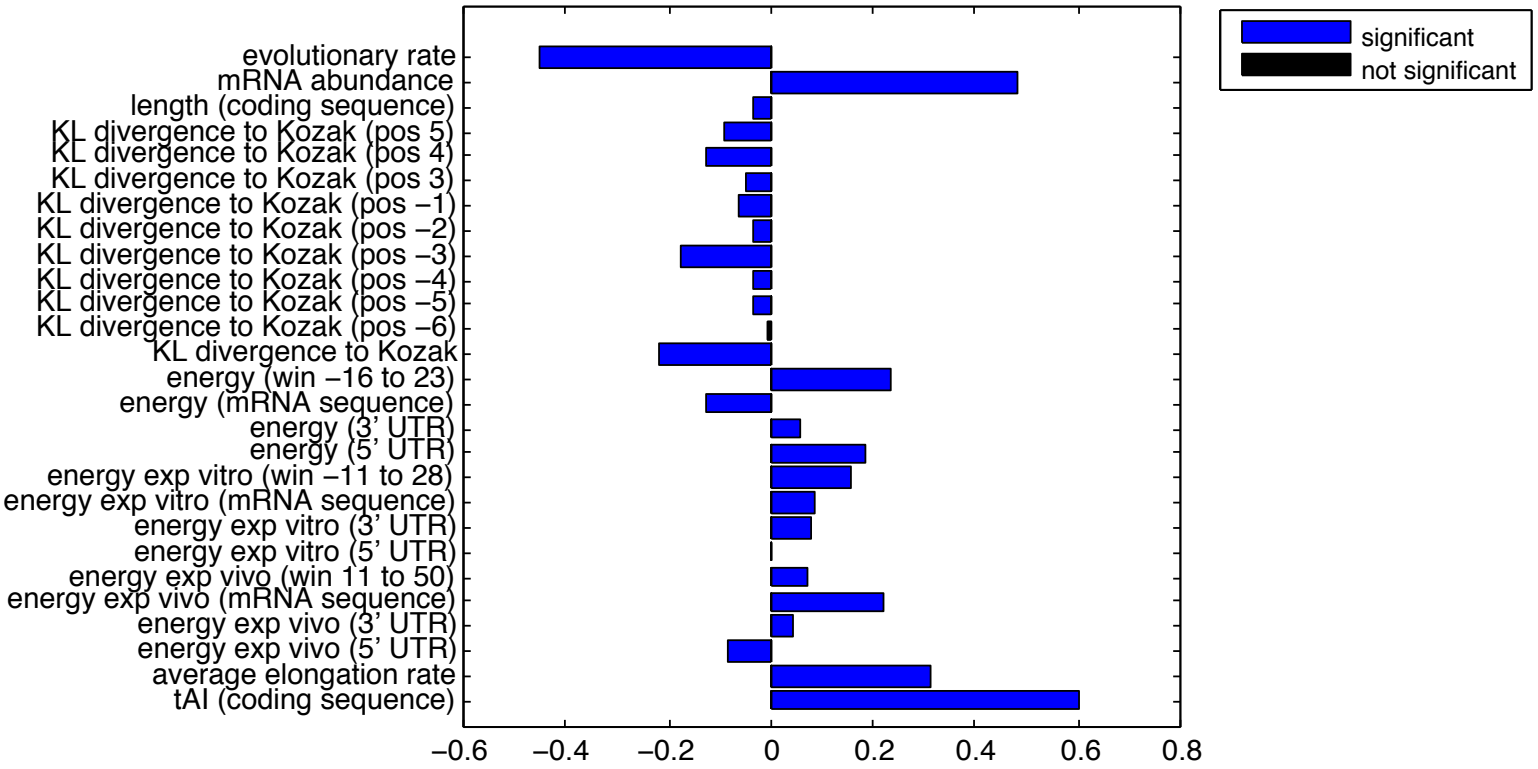


Figure S7

5b. RNA structure energy and its relationship