OPEN ACCESS Freely available online

PLOS | COMPUTATIONAL BIOLOGY

Determinants of Translation Elongation Speed and Ribosomal Profiling Biases in Mouse Embryonic Stem Cells

Alexandra Dana, Tamir Tuller*

The Department of Biomedical Engineering, Tel-Aviv University, Tel-Aviv, Israel

Abstract

Ribosomal profiling is a promising approach with increasing popularity for studying translation. This approach enables monitoring the ribosomal density along genes at a resolution of single nucleotides. In this study, we focused on ribosomal density profiles of mouse embryonic stem cells. Our analysis suggests, for the first time, that even in mammals such as *M. musculus* the elongation speed is significantly and directly affected by determinants of the coding sequence such as: 1) the adaptation of codons to the tRNA pool; 2) the local mRNA folding of the coding sequence; 3) the local charge of amino acids encoded in the codon sequence. In addition, our analyses suggest that in general, the translation velocity of ribosomes is slower at the beginning of the coding sequence and tends to increase downstream. Finally, a comparison of these data to the expected biophysical behavior of translation suggests that it suffers from some unknown biases. Specifically, the ribosomal flux measured on the experimental data increases along the coding sequence; however, according to any biophysical model of ribosomal movement lacking internal initiation sites, the flux is expected to remain constant or decrease. Thus, developing experimental and/or statistical methods for understanding, detecting and dealing with such biases is of high importance.

Citation: Dana A, Tuller T (2012) Determinants of Translation Elongation Speed and Ribosomal Profiling Biases in Mouse Embryonic Stem Cells. PLoS Comput Biol 8(11): e1002755. doi:10.1371/journal.pcbi.1002755

Editor: Sarah A. Teichmann, MRC Laboratory of Molecular Biology, United Kingdom

Received May 26, 2012; **Accepted** September 7, 2012; **Published** November 1, 2012

Copyright: © 2012 Dana, Tuller. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The study was supported by TAU. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

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

Introduction

Gene translation is the second major step of gene expression and thus has ramifications related to every biomedical discipline including human health [1,2], biotechnology [3], evolution [4–6], functional genomics [7,8] and systems biology [9,10]. One of the open questions in the field is related to the way translation efficiency is encoded in the transcript.

The most promising approach for studying gene translation is the ribosomal profiling method [11]. This approach was introduced only a few years ago but has already been successfully employed for answering various fundamental biological questions [12–19]. Specifically, ribosomal profiling has been used for: 1) understanding the mechanism of gene expression down-regulation by microRNAs [13], 2) understanding the dynamics of translation in mouse embryonic stem cells [12], 3) showing that the anti-Shine–Dalgarno sequence drives translational pausing and codon choice in bacteria [14], 4) studying the yeast meiotic program [15], 5) showing that miR-430 reduces translation before causing mRNA decay in zebrafish [17], and 6) to reveal the co-translational chaperone action of trigger factor in vivo [16].

In the current study we analyzed ribosomal profiles of mouse embryonic stem cells measured in a previous experiment [12]. The experiment output included ribosomal density measurements along hundreds of genes at a few time points, after preventing translation initiation. These data enabled us to infer the translation elongation speed in different genes, allowing us for the first time to study several biophysical aspects of translation elongation in mouse embryonic stem cells.

Results

Measuring translation elongation velocities

To study the kinetics of translation elongation in *M. musculus*, ribosome footprint profiles of isoforms expressed in embryonic stem cell were reconstructed based on a previous study [12]. Briefly, translation was halted by applying cyclohexamide. Fragments covered by ribosomes were mapped to the transcript and a baseline ribosomal read counts profile for each expressed isoform was created (see Methods). Let us denote these created profiles by RC_0^i, where i is the index of the analyzed isoform. In addition, to estimate the elongation speed of ribosomes, in three *additional* experiments harringtonine was used to stop translation initiation, while allowing ribosomes that already started translating the mRNA to continue their movement on it. Cyclohexamide was again applied 90/120/150 seconds after applying harringtonine to stop translation. In this work, the time difference between applying harringtonine and cyclohexamide for creating depleted profiles is named the '*run-off*' time.

Let us denote the ribosomal read counts obtained in each of these three experiments by $RC_{90}^i/RC_{120}^i/RC_{150}^i$ accordantly. The estimated **S**tarting **L**ocation of the depleted ribosomal profile (*SL*) was defined as the point where the ribosomal read counts profile of gene i at time point t (RC_t^i profile) reached half of the original

Author Summary

Gene translation is the process by which ribosomes translate mRNA molecules to proteins, a central process in all living organisms. Thus, understanding the biophysics of gene translation and the way its efficiency is encoded in the different features of the coding sequence has ramifications to every biomedical discipline. Recently, a new large-scale experimental approach named 'ribosomal profiling', has been developed for monitoring the ribosomal density at a resolution of single nucleotides. In this study, we analyzed ribosomal profiling data of mouse embryonic stem cells. These data enabled us to directly show that translation velocity is affected by the adaptation of codons to the tRNA pool, local mRNA folding of coding sequence, and local charge of the amino acids encoded in the coding sequence. In addition, our analyses suggest that ribosomal speed tends to be slower at the beginning of the coding sequence. Finally, we report possible biases in the 'ribosomal profiling' procedure that should be considered in future studies utilizing this method.

ribosomal read counts profile RC_0^i (Methods). Using these *SL* points, local translation elongation velocities were estimated for each analyzed isoform. Figure 1 outlines a schematic description of the method used to estimate the *SL* points, demonstrated on the uc007gge.1 isoform (see also Figure S7).

In the original work 4,994 isoforms with good read counts were found [12]. The authors noticed that the effect of harringtonine was best observed for genes longer than 750 codons, as for shorter genes the ribosomes managed to exit the mRNA for the used run-off times. Thus, only genes that were long enough (at least 750 codons) were used to infer the position of the *SL* points. In the current work, the same isoforms satisfying these conditions were analyzed, resulting in 785 processed isoforms (see Table S10). Let us define the three estimated *SL* points by x_1,x_2,x_3 corresponding to time points 90/120/150 respectively. Let us mark with dx_1,dx_2 the segments defined by $[dx_1,dx_2]$, $[dx_2,dx_3]$ accordantly, and the ribosomal average translation velocity in these segments by v_1 and v_2. The average translation velocity of a segment was estimated by dividing the segments' length by 30 seconds. For each gene and time point, various quality checks were performed to reliably estimate the position of the *SL* points (see more technical details in the Methods section). Eventually, only isoforms with *SL* points satisfying $x_1<x_2<x_3$ were selected, resulting in 692 valid isoforms out of the 785 processed isoforms (88%).

Translation elongation speed varies among genes and tends to increase along the coding sequence

Analysis of the data indicated that the median length of dx_1 was 128 codons (130±77 codons) while the median length of dx_2 was 184 codons (181±75 codons). Therefore, although the mean translation velocity of all genes is around 5.5 *codons/second* [12] (see Figure 2A and Tables S4, S6), the average translation velocity along the second segment (v_2) is larger than the average translation velocity along the first segment (v_1) (6+/−2.5 *codons/second vs.* 4.3+/−2.6 *codons/second*, Wilcoxon test $p=2.2*10^{-26}$ Figure 2A). This result remains significant under various estimations methods of these velocities.

We performed additional analyses to support the conjecture that translation elongation velocity is not similar among genes: first, the standard deviation of the estimated *SL* points was between 17% and 49% (Figure 2A–B, Table S4, S6, columns 1, 2, 3). Second, the relative difference between the two estimated velocities (calculated using $|v_2-v_1|/\min(v_1,v_2)$) resulted in a median value of 0.82 while the median value of the ratio v_2/v_1 resulted in a value of 1.37 (see also Figure 2C–D). To compare the attained results to simulated genes with uniform translation elongation rate, we simulated 692 synthetic genes with 1) lengths distribution identical to the lengths distribution of the analyzed genes, and 2) with constant codons translation efficiency (see Methods). The ribosomal profile of these genes was simulated with a biophysical model (see Methods), resulting in a much smaller difference between the calculated velocities v_1, v_2 (median = 0.01; KS-test: p-value $<1.81*10^{-271}$), as seen in Figure 2C. The ratio between the velocities v_2/v_1 was also much more moderate when calculated on these simulated ribosomal profiles (0.99+/−0.03, KS-test p-value $<1.56*10^{-295}$), as seen in Figure 2D. This comparison supports the claim that there is a high variance in the elongation speed of the analyzed genes.

