

# Mean of the Typical Decoding Rates: A New Translation Efficiency Index Based on the Analysis of Ribosome Profiling Data

Alexandra Dana\* and Tamir Tuller\*<sup>1,1</sup>

\*Department of Biomedical Engineering, Faculty of Engineering and <sup>†</sup>Segol School, Tel Aviv University, Ramat Aviv 69978, Israel

**ABSTRACT** Gene translation modeling and prediction is a fundamental problem that has numerous biomedical implementations. In this work we present a novel, user-friendly tool/index for calculating the mean of the typical decoding rates that enables predicting translation elongation efficiency of protein coding genes for different tissue types, developmental stages, and experimental conditions. The suggested translation efficiency index is based on the analysis of the organism's ribosome profiling data. This index could be used for example to predict changes in translation elongation efficiency of lowly expressed genes that usually have relatively low and/or biased ribosomal densities and protein levels measurements, or can be used for example for predicting translation efficiency of new genetically engineered genes. We demonstrate the usability of this index via the analysis of six organisms in different tissues and developmental stages. Distributable cross platform application and guideline are available for download at: [http://www.cs.tau.ac.il/~tamirtul/MTDR/MTDR\\_Install.html](http://www.cs.tau.ac.il/~tamirtul/MTDR/MTDR_Install.html)

## KEYWORDS

codon decoding rate  
codon usage bias  
gene translation  
ribosome profiling  
transcript evolution

Gene translation is a fundamental intracellular process. Thus, the ability to predict gene translation elongation efficiency (*i.e.*, a gene's translation rate) is a central challenge related to all biomedical disciplines.

Currently, there are no direct measures of genes translation efficiency that do not include “components” of other gene expression stages, such as transcription and/or posttranslational regulatory steps (*e.g.*, mRNA degradation, protein degradation, and protein synthesis rate). For example, conventional nondirect proxies of translation efficiency include messenger RNA (mRNA) levels, protein abundance, or the normalization of the two aforementioned factors. One drawback of these proxies is the fact that they are not available for the majority of organisms in various experimental conditions and tissues. In addition, these proxies are not highly reliable for lowly expressed genes, nor they can predict the translation efficiency of new engineered genes expressed in the same cell conditions.

Recently, ribosome profiling was suggested for measuring some aspects of the translation process at nucleotide resolution (Ribo-seq) (Ingolia *et al.* 2009). This method can potentially estimate the relative time ribosomes spend on the organismal mRNA molecules, at nucleotide resolution. Thus, ribosome profiles reflect *in vivo* the translation process of specific tissues and developmental stages or conditions. As a result, it was suggested to estimate the general translation efficiency of genes by calculating their mean average footprint read counts (Ingolia *et al.* 2009).

However, resulting ribosome profiles are reliable only for highly expressed genes, thus restricting the ability of the method to accurately measure translation efficiency of the remaining of the genes or to predict translation efficiency of newly engineered genes in similar cellular conditions. For example, as can be seen in Figure 2, for *S. cerevisiae*, *H. sapiens*, and *M. musculus* only 13.9–23.7% of the genes include more than 50% positions with nonzero mapped read counts; similarly, only 8.5–11.8% of their genes include mean footprint count (FC; per nucleotide) larger than 2.

Additional conventional approach/indexes for estimating translation efficiency are based on various measures of codon distribution/bias within the opening reading frame (ORF) (Sharp and Li 1987; Wright 1990; dos Reis *et al.* 2004; Fox and Erill 2010; Sabi and Tuller 2014). These indexes were found to be correlative with the protein abundance in the cell for *S. cerevisiae*, *E. coli*, and *C. elegans* (dos Reis *et al.* 2004; Tuller *et al.* 2010b; Sabi and Tuller 2014). However, these

Copyright © 2015 Dana and Tuller

doi: 10.1534/g3.114.015099

Manuscript received October 21, 2014; accepted for publication November 13, 2014; published Early Online December 1, 2014.

This is an open-access article distributed under the terms of the Creative Commons Attribution Unported License (<http://creativecommons.org/licenses/by/3.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

<sup>1</sup>Corresponding author: Tel Aviv University, Ramat Aviv 69978, Israel.

E-mail: tamirtul@post.tau.ac.il

indexes are not condition nor tissue specific and may not be directly related to translation but to other steps of gene expression and gene evolution (Sharp and Li 1987; Plotkin and Kudla 2011).

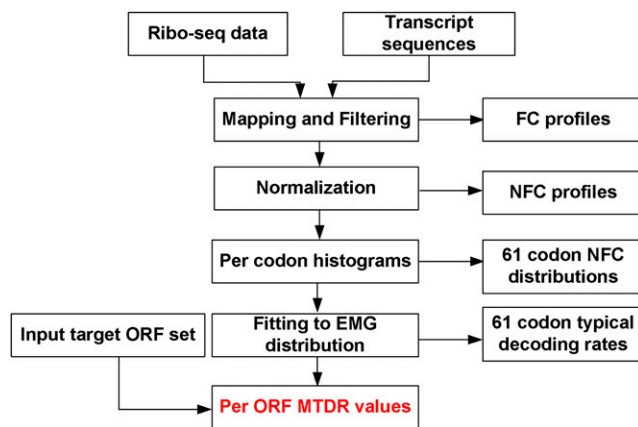
In contrast to the previous suggested indexes, the mean of the typical decoding rates (MTDR) index (Dana and Tuller 2014) is based on the estimation of the typical codon decoding times from Ribo-seq data, thus potentially capturing aspects of translation elongation in specific tissues, developmental stages, and/or conditions. Specifically, the MTDR index calculates the geometrical mean of the estimated typical nominal translation rates of a gene's codons after filtering biases and phenomena such as ribosomal traffic jams and translational pauses (Dana and Tuller 2014) (see also Figure 1 and the section *Materials and Methods*), reflecting the mean typical translation elongation rate of a gene. Thus, this index could be used to predict all genes' translation efficiency, including newly engineered genes.

In this study we estimate the typical codons translation time of additional organisms in different conditions and tissues (see Table 1) from ribosomal profiling data (see the section *Materials and Methods*). We also demonstrate the advantages of the MTDR index in predicting various gene expression measurements and compare it with previously suggested translation efficiency indexes (which are based on codon distributions). Finally, we provide a cross-platform tool for calculating the MTDR index of all ORFs in these organisms and conditions.