to translation efficiency

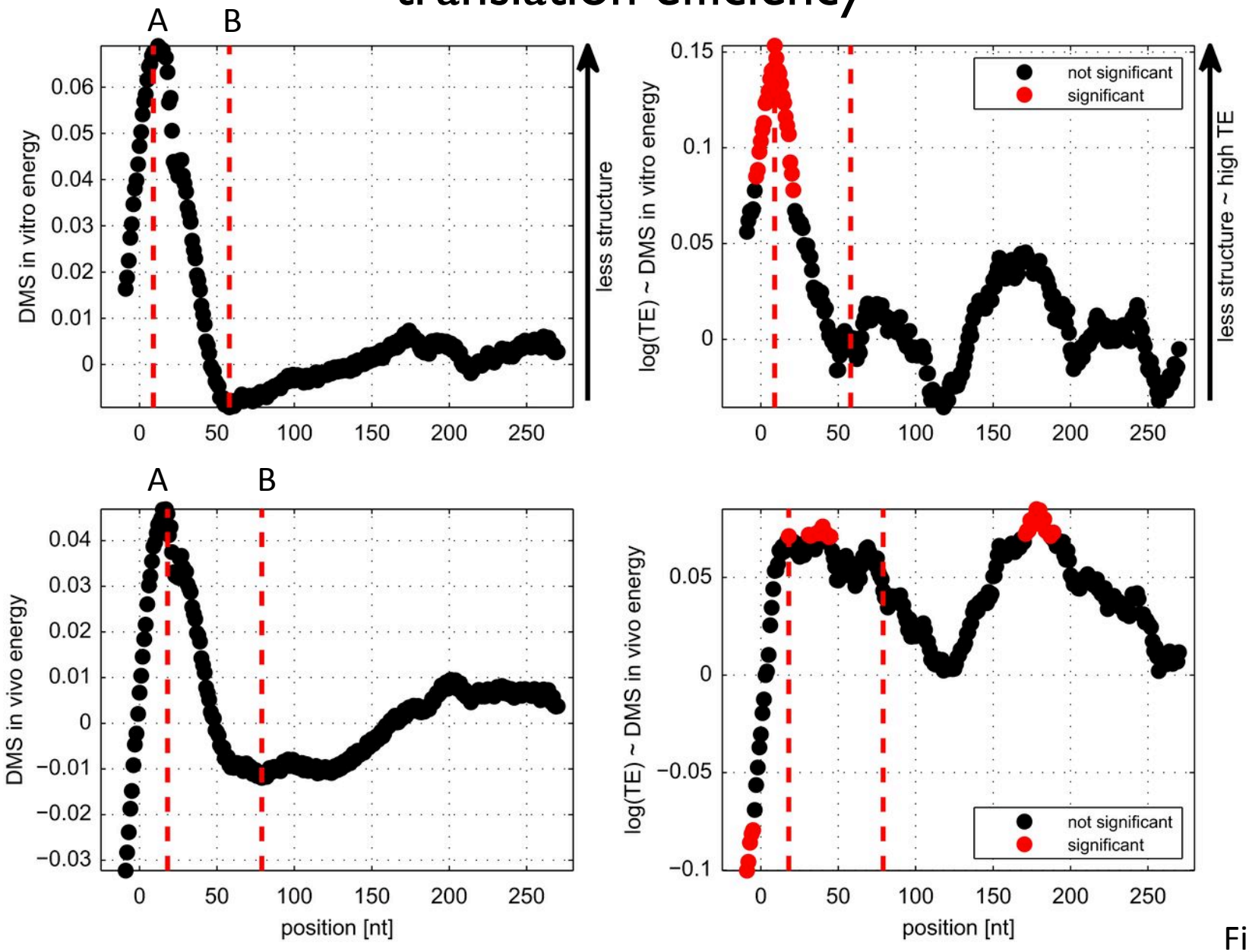


Figure 6