Estimated translation elongation velocity is significantly associated with features of the coding sequence

In order to explain the high variability among segments length, those were analyzed with respect to different features of the coding sequence, such as the adaptation to the tRNA pool (*e.g.* the tAI [20] and the CAI measure [21]), local mRNA folding energy [22] and local charge of the translated amino acids [22,23]. Specifically, codons recognized by more abundant tRNA molecules increase the tAI measure, therefore we expect longer segments to positively correlate with this measure [24]. The CAI index, which measures the frequency of codons in a segment relatively to their appearance in highly expressed genes, is also expected to positively correlate with the segment length.

In addition, it was suggested that strong local mRNA folding tends to slow down ribosomal translation elongation as it increases the time it takes the helicases to unfold the mRNA molecules [24]. Therefore, segments more strongly folded (*i.e.* with lower folding energy (FE)) are expected to be shorter. Finally, the polypeptide must traverse two negatively charged regions to exit the ribosome [22,23,25], thus charged amino acids (specifically positively charged amino acids [23]) that are encoded in the codons preceding (upstream) the currently translated codon should have electrostatic interactions with the ribosome exit tunnel [22,23,25]. Therefore, segments more positively charged are expected to be shorter. More details about the calculation of these measures appear in the Methods section.

To estimate the distinct contribution of each of the coding sequence features to the elongation speed, we calculated the correlation between the length of the segments and each of these features, when controlling for the other two features, and after binning the data (details in the Methods section). Spearman correlation between the segments length and the genes' tAI/CAI when controlling for charge and folding energy of the segments resulted in a correlation coefficient of r = 0.29/0.21 (P<0.00615/0.049) accordantly. Spearman correlation between the segments' length and their mRNA folding energy when controlling for charge and gene tAI was r = 0.42 ($P<4.72*10^{-5}$). The correlation between the segments' length and their charge when controlling for folding energy and the genes' tAI was r = −0.21 (P<0.046) (additional analyses appear in the supplementary). Thus, the results reported in the current subsection support the conjecture that the translation elongation speed is independently affected by each of the following features of the ORF: the adaptation of the ORF codons to the tRNA pools, local mRNA folding and local amino acids charge.

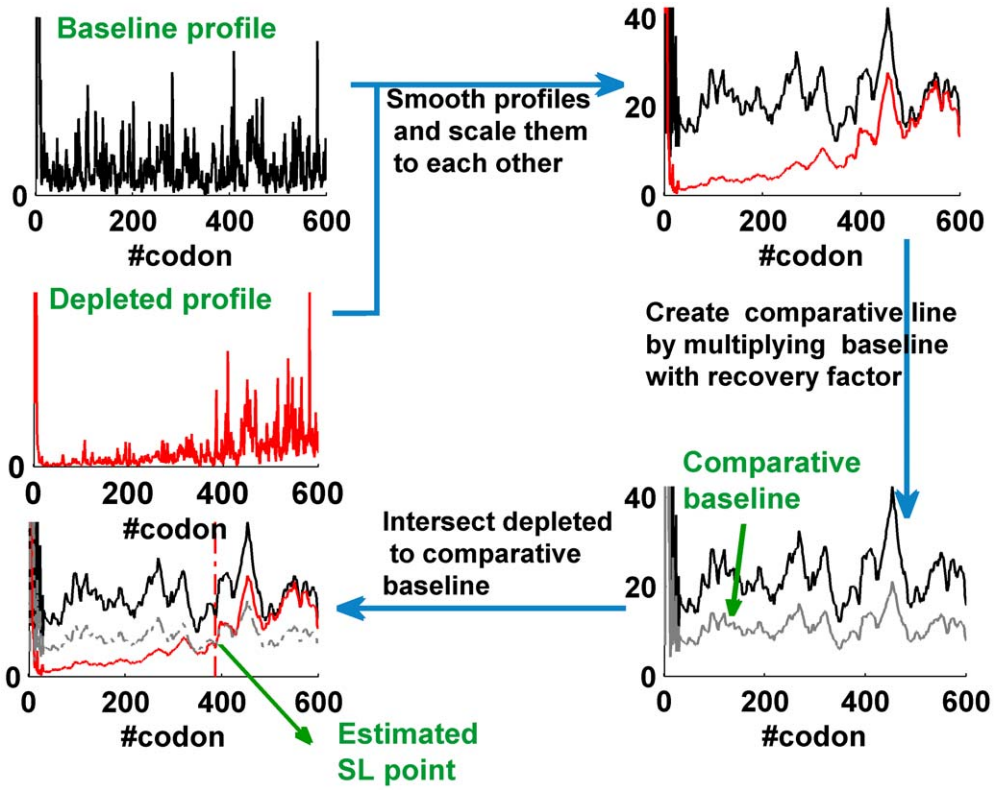

Figure 1. Estimating the *SL* point. A diagram outlining the methodology for estimating starting location points (*SL* points) demonstrated on the uc007gge.1 isoform. The black line depicts the ribosomal baseline profile while the red line represents a depleted profile (created using harringtonine to halt initiation; cychlohexamide was applied after 120 seconds). The dotted gray line represents the baseline profile multiplied by a recovery factor of 0.5, to which the depleted profile is compared. The red dotted line represents the estimated *SL* point.
doi:10.1371/journal.pcbi.1002755.g001

As mentioned in the previous section, the speed of translation elongation tends to increase along the coding sequence. Aiming at explaining this phenomenon, features measured on the first and second segment were also compared using a paired Wilcolxon test, resulting in significant values for folding energy (Wilcolxon test: $P<1.04*10^{-3}$) but not for tAI/CAI and charge. This suggests that in mouse, a possible explanation of the increase in translation speed along the coding sequence is the decrease in the strength of the mRNA folding along the coding sequence. Finally, a weak but significant correlation between the average v_1 and v_2 translation speed and the average transcripts length was observed in mouse (Spearman correlation: $r=-0.05$, $p=0.022$), supporting the conjecture that shorter genes are more efficiently translated.

Ribosomal flux inferred based on ribosomal profiling increases along the coding sequence, contradicting biophysical models of translation elongation

According to the accepted biophysical model of translation, during the elongation step ribosomes move along the coding sequence, translating each codon with a speed related to the features of the coding sequence in its vicinity and according to cellular factors such as concentrations of elongation factors and tRNA molecules. In addition, a ribosome may be delayed if a ribosome is located downstream in front of it [26]. It is also assumed that in general, ribosomal abortion during translation is relatively rare and that initiation usually occurs at the 5′UTR (*i.e.* ribosomes do not appear in the middle of the coding sequence [26]).

According to the protocol of the experiment (*e.g.* see [11], [12]), ribosomal footprint reads of a certain codon are generated when the codon is covered by ribosomes. From a biophysical perspective, slower codons are covered by ribosomes for a larger amount of time (relatively to other codons in the mRNA), creating a higher number of reads (for an illustration see Figure 3A).

In this study, for each analyzed isoform, both dx_1 and dx_2 segments were assumed to be translated in an equal time interval of 30 seconds, therefore according to the above assumption, on average, it is expected for the sum of read counts in the dx_1 and dx_2 segments (measured on the baseline profile RC_0^i) would be equal. Therefore, in each isoform the shorter segment is expected to have a higher ribosomal read count per nucleotide in comparison to the longer one.