## MATERIALS AND METHODS

### Calculating the normalized footprint count (NFC) distribution

As seen in Figure 2, the majority of genes' ribosome profiles have less than 50% of codons mapped with read counts. Therefore, to avoid analyzing unreliable ribosome profiles that could bias estimations,



**Figure 1** Diagram showing the mean of the typical decoding rates (MTDR) calculation process. Ribo-seq data are created per selected organism (condition and tissue specific). Resulting mRNA fragments are then mapped to transcript sequences resulting for each gene its ribosomal profile. Genes with mean footprint counts (FC) lower than one are filtered. To enable comparison between footprint counts of different genes with different mRNA levels and initiation rates, FC of each gene are normalized, resulting in normalized footprint count (NFC) profiles. Then for each codon type a NFC distribution is created. By fitting the NFC distribution to an exponentially modified Gaussian distribution type, the typical decoding rate of each codon is estimated. Finally, for each one of the input target open reading frames (ORFs), the mean typical decoding rates of its codons (MTDR index) is calculated.

**Table 1** Analyzed organisms and tissues/conditions

Organism	Condition/Tissue
<i>E. coli</i> (Li <i>et al.</i> 2012)	–
<i>B. subtilis</i> (Li <i>et al.</i> 2012)	–
<i>S. cerevisiae</i>	Exponential (Ingolia <i>et al.</i> 2009), (Brar <i>et al.</i> 2012) DNA replication (Brar <i>et al.</i> 2012) Recombination (Brar <i>et al.</i> 2012) Metaphase II (Brar <i>et al.</i> 2012) Anaphase (Brar <i>et al.</i> 2012) Spore packing (Brar <i>et al.</i> 2012) Spores (Brar <i>et al.</i> 2012)
<i>C. elegans</i> (Stadler <i>et al.</i> 2012)	L4 (Stadler <i>et al.</i> 2012)
	L2 L1
<i>M. musculus</i>	Embryonic stem cells (Ingolia <i>et al.</i> 2011) Neutrophils (Guo <i>et al.</i> 2010) Embryonic fibroblast (Lee <i>et al.</i> 2012)
<i>H. sapiens</i>	HEK293 (Lee <i>et al.</i> 2012)

only genes with a median FC greater than one were included in the analysis (Dana and Tuller 2014). In addition, previous studies indicated an increase of FC at the beginning of the ORF (Ingolia 2010; Ingolia *et al.* 2011) and for some organisms at the end of ORF (Li *et al.* 2012); therefore, the first and last 20 codons were excluded from the analysis. Moreover, to prevent analysis of unreliable reads, codons with FC values less than one were excluded from the analysis (Li *et al.* 2012).

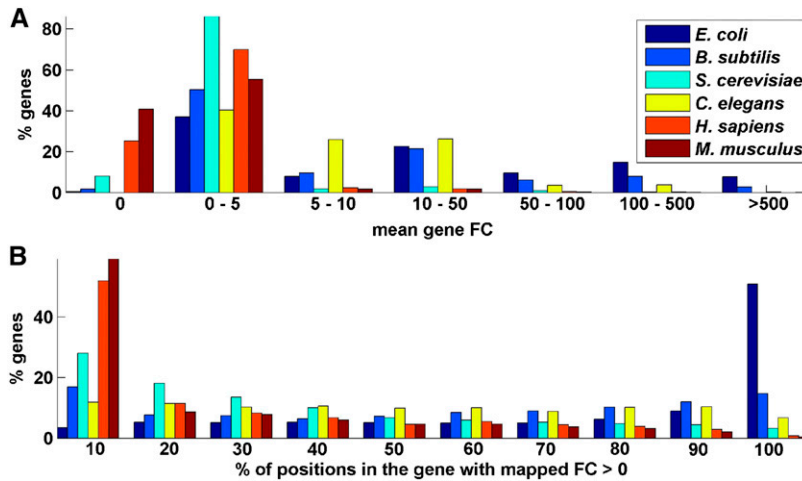
To enable comparison of footprint counts of a codon type from genes with different mRNA levels and initiation rates, FC of each codon were first normalized by the average FC of each gene (Li *et al.* 2012; Qian *et al.* 2012; Dana and Tuller 2014), resulting in NFC. This normalization enables measuring the relative time a ribosome spends translating each codon in a specific gene relative to other codons in it, while considering the total number of codons in the gene. Then, for each codon type a vector consisting of NFC values originating from all analyzed genes was generated, creating the “NFC distribution” of a codon.

### Estimating the codons' typical decoding time

Based on the characteristics of the NFC distributions, we suggest that their topology could result from a superposition of two distributions/components (Dana and Tuller 2014): the first one describes the “typical” decoding time of the ribosomes, which was modeled by a normal distribution characterized by its mean  $\mu$  and variance  $\sigma^2$ .

The second component describes relatively rare translational pauses and ribosomal interactions such as traffic jams due to the codons' different translation efficiency and was modeled by a random variable with an exponentially distribution, characterized by one parameter  $\lambda$ .

The summation of two independent normal and exponential random variables corresponding to the distributions mentioned above results in an exponentially modified Gaussian distribution. The parameters  $\mu, \sigma, \lambda$  were estimated by fitting the measured NFC distributions to the exponentially modified Gaussian distribution, under the log-likelihood criterion. The  $\mu$  parameter is referred as the typical decoding time of a codon. For more details, see Dana and Tuller (2014). Then, the MTDR index of a gene was defined as the geometrical mean of its codons translation rates.



**Figure 2** Footprint count (FC) statistics over all genes of six organisms. (A) Histogram of average read counts. (B) Histogram of percentage of positions with a positive number of mapped read counts. As can be seen in all analyzed organisms, most of the genes have very low read counts.

### Protein abundance and mRNA levels measurements

*E. coli* protein abundance measurements were downloaded from the PRIDE database (Vizcaino *et al.* 2013), and mRNA level measurements were taken from another source (Taniguchi *et al.* 2010). *C. elegans* mRNA levels were downloaded from Kirienko and Fay (2007). *B. subtilis* protein abundance measures were used as published by Chi *et al.* (2011), and mRNA levels were downloaded from another source (Nicolas *et al.* 2012). *S. cerevisiae* protein abundance measures were averaged from four quantitative large-scale measurements: two large scale measurements in two conditions (Newman *et al.* 2006), and a large-scale protein abundance measurement from two sources (Ghaemmaghami *et al.* 2003; Lee *et al.* 2011). mRNA levels were determined by averaging large scale measurements of mRNA levels (Wang *et al.* 2002; Ingolia *et al.* 2009).