5c. Estimated Kozak motif for efficient genes

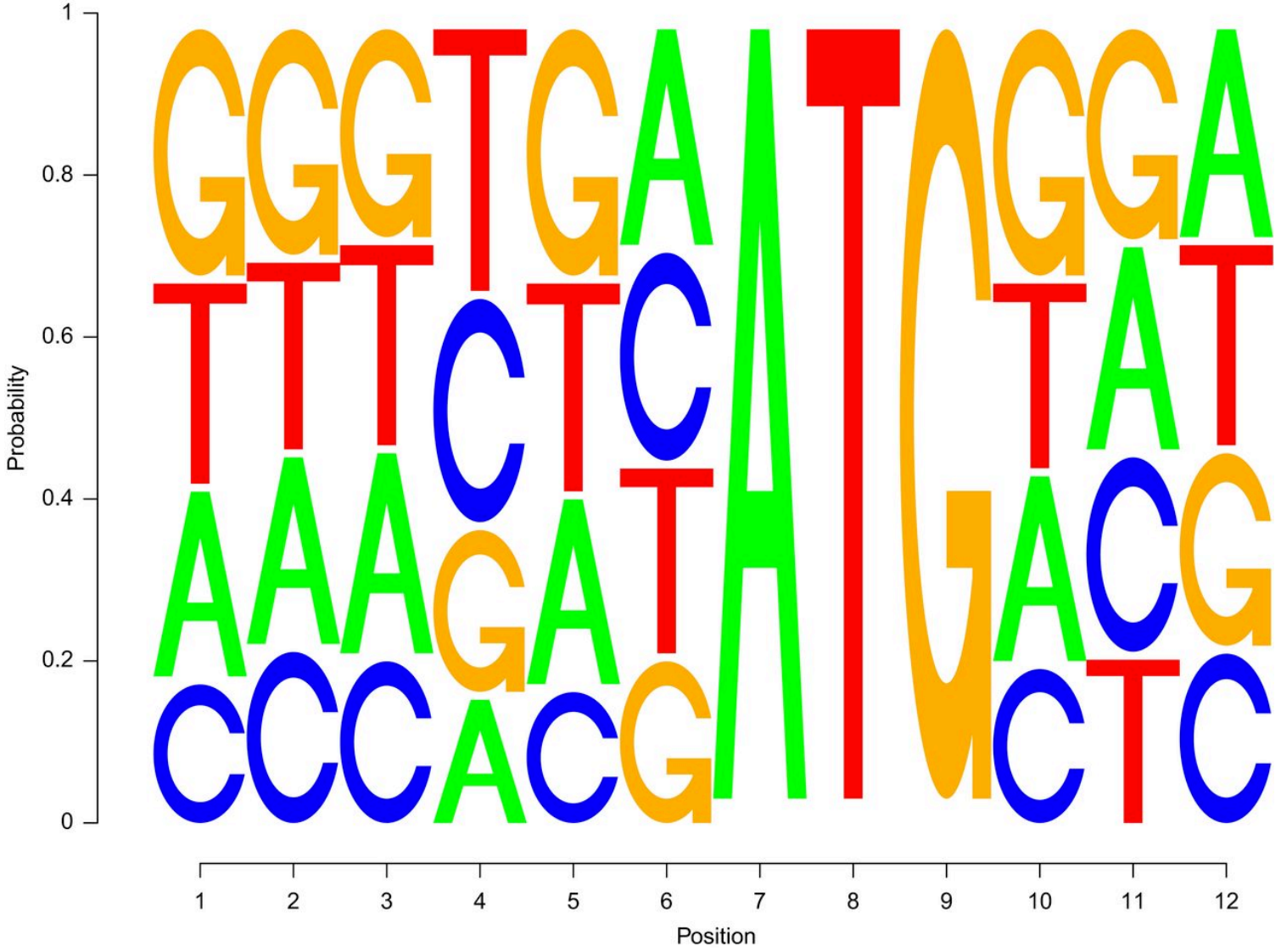


Figure 7