Let us mark the **s**um of **r**ead **c**ounts in intervals dx_1 and dx_2 by SRC_1, SRC_2 accordantly. Let us define the percentage **d**ifference between SRC_1 and SRC_2 (relatively to the minimum of SRC_1 and SRC_2) by

$$DSRC(SRC_1,SRC_2)=100\cdot\frac{|SRC_2-SRC_1|}{\min(SRC_1,SRC_2)}$$

This measure is invariant to the genes' various mRNA levels and translation initiation rates, therefore enabling comparison between all analyzed isoforms. Using the above assumption, we expect this measure to be close to zero. Figure 3B shows the histogram of the *DSRC* measure calculated both on the real ribosomal profiles and on the simulated ribosomal profiles created using the TASEP biophysical model for various initiation rate values (see Methods). However, in contrast to the made biophysical assumptions, the results indicate that for a substantial part of genes, the *DSRC*

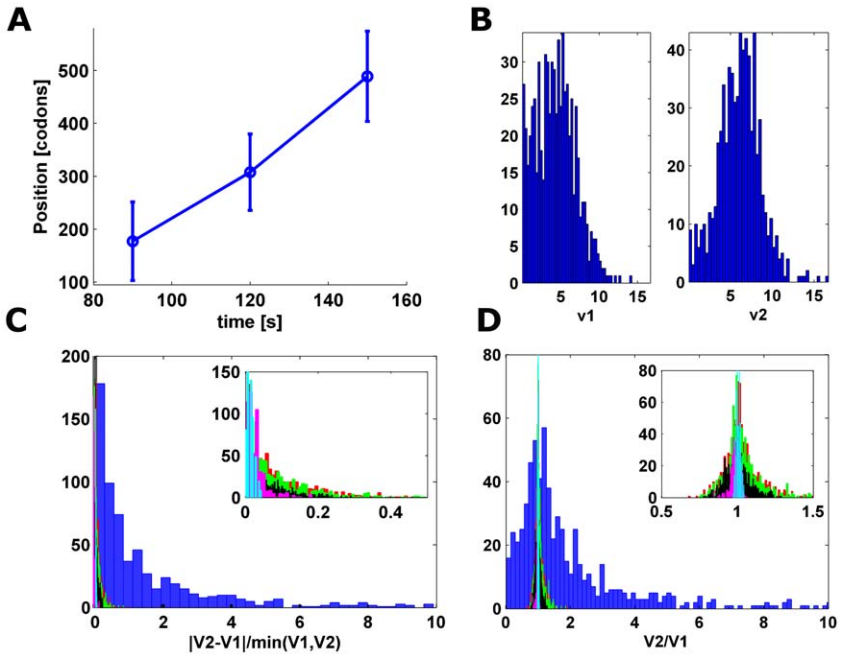

Figure 2. (A). Estimated position of the *SL* points (mean and standard deviation): v_1 = 4.3+/−2.6, v_2 = 6+/−2.5. v = 5.2+/−1.2 (Wilcoxon test: p = 2.2*10^{-26}). (B). v_1 and v_2 histograms. (C). Histogram of $|v_2-v_1|/\min(v_1,v_2)$ measure calculated on: 1) the experimental data (blue) (median value 0.82) 2) on simulated ribosomal densities of the analyzed isoforms for low/high/proportional initiation rates (green/red/black) and 3) ribosomal densities created using codons of equal translation efficiency for low/high initiation rates (magenta/teal). For the simulations, the obtained median values of the $|v_2-v_1|/\min(v_1,v_2)$ measures were 0.06/0.06/0.06/0.02/0.01, significantly lower than in the case of the experimental data (KS p-value $<6.18*10^{-153}$ in all cases). The inset shows the ratio for the simulative data only. (D). Histogram of the v_2/v_1 ratio calculated on real and simulative data. The median value of this measure for the real ribosomal profiles was 1.37, significantly higher than for the simulative data, which resulted in median values of 1/1.01/1.01/1/1.01 accordantly (KS p-value $<5.67*10^{-250}$ in all cases).
doi:10.1371/journal.pcbi.1002755.g002

measure is abnormally high (median value of 88 *vs.* 1–6 for simulative data of different levels of noise; KS test, all p-values $<3.97*10^{-215}$).

In addition, the ribosomal flux at a certain codon i along the coding sequence is defined as the multiplication of the translation velocity and density at this point ($v_i \cdot D_i$). Therefore, according to any biophysical model with negligible amount of initiation events inside the ORF, we expect the flux to be constant (*i.e.* $v_i \cdot D_i \approx v_j \cdot D_j$ for different i,j) or decrease (due to ribosomal abortion);

Let us mark the mean ribosomal read counts measured in the first and second segments by $\bar{D}_1$ and $\bar{D}_2$ respectively and the average velocity in the first and second segment by $\bar{v}_1$ and $\bar{v}_2$. If we assume that the local flux remains constant, we also expect that $v_1 \cdot D_1 \approx v_2 \cdot D_2$. Given that the average velocities of $\bar{v}_1$, $\bar{v}_2$ in both the first and second intervals were measured during the *same* time intervals, we can rewrite this relation as $dx_1 \cdot \bar{D}_1 \approx dx_2 \cdot \bar{D}_2$

Thus if

$$dx_1 \cdot \bar{D}_1 \approx dx_2 \cdot \bar{D}_2 \leftrightarrow dx_2/dx_1 \approx \bar{D}_1/\bar{D}_2 = 1/(\bar{D}_2/\bar{D}_1)$$

we would expect the correlation between $\bar{D}_2/\bar{D}_1$ and dx_2/dx_1 to be *negative*. Intuitively, for a given gene, longer segments should have relatively lower mean read counts. Indeed, the calculated ratios for the simulated densities resulted in a negative correlation

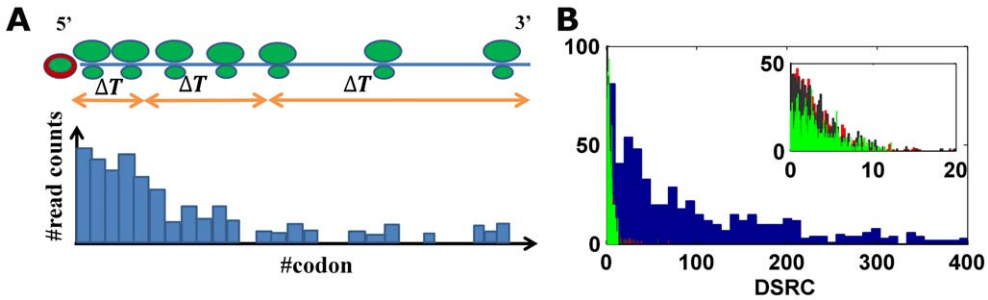

Figure 3. (A). Ribosome read counts measures according to the biophysical model. The green round shapes represent the ribosomes on the mRNA, which is depicted with a blue line. According to the biophysical model, segments of high ribosomal read counts are associated with regions more slowly translated (bottom graph). The orange double arrows represent the mRNA segments being translated in equal time intervals. (B). *DSRC* histogram calculated on real (blue) and simulated ribosomal profiles for low/high/proportional initiation rates (black/red/green) with zero noise level. The calculated median value of this measure is 88/2.46/2.39/2.38 accordantly.
doi:10.1371/journal.pcbi.1002755.g003