### Estimating the contribution of the MTDR index and other translation efficiency indexes to predicting protein abundance

To estimate whether the MTDR index improves the prediction protein levels relative to other translation efficiency indexes, we have modeled the relationship between protein abundance and the different translation efficiency indexes MTDR, transfer RNA (tRNA) adaptation index [tAI (dos Reis and Wernisch 2009)], and codon adaptation index [CAI (Sharp and Li 1987)], using a linear regressor (Seber and Lee 2012).

$$\tilde{P}A = c + w_1MTDR + w_2tAI + w_3CAI$$

where the coefficients  $c, w_1, w_2, w_3$  were estimated to minimize the mean square error difference between the prediction vector  $\tilde{P}A$  and the real protein abundance measurements  $PA$ .

For each one of the coefficients  $c, w_1, w_2, w_3$  confidence intervals (Kendall and Stuart 1979) were calculated to determine the reliability of the estimates (at 95% confidence interval). A coefficient's confidence interval that does not contain zeroes implies that it significantly

contributes to the regression, *i.e.*, the coefficient is not zero (Kendall and Stuart 1979).

## RESULTS

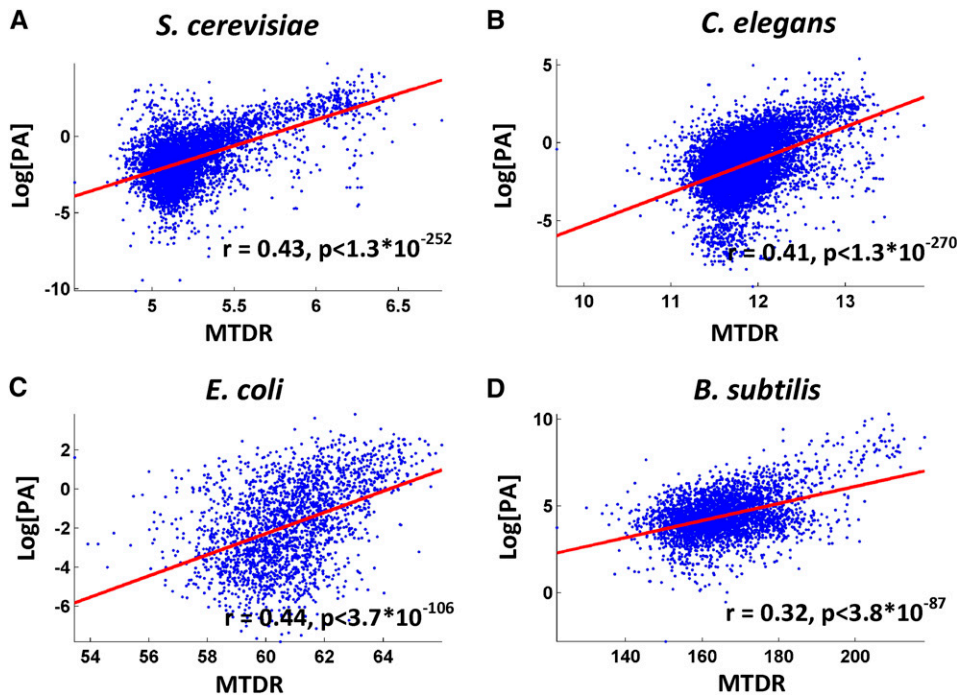
### The estimated codons decoding times correlate with measures of codon usage bias

We started our analysis by comparing the typical decoding times estimated from the ribosome profiling data ( $\mu$ ) to other estimations of codon translation efficiency estimated using additional methods which are based on codon distributions. These include: 1) decoding time based on the codon adaptation index ( $DT_{CAI}$ ), which calculates the codon bias in highly expressed genes (Sharp and Li 1987) and 2) decoding time based on the tRNA adaptation index ( $DT_{tAI}$ ) (dos Reis *et al.* 2004), which takes into consideration the number of tRNA copies in the genome recognizing each codon and additional codon/anticodon interactions. As seen in Table 2, the correlation between  $\mu$  values and the other codon decoding time estimators is significant in all analyzed organisms ( $0.42 < P < 0.83; P < 0.00065$ ).

In addition, we also compared the  $\mu/DT_{CAI}/DT_{tAI}$  values to decoding times measured by using various experimental methods. For example, the amino acid insertion time per codon in *E. coli* was estimated using known biochemistry factors (Fluitt *et al.* 2007). Spearman correlation between insertion time and  $\mu$  was 0.43 ( $P = 0.00051$ ), whereas the correlation between insertion time and  $DT_{tAI}/DT_{CAI}$  was lower: 0.35/0.31 ( $P = 0.0062/0.017$ ), supporting the conjecture that  $\mu$  values are better direct estimators of the amino acid insertion time. In another study (Chu *et al.* 2014) the authors found that for *S. cerevisiae* the GAG codon is more slowly translated than the GAA codon. This result is also supported by the  $\mu$  estimations (GAG: 0.24 vs. GAA: 0.21). The same trend was also observed for the  $DT_{CAI}$  values (GAG: 0.016 vs. GAA: 0.006) and  $DT_{tAI}$  values (GAG: 2.44 vs. GAA: 1.1). In another study (Kemp *et al.* 2013) it was shown that in *S. cerevisiae*, replacing codons the codons "CAA" by the codons "CAG," which are decoded by a rare tRNA, near the 5' end of an ORF reduces luciferase

**Table 2** Spearman correlation between  $\mu$  and various methods for estimating the typical codon decoding time

Estimation method	<i>E. coli</i>	<i>B. subtilis</i>	<i>S. cerevisiae</i>	<i>C. elegans</i>
$DT_{CAI}$	$r = 0.55$ $p = 4 \cdot 10^{-6}$	$r = 0.66$ $p = 8.8 \cdot 10^{-9}$	$r = 0.56$ $p = 3.2 \cdot 10^{-6}$	$r = 0.67$ $p = 4.3 \cdot 10^{-9}$
$DT_{tAI}$	$r = 0.42$ $p = 0.00065$	$r = 0.51$ $p = 3.1 \cdot 10^{-115}$	$r = 0.47$ $p = 0.00013$	$r = 0.83$ $p = 1.6 \cdot 10^{-16}$