5d. Correlation between log(TE) and gene-level features

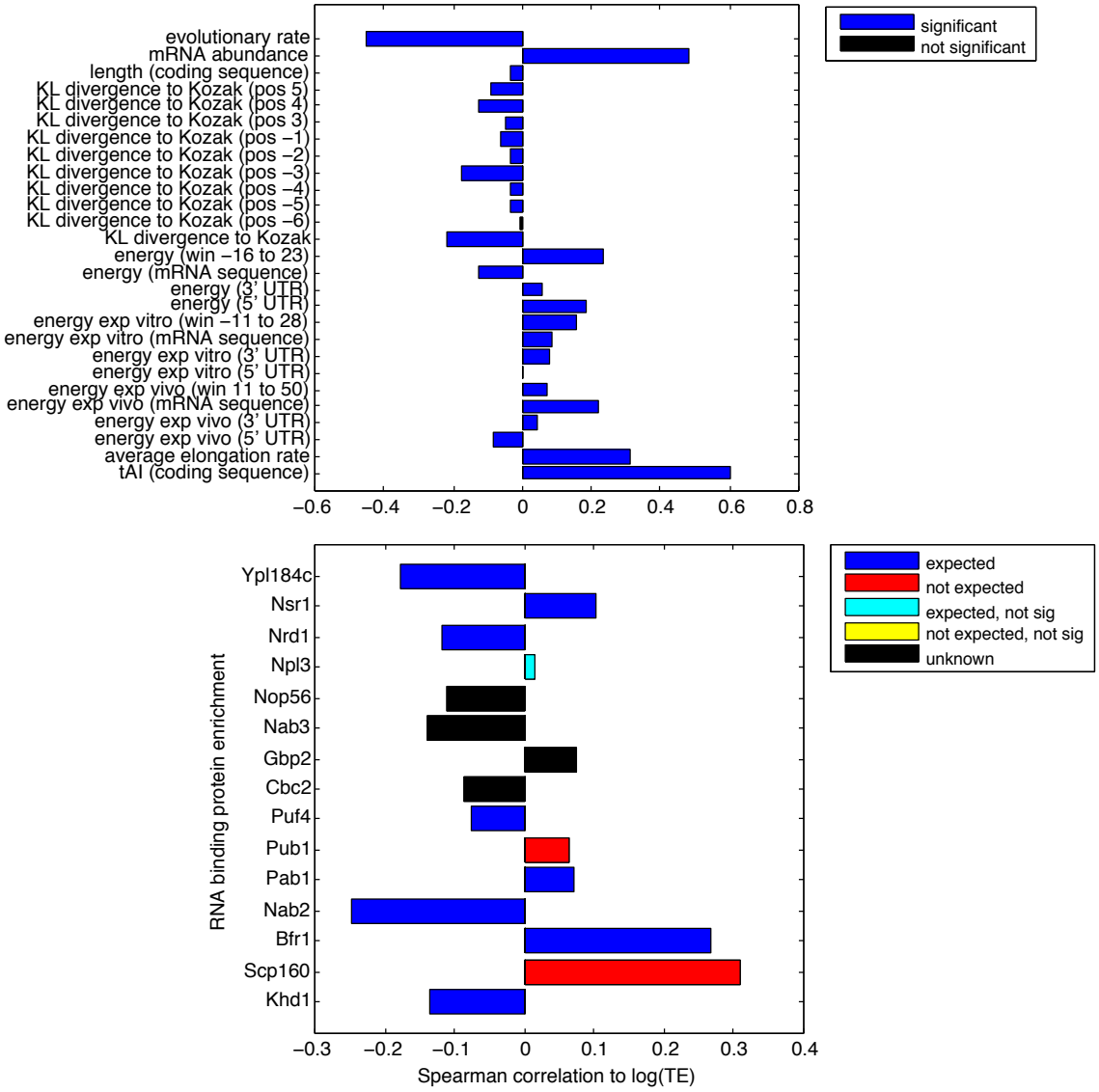


Figure S7