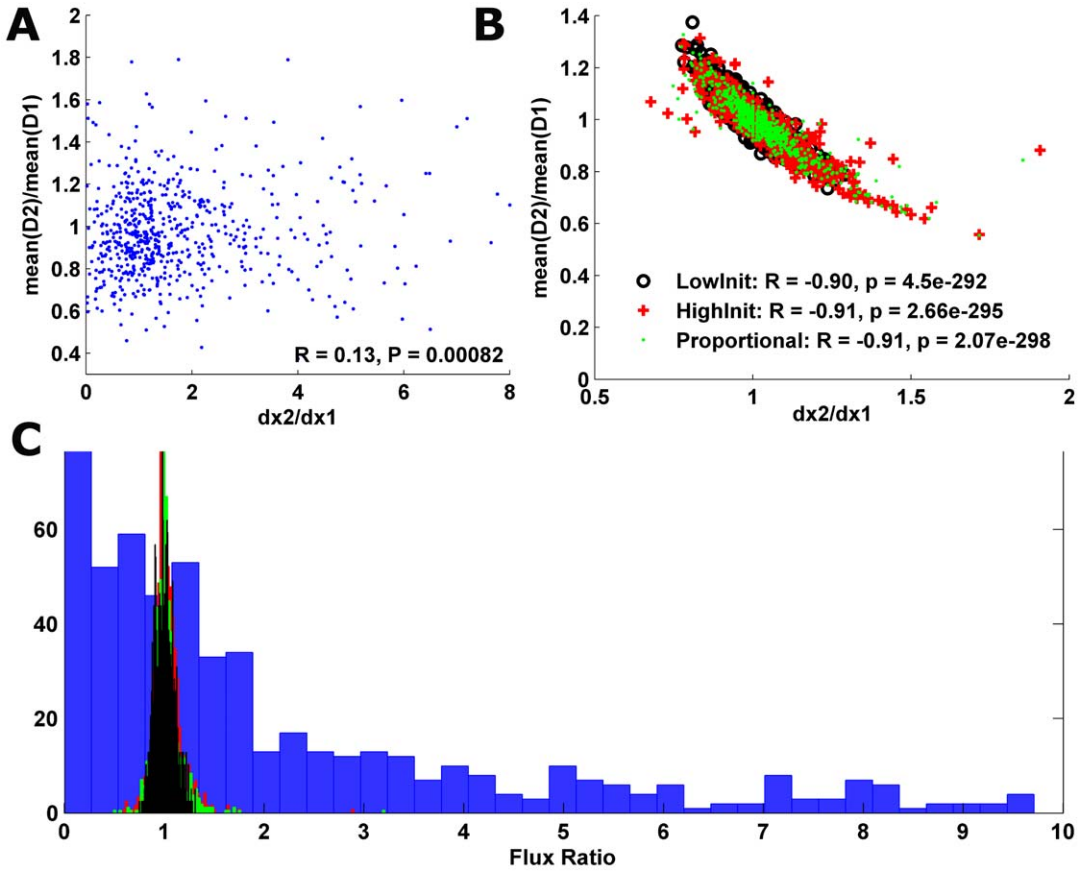

Figure 4. (A). $\bar{D}_2/\bar{D}_1$ ***vs.*** dx_2/dx_1 **measured on the real data (B).** $\bar{D}_2/\bar{D}_1$ ***vs.*** dx_2/dx_1 **measured on the simulative data for low/high/proportional initiation rates (black circles/red crosses/green dots) (C). Histogram of the flux ratio** $(dx_2\cdot\bar{D}_2)/(dx_1\cdot\bar{D}_1)$ **measured on the real (blue; median = 1.69) and on the simulative data created using low/high/proportional initiation rates (black/red/green; median = 1/1.01/1.01; KS test: p-value <9*10^{-95}).**
doi:10.1371/journal.pcbi.1002755.g004

(Figure 4B, Spearman correlation of R = −0.9, P<10^{-291}; R = −0.91, P<10^{-294}; R = −0.91, P<10^{-297}; for low/high/proportionate initiation rates). However, when measured on real ribosomal read counts profiles, the correlation between $(\bar{D}_2/\bar{D}_1)$ and dx_2/dx_1 achieved a significant *positive* value (R = 0.13, P<0.00082; Figure 4A), contradicting the accepted translation model. Finally, the flux itself $dx_i\cdot\bar{D}_i$ is expected to remain constant or decrease (due to ribosomal abortion), *i.e.* $(dx_2\cdot\bar{D}_2)/(dx_1\cdot\bar{D}_1)\leq 1$. Yet, we found that this ratio tends to *increase* $(\text{median}((dx_2\cdot\bar{D}_2)/(dx_1\cdot\bar{D}_1)) = 1.69 > 1)$.

Next, we calculated the values of all presented measures on the simulated ribosomal profiles for different initiation rate regimes (see Methods) and compared them to the values obtained when calculating them on the real ribosomal profiles. This analysis resulted in significantly different values: the $(dx_2\cdot\bar{D}_2)/(dx_1\cdot\bar{D}_1)$ measure calculated on the simulative data resulted in a median value of 1.01 (KS test in comparison to the measured data: p-value <9*10^{-95}; Figure 4C, Table S9), while the difference between the velocities v_1 and v_2 resulted in a median value of 0.06 (KS test: p-value <6.18*10^{-153}; Figure 2B). In addition, the ratio between the velocities resulted in median values of 1–1.01 (KS test: p-value <5.67*10^{-250}) (Figure 2C). Overall, the comparisons between all measures calculated on the experimental data and on the simulative ribosomal profiles created by the biophysical model point on the existence of substantial biases in the data produced by the ribosomal profiling procedure.

Discussion

In this study, we reanalyzed the ribosomal profiling data of mouse embryonic stem cells that was generated in a previous study [12]. Our analysis demonstrates that even for relatively long analyzed genes, that are not expected to be under strong selection for translation efficiency [27], in unusual tissue/conditions such as embryonic stem cells, translation elongation speed is affected by features such as the adaptation of codons to the tRNA pool, local mRNA folding, and charge.

In addition, our analysis directly shows for the first time that the translation elongation speed tends to increase along the coding sequence. The reasons for this phenomenon may be related to the fact that at the beginning of the coding sequence features such as adaptation to the tRNA pool and mRNA folding strength tend to slow down ribosomal movement (see, for example, [22,24]). This may also be related to the fact that there is a selection for lower codon bias at the beginning to reduce the costs of both missense and nonsense translational errors [28,29]. The statistical analysis performed in this study support the conjecture that the slower speed at the beginning of the coding sequence is due to stronger mRNA folding in this region. This phenomenon, however, may also be related to yet unknown properties of this process or to biases of the ribosomal profiling methods.

Finally and importantly, at least in the reported study, our analysis demonstrates the existence of some unexplained

deviations between the output of the ribosomal profiling approach and any of the accepted models of translation elongation, which assume that the rate of initiation from sites inside the ORF is negligible. This discrepancy may be explained by the fact that current models of translation elongation are inaccurate and, for example, initiation does tend to occur from sites inside coding sequences. However, the most plausible explanation is that ribosomal profiling approach, as in the case of the more traditional approaches for studying mRNA levels (*e.g.* [30]), includes experimental biases that should be further explored. Another bias of the ribosomal profiling approach which is related to the increased ribosomal density at the beginning of the ORF has been suggested recently in [12].

We also suggest a few explanations for these observed biases, while taking into consideration that there might be additional sources of bias in the ribosomal profiling protocol that are not mentioned here. For example, an insufficient number of mRNA molecules could increase the estimation errors and bias all the presented measures. Specifically, the ribosomal profiling approach produces for each gene the ribosomal positions along the mRNA molecules that have been transcripted from it and that are present in the cell at the time of the experiment. As the read counts per location of a single mRNA are stochastic, averaging them over many mRNA molecules of a gene should theoretically produce a profile that is similar to the stationary density profile of the gene. Thus, the number of mRNA copies affects the averaged profile and eventually the quality of the estimated measures mentioned in this study. In practice, genes with a relatively low number of mRNA molecules can result in highly biased profiles. Indeed, when we modified our computational simulation of the experiment to simulate a low number of mRNA molecules per gene (see Methods), the correlation between $\bar{\boldsymbol{D}}_2/\bar{\boldsymbol{D}}_1$ and dx_2/dx_1 decreased (Figures S12, S13, S14) while the *DSRC* measure increased (Figures S9, S10, S11), contrary to the expected trend.

Another source of bias may be related to the fact that the current ribosomal density protocol involves filtering some of the reads, distorting the resultant ribosomal density profiles. Specifically, by the protocol of the experiment, only short mRNA fragments that are covered by exactly one ribosome (*i.e.* monosomes) are purified for further analyses [11,31], while mRNA segments covered by polysomes are discarded. Thus, it is also possible that the reported biases are, at least partially, due to the fact that fragments that origin from ribosomes located very close to each other on the mRNA are filtered and not analyzed, creating deviated ribosomal profiles. Indeed, cases of fragmented mRNA covered with more than one ribosome as a result of very close ribosomes were reported in a previous study [32]. In addition, when only monosomal footprints were considered in the simulation (see Methods), we obtained a decrease in the correlation between the $\bar{\boldsymbol{D}}_2/\bar{\boldsymbol{D}}_1$ and dx_2/dx_1 ratios and a major increase in the *DSRC* measure (see Figures S9, S10, S11, S12, S13, S14).