**Figure 3** Protein abundance (PA) vs. mean of the typical decoding rates (MTDR) and Spearman correlation between MTDR index and protein abundance for various organisms.

expression by 60%. Indeed the estimated  $\mu$  decoding time of the CAG codon was found to be greater than of the CAA codon (0.25 vs. 0.15). An additional work (Letzring *et al.* 2010) studied the effect of synonymous codon repeats of luciferase expression in *S. cerevisiae*; to validate our  $\mu$  estimations, we calculated Spearman correlation between translation efficiency changes with respect to the wild type and the estimated  $\mu$  values and found them to be significantly correlative ( $r = -0.33, P = 0.013$ ).

### The MTDR index predicts protein levels in standard conditions with comparable quality as other codon bias measures

The usability of the suggested MTDR index was validated for four different organisms that have abundant large scale protein and mRNA levels measurements (*E. coli*, *B. subtilis*, *C. elegans*, and *S. cerevisiae*). Spearman correlation between the MTDR index to various proxies of translation efficiency such as protein abundance resulted in significant correlations ( $0.32 < r < 0.44; P < 3.8 \times 10^{-87}$ ; see also Figure 3 and Table 3). Similarly, a significant partial Spearman correlation between protein abundance and MTDR index given mRNA

levels (partial correlation) was observed ( $0.23 < r < 0.5; P < 4.2 \times 10^{-86}$ ; see Table 3).

Measures based on codon usage bias resulted in similar correlations with protein levels in standard conditions (Table 3); these results can be explained among others, by the fact that codon bias usage measures various (direct and indirect) aspects of gene expression (Chamary *et al.* 2006; Hershberg and Petrov 2008; Plotkin and Kudla 2011; Sauna and Kimchi-Sarfaty 2011) (Tuller and Zur, unpublished data), by the fact that ribosome profiling data are probably “noisier” and more biased than genomic sequence data (Dana and Tuller 2012; Gerashchenko and Gladyshev 2014), and by the fact that in this study we consider protein levels in “standard”/exponential conditions which are probably reflected more properly by codon bias usage.

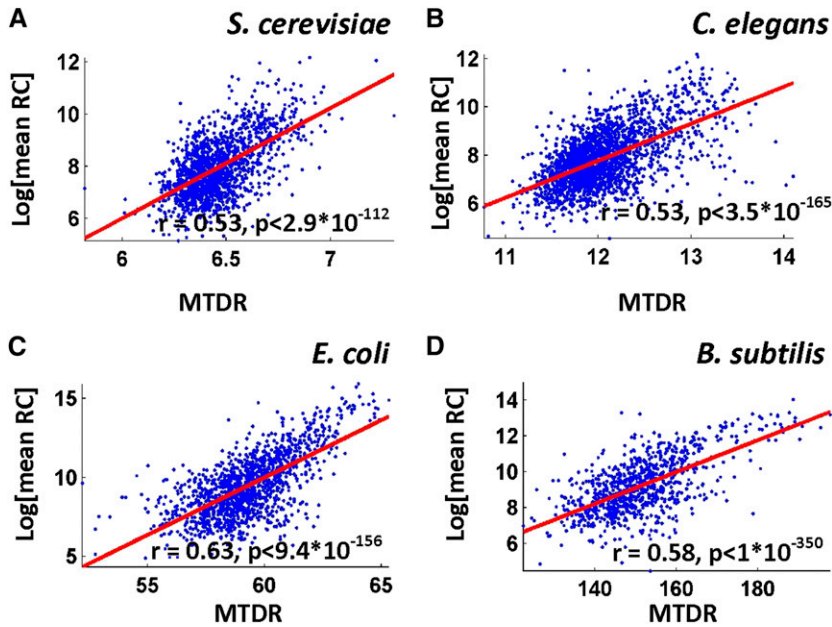
For some of the analyzed organisms, a decrease in correlation between the different translation efficiency indexes and protein abundance was observed when controlling for mRNA levels. As genes with greater mRNA levels potentially consume a greater percentage of the ribosomes in the cell, they are expected to undergo stronger selection for greater translation elongation speed (and/or other aspects of translation efficiency) to reduce ribosome utilization (Tuller *et al.*

**Table 3** Spearman correlation between protein abundance (also when controlling for mRNA levels) and various translation efficiency indexes

Organism	Correlation Between Index and PA			Correlation Between Index and PA Given mRNA Levels		
	MTDR	CAI	tAI	MTDR	CAI	tAI
<i>E. coli</i>	$r = 0.44$ $p < 3.7 \times 10^{-106}$	$r = 0.56$ $p = 3.8 \times 10^{-184}$	$r = 0.55$ $p = 1.2 \times 10^{-178}$	$r = 0.5$ $p < 1.1 \times 10^{-38}$	$r = 0.61$ $p = 3.4 \times 10^{-62}$	$r = 0.48$ $p = 3.4 \times 10^{-35}$
<i>B. subtilis</i>	$r = 0.32$ $p < 3.8 \times 10^{-87}$	$r = 0.37$ $p = 1.5 \times 10^{-117}$	$r = 0.41$ $p = 4.7 \times 10^{-146}$	$r = 0.26$ $p < 3.7 \times 10^{-43}$	$r = 0.280$ $p = 7.1 \times 10^{-67}$	$r = 0.32$ $p = 8.6 \times 10^{-108}$
<i>C. elegans</i>	$r = 0.41$ $p < 1 \times 10^{-270}$	$r = 0.52$ $p = 0$	$r = 0.56$ $p = 0$	$r = 0.23$ $p < 4.2 \times 10^{-86}$	$r = 0.35$ $p = 3.7 \times 10^{-206}$	$r = 0.41$ $p = 6.1 \times 10^{-293}$
<i>S. cerevisiae</i>	$r = 0.43$ $p < 1.3 \times 10^{-252}$	$r = 0.44$ $p = 1.1 \times 10^{-275}$	$r = 0.47$ $p = 2.6 \times 10^{-314}$	$r = 0.3$ $p < 1.2 \times 10^{-102}$	$r = 0.33$ $p = 5.3 \times 10^{-125}$	$r = 0.37$ $p = 1.1 \times 10^{-66}$

PA, protein abundance; MDTR, mean of the typical decoding rates; CAI, codon adaptation index; tAI, tRNA adaptation index.





**Figure 4** Ribosomal load, as estimated by the mean footprint counts per gene vs. mean of the typical decoding rates (MTDR) index for various organisms and Spearman correlation between these factors.

2010a; Plotkin and Kudla 2011). Therefore we expect to see a positive correlation between measures of translation efficiency and mRNA levels (and not only with protein abundance); as a result, the partial correlation between protein levels and translation efficiency indexes decreases when controlling for mRNA levels.

Because the correlation between the MTDR index and protein abundance was found to be similar to the correlation between protein abundance and the other two translation efficiency indexes, we wanted to assess whether the MTDR index provides *additional* information with respect to the other two. To this end, we calculated a linear regressor based on tAI, CAI, and MTDR for predicting protein abundance levels (see the section *Materials and Methods*). Then, we checked the 95% confidence intervals of each one of the translation efficiency indexes and found that they do not include the value 0 for

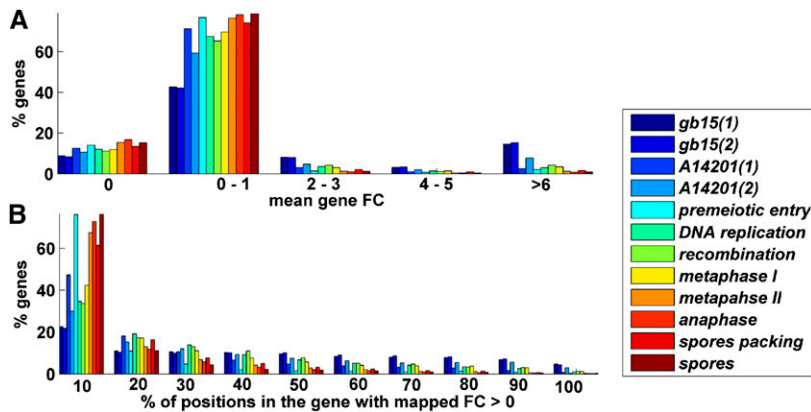
the variable MTDR (see the section *Materials and Methods*), demonstrating that the MTDR index contributes additional information to protein abundance prediction given the other two indexes.

Finally, it should be noted that in endogenous genes it is impossible to prove the direction of causality between the various translation efficiency indexes and protein abundance based on correlation; a correlation may suggest that 1) codons with greater  $\mu$  values or codons that are recognized by more abundant tRNA molecule tend to improve translation rate and thus increase protein levels (e.g., Letzring *et al.* 2010; Tuller *et al.* 2010b); 2) genes with greater expression levels are selected to have codons with a greater MTDR/tAI index due to reasons not directly related, increasing the number of proteins per mRNA [for example, global ribosomal allocation (Kudla *et al.* 2009)].

**Table 4** Correlation between MTDR index and ribosomal load for the different tissues/conditions of the analyzed organisms

Organism	Correlation Between MTDR Index and Ribosomal Load	Correlation Between tAI Index and Ribosomal Load	Correlation between CAI Index and Ribosomal Load
<i>E. coli</i> (Li <i>et al.</i> 2012)	$r = 0.63$ $p = 9.4 \times 10^{-156}$	$r = 0.56$ $p = 1.3 \times 10^{-117}$	$r = 0.60$ $p = 5.7 \times 10^{-142}$
<i>B. subtilis</i> (Li <i>et al.</i> 2012)	$r = 0.58$ $p = 0$	$r = 0.33$ $p = 0$	$r = 0.56$ $p = 0$
<i>S. cerevisiae</i> – Exponential (Ingolia <i>et al.</i> 2009),	$r = 0.53$ $p = 2.9 \times 10^{-112}$	$r = 0.62$ $p = 5.5 \times 10^{-163}$	$r = 0.65$ $p = 3 \times 10^{-184}$
<i>S. cerevisiae</i> – Exponential (gb15) (Brar <i>et al.</i> 2012)	$r = 0.32$ $p = 1 \times 10^{-22}$	$r = 0.29$ $p = 2.3 \times 10^{-18}$	$r = -0.32$ $p = 1.3 \times 10^{-22}$
<i>S. cerevisiae</i> – Exponential (A14201) (Brar <i>et al.</i> 2012)	$r = 0.80$ $p = 1.4 \times 10^{-131}$	$r = 0.77$ $p = 7.9 \times 10^{-116}$	$r = 0.80$ $p = 1.1 \times 10^{-134}$
<i>S. cerevisiae</i> – DNA replication (Brar <i>et al.</i> 2012)	$r = 0.26$ $p = 1.9 \times 10^{-8}$	$r = 0.11$ $p = 0.023$	$r = 0.09$ $p = 0.048$
<i>S. cerevisiae</i> – Recombination (Brar <i>et al.</i> 2012)	$r = 0.31$ $p = 2.4 \times 10^{-12}$	$r = 0.20$ $p = 7.6 \times 10^{-6}$	$r = 0.20$ $p = 1.4 \times 10^{-5}$
<i>S. cerevisiae</i> – Metaphase II (Brar <i>et al.</i> 2012)	$r = 0.46$ $p = 4.3 \times 10^{-8}$	$r = -0.12$ $p = 0.15$	$r = -0.12$ $p = 0.18$
<i>S. cerevisiae</i> – Anaphase (Brar <i>et al.</i> 2012)	$r = 0.23$ $p = 0.027$	$r = 0.10$ $p = 0.31$	$r = 0.11$ $p = 0.26$
<i>S. cerevisiae</i> – Spore packing (Brar <i>et al.</i> 2012)	$r = 0.50$ $p = 0$	$r = -0.15$ $p = 0.045$	$r = -0.16$ $p = 0.029$
<i>S. cerevisiae</i> – Spores (Brar <i>et al.</i> 2012)	$r = 0.35$ $p = 0.00015$	$r = -0.11$ $p = 0.25$	$r = -0.08$ $p = 0.39$
<i>C. elegans</i> – L4 (Stadler <i>et al.</i> 2012)	$r = 0.53$ $p = 3.5 \times 10^{-165}$	$r = 0.45$ $p = 1.8 \times 10^{-112}$	$r = 0.49$ $p = 1.9 \times 10^{-132}$
<i>C. elegans</i> – L2 (Stadler <i>et al.</i> 2012)	$r = 0.48$ $p = 1.6 \times 10^{-197}$	$r = 0.47$ $p = 1.4 \times 10^{-185}$	$r = 0.51$ $p = 1.4 \times 10^{-219}$
<i>C. elegans</i> – L1 (Stadler <i>et al.</i> 2012)	$r = 0.51$ $p = 5.9 \times 10^{-10}$	$r = 0.42$ $p = 2.8 \times 10^{-72}$	$r = 0.46$ $p = 9.3 \times 10^{-86}$
<i>M. musculus</i> – Embryonic stem cells (Ingolia <i>et al.</i> 2011)	$r = 0.02$ $p = 0.35$	$r = 0.07$ $p = 0.0095$	$r = 0.07$ $p = 0.0046$
<i>M. musculus</i> – Neutrophils (Guo <i>et al.</i> 2010)	$r = 0.22$ $p = 1.2 \times 10^{-13}$	$r = 0.07$ $p = 0.012$	$r = 0.12$ $p = 5.4 \times 10^{-5}$
<i>M. musculus</i> – Embryonic fibroblast (Lee <i>et al.</i> 2012)	$r = 0.41$ $p = 4.1 \times 10^{-15}$	$r = 0.22$ $p = 3.6 \times 10^{-5}$	$r = 0.22$ $p = 3.4 \times 10^{-5}$
<i>H. sapiens</i> – HEK293 (Lee <i>et al.</i> 2012)	$r = 0.17$ $p = 1.9 \times 10^{-9}$	$r = 0.15$ $p = 9.2 \times 10^{-8}$	$r = 0.15$ $p = 3.4 \times 10^{-8}$