The deviations from the accepted biophysical model could also be explained by the non-uniform effect of the harringtonine/cyclohexamide substances on the different mRNA molecules, causing uneven run-off times, and distorting the location of the *SL* points. The simulation of this possible experimental bias (see details in Methods) also resulted in an increased *DSCR* and a decrease in the correlation between $\bar{\boldsymbol{D}}_2/\bar{\boldsymbol{D}}_1$ and dx_2/dx_1 (Figures S15, S16, S17).

Finally, complex relations between the sequence features, their effect on ribosomal density and on the output of the ribosomal profiling approach may also contribute to the deviation from the biophysical model. For example, it was suggested that elongation speed and ribosomal density are affected by the strength of the local folding of the mRNA (stronger folding→slower elongation speed→high ribosomal density) [22]. However, it is also possible that stronger mRNA folding decreases the efficiency of footprint production in the ribosomal profiling protocol (*e.g.* the efficiency of RNase activity decreases for mRNA fragments with strong folding; *e.g.* see [33]), contributing to a distorted ribosomal density profiles.

Nonetheless, currently, the ribosomal profiling approach is the major method for studying gene translation, therefore understanding these biases and accurately correcting them should significantly affect studies in various biomedical disciplines. As was demonstrated in this study, one possible direction for detecting such biases is by comparing the ribosomal profiling outcome to the computational biophysical models using statistical analysis. We believe that such approach will be used in the future for employing filters and normalization procedures that are inversed to the noise/bias obtained in the experimental procedure and for adjusting the experimental procedure itself.

Methods

Reconstructing the genes' ribosomal profiles

Sequencing data were downloaded from the GEO database (accession number GSE30839) [12]. We analyzed all data related to the study of the kinetics of translation elongation. The specific processed files are summarized in Table S1.

Sequenced reads comprise short RNA fragments of different lengths; therefore, a generated linker sequence (CTGTAGGCACCATCAATTCGTATGCCGTCTTCTGCTTGAA) was attached to enable the recovery of the original fragment. More details of this method appear in the original work [12]. In this study, linkers were first detected and removed from the published fragments and only then aligned to transcripts. The start location of the linker was estimated to be between the 20–36 *nt* of the RNA fragment. Next, the distance between the estimated linker and the published linker was calculated (in terms of number of different nucleotides); a valid linker was accepted if this distance differed by up to two nucleotides. If no valid linker was found, the fragment was rejected. Table S2 summarizes the number of fragments published by Ingolia *et al.* (see Table S1, column 2) and the percentage of processed fragments after removing the attached linker (column 3).

Aligning the fragments directly to the genome resulted in a high number of ambiguous matches. Therefore, fragments were aligned to known transcripts (exons) and spliced junctions. The *M. musculus* transcripts were derived from the UCSC Genes data set [34] and the alignment was performed using the Bowtie software [35], allowing up to two mismatches.

As mentioned by Ingolia *et al.*, fragments of different lengths tend to have slighter different A site locations, therefore the beginning of the A site for fragments of 29–30/31–33/34–35 nt was defined to begin +15/+16/+17 *nt* relatively to the 5′ end of the fragment. Additional details about this topic appear in the original work [12].

As summarized in Table S2, part of the processed fragments matched to more than one location. To overcome multiple mapping of a single fragment, we performed the following procedure: first, only fragments aligning to a single location were mapped. In the second iteration, for all fragments aligning to more than one location, the mean read counts in the region of the possible locations was calculated (10 *nt* before and after the location of the A site for each possible location). These mean read counts defined the probability of an ambiguous fragment to be aligned to only one of the locations.

For each isoform, nucleotide read counts profiles were reconstructed by assembling read counts of relevant exons and spliced junctions. Codon reads were calculated by averaging the obtained reads of each three non-overlapping consecutive nucleotides.

Estimating the position of the ribosomes at each time point by the original method

In the *original* work, the RC_t^i profiles were smoothed using an averaging window of five codons and normalized by the average read counts of codons 800–1000. This normalization assumed that read counts in regions not affected by harringtonine (codons 800–1000) have a similar value (for each one of the run-off profiles apart). When assuming the experiment is reproducible, *i.e.* ribosomal read counts of all RC_t^i profiles are similar after the first 750 codons (the harringtonine effect did not extend beyond this point for any isoform in the experiment [12]), it is possible to estimate the **S**tarting **L**ocation (*SL*) point of a depleted profile RC_t^i by comparing it to the baseline profile, RC_0^i. The *SL* of the depleted ribosomal profile of each isoform was defined as the position beyond the first 40 codons, where the normalized ribosomal density profile RC_t^i exceeded a value of 0.5. In this work, this parameter is defined as the recovery factor. Isoforms with *SL* points not satisfying $x_1 < x_2 < x_3$ (see Table S3) were discarded. When smoothing the profiles with longer averaging windows, the number of isoforms with non-physical *SL* points reduced to 141 (out of 785, see also Table S3).

Estimating the position of the ribosomes at each time point by the new method

Further study of the nature of ribosomal profiles revealed that the original *SL* estimation method suffers from some difficulties: the results presented in Figure S3-A show that read counts in regions not affected by harringtonine (beyond the 750$^{\text{th}}$ codon, excluding the last 20 codons) have a high variability, therefore their average read count value cannot be used for normalizing the ribosomal profiles. In addition, in the original method the *SL* point was defined as the location where the run-off profile exceeded the threshold value 0.5. This criterion assumes again that RC_0^i profiles are relatively homogenous, and small spikes caused by noises can be filtered by first smoothing them. However, the results in Figure S3-B show that different profiles have a high read counts variability, also suggesting that ribosomal read counts could be position dependent, making the comparison of the run-off profile to a static threshold of 0.5 problematic.

To overcome these issues, in the current work we suggested scaling each run-off profile to the baseline profile by a dynamic factor that derives from the read counts beyond the 750$^{\text{th}}$ codon of both profiles (excluding the last 20 codons). This factor is set to minimize the distance between these regions. In the current study, we also tested the effect of the smoothing window size (10/15/20/25/30 codons) on the number of genes with physical *SL* points, as presented in Table S5. The *SL* location of each isoform was defined as the position beyond the first 40 codons, where the ribosomal density profile RC_t^i exceeded the value of the RC_0^i profile multiplied by the recovery factor. This created a dynamic threshold for the run-off profiles to be compared to. The influence of the recovery factor on the number of genes with physical *SL* points was also evaluated, as presented in Table S6. In addition, to improve robustness of the method to local bursts of noise, an *SL* point was defined to be valid if 50% of the next 20 points could also exceed the dynamic threshold. The optimal smoothing window size and recovery factor were selected to maximize the number of genes whose *SL* points were physically estimated ($x_1 < x_2 < x_3$), resulting in a window size of 30 codons and a recovery factor of 0.5 (see Table S3, S4, S5, S6).

To compare between the methods' ability to correctly estimate *SL* points in a noisy environment, both the original [12] and the newly suggested methods were also evaluated on synthetic data created using the TASEP model (*e.g.* see [22]). *SL* points were estimated for different run-off times and different levels of additive noise (see Methods, evaluating the error rate of the *SL* points). Figures S4, S5, S6 show the mean and standard deviation estimation error as function of noise level and size of the smoothing window for both estimation methods. As seen from the results, on the simulative data the newly suggested method achieved a lower estimation error for all levels of noise and smoothing window sizes.

For comparison, in this work, the various tested measures were calculated based on *SL* points estimated using both methods. The smoothing window size was set to 30 codons and the recovery factor was set to 0.5. The figures in the main text were generated using the new method with these parameters. More details appear in Text S1.

Calculating the average folding energy of a segment

Folding energy (FE) of a nucleotide was defined as the folding energy of a 40 *nt* segment, starting from the current nucleotide. The segment's FE was calculated using the rnafold Matlab function [36]. The FE of a gene (segment) was defined as the average folding energy of its nucleotides.