MDTR, mean of the typical decoding rates; tAI, tRNA adaptation index; CAI, codon adaptation index.



**Figure 5** Footprint count (FC) statistics for *S. cerevisiae* different meiosis stages. (A) Histogram of average read counts. (B) Histogram of percentage of positions with a positive number of mapped read counts. As can be seen in all of the organisms, most of the genes have very low read counts.

### Using the MTDR index for predicting ribosomal load

To demonstrate that the MTDR index can predict ribosomal densities of each one of the analyzed organisms we used 50% of the highly expressed genes to compute the typical codon decoding times. Next, we computed the MTDR index of the genes in the rest 50% of the highly expressed genes (that were not used for computing the typical decoding times). Finally, we computed the correlation between the MTDR index and the actual mean read count of these genes. As can be seen in Figure 4 and in Table 4, in almost all analyzed organisms/conditions the correlations are positive and significant (the top correlation is 0.8,  $P = 1.4 \times 10^{-131}$ ). This result demonstrates that the MTDR index could be used as a good predictor of the ribosome load (number of ribosomes per mRNA \* number of mRNA molecules), probably since highly translated genes tend to undergo selection for greater codon elongation rate, for example, to improve ribosomal allocation and translation cost (Kudla *et al.* 2009; Tuller *et al.* 2010a). For comparison, we also calculated the correlation between the tAI and CAI indexes and ribosomal load (see Table 4). These indexes resulted in similar correlations for the different “typical”/exponential stage conditions, however, produced a lower correlation in atypical conditions such as the different *S. cerevisiae* meiosis stages and for some *M. musculus* tissue types.

It should be mentioned that a decrease in this correlation was observed for non-exponential stages in *S. cerevisiae* and in greater eukaryotes such as *H. sapiens* and *M. musculus*. This could result from: 1) a greater level of noise (see Figure 5) and biases caused by additional and/or more complicated biological mechanisms (*e.g.*, due substantial alternative splicing in mammals the mapping of reads to exons is less trivial) (Engstrom *et al.* 2013); 2) the fact that mammals have smaller effective population size and thus lower selection pressure related to some translation aspects (Charlesworth 2009; dos Reis and Wernisch 2009); and 3) since growth rate is strongly related to fitness in unicellular organisms but not in mammals, there is lower effect on ribosomal allocation on the organisms fitness in mammals than in the rest of the analyzed organism (Rocha 2004; dos Reis and Wernisch 2009).

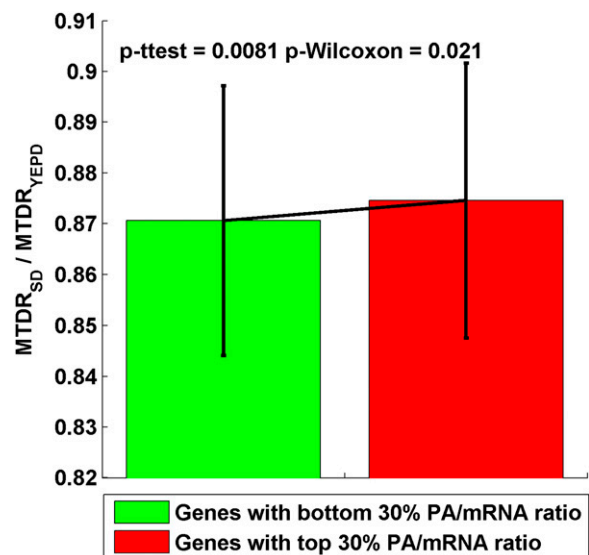
### Assessing the ability of the MTDR index to predict translation efficiency in different experimental conditions

Codon bias indexes (*e.g.*, CAI, tAI) are based solely on static information encoded in the genome or general chemical properties; thus, they cannot differentiate among different experimental conditions. One of the major advantages of the MTDR index is that it is condition specific. To demonstrate this advantage, we estimated the typical

decoding times of *S. cerevisiae* in starvation conditions (Ingolia *et al.* 2009). Translation efficiency (defined as protein abundance normalized by mRNA levels) was calculated for rich (yeast extract peptone dextrose) and minimal (synthetic defined) media [protein abundance and mRNA levels measured per cell were taken from a previous study (Newman *et al.* 2006)].

We found that the MTDR values of the genes with top/bottom 30% translation efficiency ratios are significantly different (*t*-test:  $P = 0.0081$ ; Wilcoxon test:  $P = 0.021$ ): genes with a greater change in their translation efficiency have a greater change in their estimated MTDR index (in the same direction). This result demonstrates the ability of the suggested index to estimate gene expression in different conditions (Figure 6).

In addition, we also found that Spearman correlation between the MTDR ratio and ribosomal load ratio for *different tissue/developmental conditions* in *C. elegans*, *S. cerevisiae*, and *M. musculus* (Table 5) was significantly positive, ( $0.17 < r < 0.72$ ;  $P < 0.0003$ ), indicating that the MTDR index could predict changes in ribosomal load, contrary to the static tAI and CAI indexes.



**Figure 6** The ration between mean of the typical decoding rates (MTDR) index calculated in starvation (synthetic defined [SD]) and rich medium (yeast extract peptone dextrose [YEPD]) for genes with bottom/up 30% PA/mRNA ratio in SD/YEPD conditions (green/red).

■ **Table 5 Spearman correlation between MTDR index ratios and ribosomal load ratio for different tissues/developmental stages in *C. elegans*, *S. cerevisiae* (partial list) and *M. musculus***

Compared Conditions	Corr(MTDR1/MTDR2, meanRC1/meanRC2)	Compared Conditions	Corr(MTDR1/MTDR2, meanRC1/meanRC2)
<i>C. elegans</i> L4 vs. L2	$r = 0.53$ $P = 0$	<i>S. cerevisiae</i> premeiotic entry vs. metaphase I	$r = 0.40$ $P = 0.00045$
<i>C. elegans</i> L4 vs. L1	$r = 0.51$ $P = 4.5 \times 10^{-52}$	<i>S. cerevisiae</i> premeiotic entry vs. spore packing	$r = 0.33$ $P = 0.044$
<i>C. elegans</i> L2 vs. L1	$r = 0.51$ $P = 0$	<i>S. cerevisiae</i> premeiotic entry vs. spores	$r = 0.56$ $P = 0.00042$
<i>S. cerevisiae</i> gb15 vs. A14201	$r = 0.59$ $P = 5.8 \times 10^{-8}$	<i>S. cerevisiae</i> DNA replication vs. metaphase II	$r = 0.44$ $P = 0.014$
<i>S. cerevisiae</i> gb15 vs. premeiotic entry	$r = 0.42$ $P = 0.00011$	<i>S. cerevisiae</i> recombination vs. metaphase II	$r = 0.54$ $P = 3.7 \times 10^{-5}$
<i>S. cerevisiae</i> gb15 vs. recombination	$r = 0.32$ $P = 2 \times 10^{-6}$	<i>S. cerevisiae</i> metaphase I vs. metaphase II	$r = 0.45$ $P = 6.7 \times 10^{-5}$
<i>S. cerevisiae</i> gb15 vs. metaphase II	$r = 0.55$ $P = 5.2 \times 10^{-6}$	<i>S. cerevisiae</i> metaphase II vs. anaphase	$r = 0.45$ $P = 0.0031$
<i>S. cerevisiae</i> gb15 vs. spore packing	$r = 0.34$ $P = 0.00074$	<i>S. cerevisiae</i> metaphase II vs. spore packing	$r = 0.38$ $P = 0.055$
<i>S. cerevisiae</i> gb15 vs. spores	$r = 0.45$ $P = 0.0013$	<i>S. cerevisiae</i> spore packing vs. spores	$r = 0.52$ $P = 0.00079$
<i>S. cerevisiae</i> A14201 vs. premeiotic entry	$r = 0.72$ $P = 0.0031$	<i>M. musculus</i> embryonic stem cells vs. neutrophils	$r = 0.17$ $P = 0.0003$
<i>S. cerevisiae</i> A14201 vs. recombination	$r = 0.56$ $P = 6.5 \times 10^{-5}$	<i>M. musculus</i> embryonic stem cells vs. fibroblast	$r = 0.25$ $P = 0.00074$
<i>S. cerevisiae</i> A14201 vs. spore packing	$r = 0.56$ $P = 0.0096$	<i>M. musculus</i> neutrophils vs. fibroblast	$r = 0.31$ $P = 0.00023$

## APPLICATION

The MTDR application enables calculating the translation efficiency of various genes according to their ORF sequence. The input of the application includes the selection of one of the organisms and its tissue/condition and a file containing the ORFs of the requested genes (in FASTA format or text format where each ORF is defined in a separate line). ORFs could also be straightforwardly inserted in a textbox. The application returns an output file which includes for each of the ORFs their MTDR index. The available organisms and conditions/tissues are depicted in Table 1. The distributable cross platform application and user manual are available for download at: [http://www.cs.tau.ac.il/~tamirtul/MTDR/MTDR\\_Install.html](http://www.cs.tau.ac.il/~tamirtul/MTDR/MTDR_Install.html)

## ACKNOWLEDGMENTS

This research is partially supported by research grants from the German-Israeli Foundation (GIF) and Israel Cancer Research Fund (ICRF).