Calculating the average tAI measure of a segment

Codon tAI values were calculated according to [20], using tRNA copy numbers published in http://gtrnadb.ucsc.edu/Mmusc10/. The tAI value of a segment was calculated using:

$$tAI_g = \left(\prod_{k=1}^{l_g} w_{ikg}\right)^{\frac{1}{l_g}}$$

Where w_{ikg} is the relative adaptiveness of codon of type i, j the index of the codon and l_g the number of codons in segment g. Let $tCGN_{ij}$ be the copy number of the j^{th} anti-codon that recognizes the i^{th} codon, and let S_{ij} be the selective constraint of the codon/anti-codon coupling efficiency. Then, the absolute adaptiveness value of a codon is defined by

$$W_i = \sum_{j=1}^{n_i} (1 - S_{ij}) tCGN_{ij}$$

The relative adaptiveness value of a codon w_i is obtained by normalizing W_i with the maximal W_i value among its 61 values (for specific values see Table S10).

Calculating the average CAI of a segment

To calculate CAI of a segment, codons were ranked according to their usage in ribosomal proteins $\{f_i\}_{i=1}^{61}$ (Table S10). Using these frequencies, the CAI of a segment was similarly defined in the following manner:

$$CAI_g = \left(\prod_{k=1}^{l_g} f_{ikg}\right)^{\frac{1}{l_g}}$$

Calculating the average charge measure of a segment

For each gene, a vector of charges was defined by assigning +1 to positively charged amino acids (Arg and Lys) and −1 to negatively charged amino acids (Asp and Glu). The charge of other amino acids was set to 0. A sliding window of 40 codons was applied on the charge vector to smoothen the charge effect on the mRNA. The overall charge of a segment was defined as the sum of its charges.

Simulating ribosomal densities

To enable analysis of various features in a simulated environment, ribosomal densities of the analyzed isoforms in this work were calculated using the TASEP biophysical translation model, previously used in different studies (*e.g.* [22,37]). The mRNA was modeled using a lattice of N sites, representing the number of codons of the isoform. Each ribosome was defined to cover 11 codons and the A site was located at the sixth codon. During translation, any codon could be covered at a time by a single ribosome at most. In each step of the simulation, a single ribosome was allowed to attach itself to the lattice or advance to the next codon if the first/next six codons were not occupied. The time between initiation attempts was set to be exponentially distributed with a constant rate λ. Similarly, the time between jump attempts from site i to site $i+1$ was assumed to be exponentially distributed with rate λ_i.

The time between events, (initiation or jumping between sites) is therefore exponentially distributed (minimum of exponentially distributed random variables) with rate:

$$\mu(n_i)=\lambda+\sum_{i=1}^{N} n_i\lambda_i$$

where i describes the site (codon) number on the lattice and $n_i=1$ if codon i is being translated, otherwise $n_i=0$. Therefore the initiation probability is given by $\lambda/\mu(n_i)$ and the probability of a ribosome to jump from site i to $i+1$ is given by $n_i\lambda_i/\mu(n_i)$.

The λ_i parameter was determined for each codon type according to its translation efficiency, estimated by the tAI measure (for specific values see Table S10). The initiation rate λ was studied for different values, depicting different initiation rate regimes.

To achieve an initial scattering of the ribosomes on the mRNA, 10^6 simulations steps (events) were performed. This number of steps was selected to enable full initial steady state ribosomal cover for the analyzed genes. In general, longer genes or genes with low initiation rates (relative to the gene's codon translation efficiencies) require a higher number of simulation steps to achieve this condition.

To calculate ribosomal density profiles, we simulated another 10^7 steps. In each step, the simulation updated the time each site was translated by a ribosome. The final vector of times representing the total time a site was translated by a ribosome was then normalized by the total time of the simulation. In addition, the final scattering location of the ribosomes on the mRNA was saved.

Simulating ribosomal densities for different run-off times

Simulated ribosomal profiles were created by using three different initiation rate regimes (λ): low, high and proportional to the genes' mean ribosomal read counts. The low initiation rate was set to be 10% of lowest codon translation rate (based on the tAI measure), while the high initiation rate was set to be twice the value of the highest codon translation rate (based on the tAI measure).

Proportional initiation rates were set for each isoform according to its measured mean ribosomal read counts (excluding the first 60 and last 40 codons). This initiation rate type assumed that in general, genes with higher mRNA and ribosomal densities levels (thus higher ribosomal read counts) are more highly expressed, therefore their initiation rate should be higher. Thus, for this regime initiation rate of the isoform with the lowest mean read counts was set as half of the slowest codon translation rate, while the initiation rate of the isoform with the highest mean ribosomal read counts value was set to twice the value of the highest codon translation rate. Initiation rates for the rest of the genes were set with equal distance between these two extremes, according to the genes' mean ribosomal read counts.

To simulate ribosomal profiles for different run-off times, the TASEP model was run 10^6 simulations steps to achieve a steady state ribosomal spread on the mRNA. Initiation halting was simulated for 100 different run-off times, defined by

$$\Delta T, 10\Delta T, \ldots 1000\Delta T$$

where ΔT was defined to be the maximal translation time of a codon (based on the tAI measure).

To simulate numerous mRNA copies per gene, for each run-off time and analyzed gene, 500 ribosomal density profiles were calculated and those were averaged with equal weight to obtain a representative ribosomal profile for each gene and run-off time. More details appear in Text S1.

Simulating ribosomal densities for different run-off times for genes with codons of equal translation efficiency

In the original work, it was claimed that translation elongation is constant throughout the translation of the mRNA. To test this hypothesis, we created synthetic genes using the length of the analyzed genes in this work, but with codons of equal translation efficiency, which was set as the mean tAI value of the codons calculated in *M. musculus*. Using the TASEP model, the ribosomal profile of each one of the synthetic genes was created for different run-off times $[\Delta T, 10\Delta T, \ldots 1000\Delta T]$ for low and high initiation rates. More details appear in Text S1.

Evaluating the error rate of methods that estimate *SL* points

To allow accuracy evaluation of the original and new method for estimating *SL* points, ribosomal density profiles with specific run-off times were created, as previously described. To test the robustness of the estimation method for different levels of noise, additive uniformly distributed noise of different levels was added prior to estimating the *SL* points of each analyzed gene. The noise level added to each gene was selected to be proportional to its maximal *simulated* ribosomal density, such that

$$N\sim U\left[\frac{-\alpha\max[RC_x^i]}{4},\frac{\alpha\max[RC_x^i]}{4}\right],\quad \alpha=[\frac{1}{40},\frac{2}{40},..1]$$

Let us mark by $\hat{x}_\alpha$ the estimated *SL* location for a noise level characterized by α. The estimation error is then defined by

$$err(\hat{x}_\alpha)=|\hat{x}_\alpha-\hat{x}_0|$$

The *SL* points for all simulated genes for run-off times of $[20\Delta T, 50\Delta T, 80\Delta T, \ldots\ 200\Delta T]$ were calculated for the above

noise levels. The general estimation error for a given noise level was defined as the average estimation error for all tested genes and run-off times. More details in Text S1.

Calculating different measures on the simulated ribosomal densities

For each simulated ribosomal profile (based on the real analyzed genes) and various initiation rates (low/high/proportional) the estimated *SL* points were calculated for run-off times of $[150\Delta T, 200\Delta T, 250\Delta T]$. These points were selected to resemble the real aggregated profiles (see Figures S1, S2).

These *SL* points were used for calculating the ratio between the estimated velocities v_2 and v_1, analysis of the *DSRC* measure and correlation between the ratio of the mean read counts and the ratio of the segments length.

In addition, these measures were also calculated for the simulated ribosomal profiles of genes composed of codons with equal translation efficiency (same run-off times as described above), for low and high initiation rate. More details appear in Text S1.

Simulating the influence of removing fragments covered by polysomes on the obtained ribosomal densities

To simulate ribosomal densities profiles obtained after filtering long fragments (created by adjacent ribosomes), for each simulated mRNA copy, ribosomal read counts were considered only for fragments covered by ribosomes that had a least one codon gap between themselves and their neighboring ribosomes, on both sides (using the final ribosome scattering on the mRNA). More details appear in Text S1.