## LITERATURE CITED

- Brar, G. A., M. Yassour, N. Friedman, A. Regev, N. T. Ingolia *et al.*, 2012 High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science* 335, 552–557.
- Chamary, J. V., J. L. Parmley, and L. D. Hurst, 2006 Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nat. Rev. Genet.* 7: 98–108.
- Charlesworth, B., 2009 Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* 10: 195–205.
- Chi, B. K., K. Gronau, U. Mader, B. Hessling, D. Becher *et al.*, 2011 S-bacillithiolation protects against hypochlorite stress in *Bacillus subtilis* as revealed by transcriptomics and redox proteomics. *Mol. Cell Proteomics* 10: M111 009506.
- Chu, D., E. Kazana, N. Bellanger, T. Singh, M. F. Tuite *et al.*, 2014 Translation elongation can control translation initiation on eukaryotic mRNAs. *EMBO J.* 33: 21–34.
- Dana, A., and T. Tuller, 2012 Determinants of translation elongation speed and ribosomal profiling biases in mouse embryonic stem cells. *PLOS Comput. Biol.* 8: e1002755.
- Dana, A., and T. Tuller, 2014 The effect of tRNA levels on decoding times of mRNA codons. *Nucleic Acids Res.* 42: 9171–9181.
- dos Reis, M., R. Sava, and L. Wernisch, 2004 Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res.* 32: 5036–5044.
- dos Reis, M., and L. Wernisch, 2009 Estimating translational selection in eukaryotic genomes. *Mol. Biol. Evol.* 26: 451–461.
- Engstrom, P. G., T. Steijger, B. Sipos, G. R. Grant, A. Kahles *et al.*, 2013 Systematic evaluation of spliced alignment programs for RNA-seq data. *Nat. Methods* 10: 1185–1191.
- Fluitt, A., E. Pienaar, and H. Viljoen, 2007 Ribosome kinetics and aa-tRNA competition determine rate and fidelity of peptide synthesis. *Comput. Biol. Chem.* 31: 335–346.
- Fox, J. M., and I. Erill, 2010 Relative codon adaptation: a generic codon bias index for prediction of gene expression. *DNA Res.* 17: 185–196.
- Gerashchenko, M. V., and V. N. Gladyshev, 2015 Translation inhibitors cause abnormalities in ribosome profiling experiments. *Nucleic Acids Res.* 42: e134.
- Ghaemmaghami, S., W. K. Huh, K. Bower, R. W. Howson, A. Belle *et al.*, 2003 Global analysis of protein expression in yeast. *Nature* 425: 737–741.
- Guo, H., N. T. Ingolia, J. S. Weissman, and D. P. Bartel, 2010 Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* 466, 835–840.
- Hershberg, R., and D. A. Petrov, 2008 Selection on codon bias. *Annu. Rev. Genet.* 42: 287–299.
- Ingolia, N. T., 2010 Genome-wide translational profiling by ribosome footprinting. *Methods Enzymol.* 470: 119–142.
- Ingolia, N. T., S. Ghaemmaghami, J. R. Newman, and J. S. Weissman, 2009 Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324: 218–223.
- Ingolia, N. T., L. F. Lareau, and J. S. Weissman, 2011 Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147: 789–802.
- Kemp, A. J., R. Betney, L. Ciandrini, A. C. Schwenger, M. C. Romano *et al.*, 2013 A yeast tRNA mutant that causes pseudohyphal growth exhibits reduced rates of CAG codon translation. *Mol. Microbiol.* 87: 284–300.
- Kendall, M., and A. Stuart, 1979 *The Advanced Theory of Statistics. Vol. 2, Ed. 1. Inference and Relationship.* Griffin, London.
- Kirienko, N. V., and D. S. Fay, 2007 Transcriptome profiling of the *C. elegans* Rb ortholog reveals diverse developmental roles. *Dev. Biol.* 305: 674–684.
- Kudla, G., A. W. Murray, D. Tollervey, and J. B. Plotkin, 2009 Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* 324: 255–258.
- Lee, M. V., S. E. Topper, S. L. Hubler, J. Hose, C. D. Wenger *et al.*, 2011 A dynamic model of proteome changes reveals new roles for transcript alteration in yeast. *Mol. Syst. Biol.* 7: 514.
- Lee, S., B. Liu, S. X. Huang, B. Shen, S. B. Qian *et al.*, 2012 Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proceedings of the National Academy of Sciences* 109, E2424–E2432.
- Letzring, D. P., K. M. Dean, and E. J. Grayhack, 2010 Control of translation efficiency in yeast by codon-anticodon interactions. *RNA* 16: 2516–2528.

- Li, G. W., E. Oh, and J. S. Weissman, 2012 The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* 484: 538–541.
- Newman, J. R., S. Ghaemmaghami, J. Ihmels, D. K. Breslow, M. Noble *et al.*, 2006 Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* 441: 840–846.
- Nicolas, P., U. Mader, E. Dervyn, T. Rochat, A. Leduc *et al.*, 2012 Condition-dependent transcriptome reveals high-level regulatory architecture in *Bacillus subtilis*. *Science* 335: 1103–1106.
- Plotkin, J. B., and G. Kudla, 2011 Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.* 12: 32–42.
- Qian, W., J. R. Yang, N. M. Pearson, C. Maclean, and J. Zhang, 2012 Balanced codon usage optimizes eukaryotic translational efficiency. *PLoS Genet.* 8: e1002603.
- Rocha, E. P., 2004 Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. *Genome Res.* 14: 2279–2286.
- Sabi, R., and T. Tuller, 2014 Modelling the efficiency of codon-tRNA interactions based on codon usage bias. *DNA Res.* 21: 511–526.
- Sauna, Z. E., and C. Kimchi-Sarfaty, 2011 Understanding the contribution of synonymous mutations to human disease. *Nat. Rev. Genet.* 12: 683–691.
- Seber, G. A., and A. J. Lee, 2012 *Linear Regression Analysis*. John Wiley & Sons, New York.
- Sharp, P. M., and W. H. Li, 1987 The codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 15: 1281–1295.
- Stadler, M., K. Artiles, J. Pak, and A. Fire, 2012 Contributions of mRNA abundance, ribosome loading, and post- or peri-translational effects to temporal repression of *C. elegans* heterochronic miRNA targets, *Genome Res.* 22, 2418–2426. doi: 2410.1101/gr.136515.136111. Epub 132012 Aug 136511.
- Taniguchi, Y., P. J. Choi, G. W. Li, H. Chen, M. Babu *et al.*, 2010 Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* 329: 533–538.
- Tuller, T., A. Carmi, K. Vestsigian, S. Navon, Y. Dorfan *et al.*, 2010a An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* 141: 344–354.
- Tuller, T., Y. Y. Waldman, M. Kupiec, and E. Ruppin, 2010b Translation efficiency is determined by both codon bias and folding energy. *Proc. Natl. Acad. Sci. USA* 107: 3645–3650.
- Tuller, T., and H. Zur, 2014 Multiple Roles of the Coding Sequence 5' End in Gene Expression Regulation. To appear in *Nucleic Acids Res.*
- Vizcaino, J. A., R. G. Cote, A. Csordas, J. A. Dienes, A. Fabregat *et al.*, 2013 The PRoteomics IDentifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* 41: D1063–D1069.
- Wang, Y., C. L. Liu, J. D. Storey, R. J. Tibshirani, D. Herschlag *et al.*, 2002 Precision and functional specificity in mRNA decay. *Proc. Natl. Acad. Sci. USA* 99: 5860–5865.
- Wright, F., 1990 The 'effective number of codons' used in a gene. *Gene* 87: 23–29.

Communicating editor: A. M. Dudley