Simulating non-uniform effect of harringtonine

To simulate a non-uniform effect of the propagation time of harringtonine, the analyzed isoforms were simulated using the TASEP model for low initiation rate (this regime results in profiles similar to the real measured profiles, see Figure S1, S2). For each gene, initiation halting was calculated for the following run-off times $[\Delta T, 2\Delta T, \ldots 750\Delta T]$ when using 500 mRNA copies per gene. Let us denote the ribosomal profile of gene i calculated for the mRNA copy j and run off time t by RD_{ij}^t. Let us denote the aggregated profile of gene i for the run-off time by RD_i^t. The non-uniform effect of harringtonine was simulated for each gene by aggregating different run-off profiles is the following manner:

$$RD_i^t = \frac{1}{J}\sum_{j=1}^{J} RD_{ij}^{\max[1,t-\Delta t_j]}$$

Where J is the number of mRNA copies simulated per gene and Δt_j is a random variable, such that $0 \le \Delta t_j \le K$. The simulation was calculated for $K = [0, 10\Delta T, 20\Delta T, \ldots 200\Delta T]$. The higher the K value, the more prominent the effect of the non-uniform propagation time of haringtonine. More details appear in Text S1.

Calculating correlations between various measurements and computation of p-values

The comparison between the translation velocity v_1 and v_2 was done using the paired Wilcoxon test, as supplied in the Matlab 2011b software. The comparison between the $|v_2 - v_1| / \min(v_1, v_2)$, v_2/v_1, *DSRC*, $(dx_2 \cdot \bar{D}_2)/(dx_1 \cdot \bar{D}_1)$ measures, calculated on the real ribosomal profiles and on the simulated ribosomal profiles, was done using the two samples Kolmogorov-Smirnov (KS)-test. The correlation between the segments' length and their tAI/CAI/FE/charge properties was calculated using partial Spearman correlation, as supplied in the Matlab 2011b software. The comparison between the translation velocities of segments with top/bottom 20%–50% of the tAI/CAI/FE/charge properties that appear in the supplementary results was calculated using the unpaired t-test and the two samples KS-test. The value of the tAI/CAI/FE/charge in the first and second segment (dx_1 and dx_2) was also compared using a Wilcolxon test. The correlation between the tAI/CAI and gene length was calculated using Spearman correlation.

Before we performed partial correlation between tAI/CAI/FE/charge measurements and segment length we binned the data in the following manner: first, segments were sorted according to their length and then divided into bins of 15 samples. For each bin, the average length/tAI/CAI/folding energy/charge was calculated in order to reduce noise.

Supporting Information

Figure S1 Reconstructed ribosomal profiles using real fragments, for different run-off times – average view.
(TIF)

Figure S2 Simulated ribosomal profiles for different run-off times – average view.
(TIF)

Figure S3 Histogram of the normalized standard deviation (STD) calculated for genes with good reads and with at least 1000 codons. The standard deviation was calculated using the real ribosomal profiles RC_0^i (blue) and based on simulative profiles created using the TASEP model. We considered different initiation rate regimes for the TASEP - low (red), high (black) and proportional (green). (A.) Normalized STD calculated on read counts of codons 730–1000. (B.) Normalized STD calculated on read counts of all codons, except for the first 40 and last 20 codons.
(TIF)

Figure S4 Estimation errors of the old (red) and the newly suggested estimation method (blue) as function of different noise levels, created with a TASEP simulation with *low* initiation rates.
(TIF)

Figure S5 Estimation errors of the old (red) and the newly suggested estimation method (blue) as function of different noise levels, created with a TASEP simulation with *high* initiation rates.
(TIF)

Figure S6 Estimation errors for the old (red) and the newly suggested estimation method (blue) as function of different noise levels, created with a TASEP simulation with *proportional* initiation rates.
(TIF)

Figure S7 Estimated *SL* points using both the old and the newly suggested methods on the ribosomal read counts profile of isoform uc007gge.1.
(TIF)

Figure S8 Explaining the semgents' length by using their tAI/CAI/folding energy/charge values. Segments were divided into two groups (top/bottom 20%(black)/30%(red)/40%(blue)/50% (green)) according to their genes' (A.) tAI, (B.) CAI and segments' (C.) folding energy and (D.) charge values.
(TIF)

Figure S9 *DSRC* measure calculated for simulated ribosomal profiles for low and high initiation rates; for each isoform we simulated 20 mRNAs. (A.) Read count profiles created using a low initialization rate, constructed with all fragments or (B.) only fragments covered by monosomes. (C.) Read count profiles created using high initialization rate, constructed with all fragments or (D.) with fragments covered only by monosomes.
(TIF)

Figure S10 *DSRC* measure calculated for simulated ribosomal profiles for low and high initiation rates; for each isoform we simulated 50 mRNAs. (A.) Read count profiles created using a low initialization rate, constructed with all fragments or (B.) only fragments covered by monosomes. (C.) Read count profiles created using high initialization rate, constructed with all fragments or (D.) with fragments covered only by monosomes.
(TIF)

Figure S11 *DSRC* measure calculated for simulated ribosomal profiles for low and high initiation rates; for each isoform we simulated 500 mRNAs. (A.) Read count profiles created using a low initialization rate, constructed with all fragments or (B.) only fragments covered by monosomes. (C.) Read count profiles created using high initialization rate, constructed with all fragments or (D.) with fragments covered only by monosomes.
(TIF)

Figure S12 Spearman correlation between $\bar{D}_2/\bar{D}_1$ and dx_2/dx_1 calculated for simulated ribosomal profiles for low and high initiation rates; for each isoform we simulated 20 mRNAs. (A.) Read count profiles created using a low initialization rate, constructed with all fragments or (B.) with fragments only covered by monosomes. (C.) Read count profiles created using high initialization rate, constructed with all fragments or (D.) with fragments covered only by monosomes.
(TIF)

Figure S13 Spearman correlation between $\bar{D}_2/\bar{D}_1$ and dx_2/dx_1 calculated for simulated ribosomal profiles for low and high initiation rates; for each isoform we simulated 50 mRNAs. (A.) Read count profiles created using a low initialization rate, constructed with all fragments or (B.) with fragments only covered by monosomes. (C.) Read count profiles created using high initialization rate, constructed with all fragments or (D.) with fragments covered only by monosomes.
(TIF)

Figure S14 Spearman correlation between $\bar{D}_2/\bar{D}_1$ and dx_2/dx_1 calculated for simulated ribosomal profiles for low and high initiation rates; for each isoform we simulated 500 mRNAs. (A.) Read count profiles created using a low initialization rate, constructed with all fragments or (B.) with fragments only covered by monosomes. (C.) Read count profiles created using high initialization rate, constructed with all fragments or (D.) with fragments covered only by monosomes.
(TIF)

Figure S15 Simulating the effect of unequal propagation time of harringtonine. (A.) $K=0$ (B.) $K=40\Delta T$ (C.) $K=90\Delta T$ (D.) $K=140\Delta T$. As can be seen from the results, an increased non-uniform harringtonine effect disturbs the profiles and decreases the slope of the run-off profiles.
(TIF)

Figure S16 Estimating the bias of *SL* points caused by non-uniform effect of harringtonine. (A.) Mean and standard deviation of the *SL* points were calculated for each of the tested K values, in comparison to the *SL* points calculated for $K=0$. (B.) Velocities ratio v_2/v_1 were calculated as function of the intensity of the non-uniform effect. As seen from the figure, for higher K the bias of the estimated *SL* points increases; however, the ratio between the estimated velocities is almost not affected.
(TIF)

Figure S17 Calculating *DSRC* and the correlation between $\bar{D}_2/\bar{D}_1$ and dx_2/dx_1 under the effect of unequal propagation times of harringtonine. As seen from the results, a higher non-uniform effect of harringtonine increases the *DSRC* measure and decreases the correlation between the $\bar{D}_2/\bar{D}_1$ and the dx_2/dx_1 measures.
(TIF)

Table S1 Description of the analyzed data.
(DOCX)

Table S2 Alignment results.
(DOCX)

Table S3 Estimated *SL* locations using the old estimation method. *SL* points locations were calculated for a recovery factor of 0.5 for profiles smoothed with averaging windows of different lengths (codon units).
(DOCX)

Table S4 Estimated *SL* locations using the old estimation method. *SL* points locations were calculated for different recovery factors for profiles smoothed with an averaging window of 30 codons.
(DOCX)

Table S5 Estimated *SL* locations using the new estimation method. *SL* points were calculated for a recovery factor of 0.5 for profiles smoothed with averaging windows of different lengths (codon units).
(DOCX)

Table S6 Estimated *SL* locations using the new estimation method. *SL* points were calculated for different recovery factors for profiles smoothed with an averaging window of 30 codons.
(DOCX)

Table S7 Explaining the segments' length by using various features of the coding sequence. Segments were divided into top/bottom 20%/30%/40%/50% according to their genes' tAI index/CAI index/segments' folding energy/segments' charge and were compared by using an unpaired t-test and two samples KS-test.
(DOCX)

Table S8 DSRC values and Spearman correlation between $\bar{D}_2/\bar{D}_1$ and dx_2/dx_1 for different recovery factors, when using both estimation methods. Ribosomal densities were smoothed for all profiles using a window of 30 codons.
(DOCX)

Table S9 Flux ratios (mean and median values) for a recovery factor of 0.5, using both the old and new estimation methods. Ribosomal densities were smoothed for all profiles using a window of 5–30 codons.
(DOCX)

Table S10 Sheet 1: Isoforms in the analysis of this study were selected according to the criterion presented in the original work of Ingolia *et al.*, 2011. Sheet 2: Calculated codons frequencies, based on ribosomal proteins (second column) and codons tAI value, calculation based on tRNA copy numbers (third column). Sheet 3: Selected ribosomal proteins for calculating the codon frequencies presented in the second sheet, second column.
(XLSX)

Text S1 Supplementary methods.
(DOCX)

Author Contributions

Conceived and designed the experiments: TT AD. Analyzed the data: AD TT. Wrote the paper: TT AD.

References

1. Kimchi-Sarfaty C, Oh JM, Kim I-W, Sauna ZE, Calcagno AM, et al. (2007) A "Silent" Polymorphism in the MDR1 Gene Changes Substrate Specificity. Science 315: 525–528.
2. Comeron JM (2006) Weak selection and recent mutational changes influence polymorphic synonymous mutations in humans. Proc Natl Acad Sci U S A 103: 6940–6945.
3. Gustafsson C, Govindarajan S, Minshull J (2004) Codon bias and heterologous protein expression. Trends Biotechnol 22: 346–353.
4. Drummond DA, Wilke CO (2008) Mistranslation-Induced Protein Misfolding as a Dominant Constraint on Coding-Sequence Evolution. Cell 134: 341–352.
5. Shah P, Gilchrist MA (2010) Explaining complex codon usage patterns with selection for translational efficiency, mutation bias, and genetic drift. Proc Natl Acad Sci U S A 108: 10231–10236.
6. Chamary JV, Parmley JL, Hurst LD (2006) Hearing silence: non-neutral evolution at synonymous sites in mammals. Nat Rev Genet 7: 98–108.
7. Zhang F, Saha S, Shabalina SA, Kashina A (2010) Differential Arginylation of Actin Isoforms Is Regulated by Coding Sequence-Dependent Degradation. Science 329: 1534–1537.
8. Frenkel-Morgenstern M, Danon T, Christian T, Igarashi T, Cohen L, et al. (2012) Genes adopt non-optimal codon usage to generate cell cycle-dependent oscillations in protein levels. Mol Syst Biol 8: 572.
9. Vogel C, Abreu Rde S, Ko D, Le SY, Shapiro BA, et al. (2010) Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. Mol Syst Biol 6: 1–9.
10. Bahir I, Fromer M, Prat Y, Linial M (2009) Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences. Mol Syst Biol 5: 1–14.
11. Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324: 218.
12. Ingolia NT, Lareau LF, Weissman JS (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell 147: 789–802.
13. Guo H, Ingolia NT, Weissman JS, Bartel DP (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature 466: 835–840.
14. Li GW, Oh E, Weissman JS (2012) The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. Nature 484: 538–41.
15. Brar GA, Yassour M, Friedman N, Regev A, Ingolia NT, et al. (2012) High-resolution view of the yeast meiotic program revealed by ribosome profiling. Science 335: 552–557.
16. Oh E, Becker AH, Sandikci A, Huber D, Chaba R, et al. (1295) Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. Cell 147: 1295–1308.
17. Bazzini AA, Lee MT, Giraldez AJ (2012) Ribosome Profiling Shows That miR-430 Reduces Translation Before Causing mRNA Decay in Zebrafish. Science: 336: 233–7.
18. Stadler M, Fire A (2011) Wobble base-pairing slows in vivo translation elongation in metazoans. Rna 17: 2063–2073.
19. Reid DW, Nicchitta CV (2012) Primary role for endoplasmic reticulum-bound ribosomes in cellular translation identified by ribosome profiling. J Biol Chem 287: 5518–5527.
20. dos Reis M, Savva R, Wernisch L (2004) Solving the riddle of codon usage preferences: a test for translational selection. Nucleic Acids Res 32: 5036–5044.
21. Sharp PM, Li WH (1987) The codon Adaptation Index–a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res 15: 1281–1295.
22. Tuller T, Veksler-Lublinsky I, Gazit N, Kupiec M, Ruppin E, et al. (2011) Composite Effects of Gene Determinants on the Translation Speed and Density of Ribosomes. Genome Biol 12: R110.
23. Lu J, Deutsch C (2008) Electrostatics in the Ribosomal Tunnel Modulate Chain Elongation Rates. J Mol Biol 384: 73–86.
24. Tuller T, Carmi A, Vestsigian K, Navon S, Dorfan Y, et al. (2010) An evolutionarily conserved mechanism for controlling the efficiency of protein translation. Cell 141: 344–354.
25. Trylska J, Konecny R, Tama F, Brooks CL, 3rd, McCammon JA (2004) Ribosome motions modulate electrostatic properties. Biopolymers 74: 423–431.
26. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, et al. (2002) Molecular Biology of the Cell. New York.
27. Eisenberg E, Levanon EY (2003) Human housekeeping genes are compact. Trends Genet 19: 362–365.
28. Stoletzki N, Eyre-Walker A (2007) Synonymous codon usage in Escherichia coli: selection for translational accuracy. Mol Biol Evol 24: 374–381.
29. Huang Y, Koonin EV, Lipman DJ, Przytycka TM (2009) Selection for minimization of translational frameshifting errors as a factor in the evolution of codon usage. Nucleic Acids Res 37: 6799–6810.
30. Scherer A (2009) Batch Effects and Noise in Microarray Experiments: Sources and Solutions. Chichester, UK: John Wiley & Sons.
31. Ingolia NT (2010) Genome-wide translational profiling by ribosome footprinting. Methods Enzymol 470: 119–142.
32. Wolin SL, Walter P (1988) Ribosome pausing and stacking during translation of a eukaryotic mRNA. EMBO J 7: 3559–3569.
33. Le Derout J, Regnier P, Hajnsdorf E (2002) Both temperature and medium composition regulate RNase E processing efficiency of the rpsO mRNA coding for ribosomal protein S15 of Escherichia coli. J Mol Biol 319: 341–349.
34. Hsu F, Kent WJ, Clawson H, Kuhn RM, Diekhans M, et al. (2006) The UCSC Known Genes. Bioinformatics 22: 1036–1046.
35. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10: R25.
36. Mathews DH, Sabina J, Zuker M, Turner DH (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J Mol Biol 288: 911–940.
37. Shaw LB, Zia RK, Lee KH (2003) Totally asymmetric exclusion process with extended objects: a model for protein synthesis. Phys Rev E Stat Nonlin Soft Matter Phys 68: 021910.