

RESEARCH ARTICLE

The extent of ribosome queuing in budding yeast

Alon Diamant¹, Anna Feldman¹, Elisheva Schochet², Martin Kupiec³, Yoav Arava⁴, Tamir Tuller^{1,5*}

1 Biomedical Engineering Dept., Tel Aviv University, Tel Aviv, Israel, **2** The Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv, Israel, **3** Dept. of Molecular Microbiology and Biotechnology, Tel Aviv University, Tel Aviv, Israel, **4** Biology Dept., Technion-Israel Institute of Technology, Haifa, Israel, **5** The Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel

☞ These authors contributed equally to this work.

* tamirtul@post.tau.ac.il



Abstract

Ribosome queuing is a fundamental phenomenon suggested to be related to topics such as genome evolution, synthetic biology, gene expression regulation, intracellular biophysics, and more. However, this phenomenon hasn't been quantified yet at a genomic level. Nevertheless, methodologies for studying translation (e.g. ribosome footprints) are usually calibrated to capture only single ribosome protected footprints (mRPFs) and thus limited in their ability to detect ribosome queuing. On the other hand, most of the models in the field assume and analyze a certain level of queuing. Here we present an experimental-computational approach for studying ribosome queuing based on sequencing of RNA footprints extracted from pairs of ribosomes (dRPFs) using a modified ribosome profiling protocol. We combine our approach with traditional ribosome profiling to generate a detailed profile of ribosome traffic. The data are analyzed using computational models of translation dynamics. The approach was implemented on the *Saccharomyces cerevisiae* transcriptome. Our data shows that ribosome queuing is more frequent than previously thought: the measured ratio of ribosomes within dRPFs to mRPFs is 0.2–0.35, suggesting that at least one to five translating ribosomes is in a traffic jam; these queued ribosomes cannot be captured by traditional methods. We found that specific regions are enriched with queued ribosomes, such as the 5'-end of ORFs, and regions upstream to mRPF peaks, among others. While queuing is related to higher density of ribosomes on the transcript (characteristic of highly translated genes), we report cases where traffic jams are relatively more severe in lowly expressed genes and possibly even selected for. In addition, our analysis demonstrates that higher adaptation of the coding region to the intracellular tRNA levels is associated with lower queuing levels. Our analysis also suggests that the *Saccharomyces cerevisiae* transcriptome undergoes selection for eliminating traffic jams. Thus, our proposed approach is an essential tool for high resolution analysis of ribosome traffic during mRNA translation and understanding its evolution.

OPEN ACCESS

Citation: Diamant A, Feldman A, Schochet E, Kupiec M, Arava Y, Tuller T (2018) The extent of ribosome queuing in budding yeast. *PLoS Comput Biol* 14(1): e1005951. <https://doi.org/10.1371/journal.pcbi.1005951>

Editor: John McCarthy, University of Warwick, UNITED KINGDOM

Received: September 27, 2017

Accepted: January 5, 2018

Published: January 29, 2018

Copyright: © 2018 Diamant et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper, its Supporting Information files, and GEO (accession GSE107718).

Funding: AD is grateful to the Azrieli Foundation for the award of an Azrieli Fellowship. This study was supported in part by a fellowship from the Edmond J. Safra Center for Bioinformatics at Tel Aviv University. TT and MK were supported by the Israel Cancer Research Fund (ICRF). 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.

Author summary

During translation, multiple ribosomes may translate the same mRNA. The density of ribosomal traffic across the transcript poses several open questions, such as how often a ribosome's path is blocked by a second ribosome, do queues of multiple ribosomes typically form on mRNAs and what is their effect on the overall translation rate of an mRNA. However, this phenomenon hasn't been quantified yet at a genomic level. Nevertheless, methodologies for monitoring translation are limited in their ability to detect ribosome queuing. On the other hand, most of the models in the field assume and analyze a certain level of queuing. Here we present an experimental-computational approach for studying ribosome queuing based on sequencing of RNA footprints extracted from pairs of adjacent translating ribosomes, and a computational model of translation dynamics. Our data shows that ribosome queuing in *Saccharomyces cerevisiae* is more frequent than previously thought, suggesting that at least one to five translating ribosomes is in a traffic jam; these queued ribosomes cannot be captured by traditional methods. Our analysis also suggests that the *S. cerevisiae* transcriptome undergoes selection for eliminating traffic jams, while specific regions and genes may possibly be under selection for increased queuing.

Introduction

Understanding the dynamics of protein translation is a fundamental question in biology, and has been extensively studied using experimental and computational methods in recent years [1,2]. During translation, multiple ribosomes may translate the same mRNA. The density of ribosomal traffic across the transcript poses several open questions, such as how often a ribosome's path is blocked by a second ribosome, do queues of multiple ribosomes typically form on mRNAs and what is their effect on the overall translation rate of an mRNA.

Computational and mathematical modeling of translation dynamics was first proposed in the 1960s [3–8]. A fundamental aspect of these models is ribosome queuing. Specifically, the totally asymmetric simple exclusion process (TASEP) and variants thereof have been widely used to study ribosome traffic [2,9]. This model assumes that the movement of the ribosomes is only in one direction, from the 5' end to the 3' end, a ribosome can only hop to the next position on the transcript at some known rate, and only if it is free from other ribosomes. Thus, it enables a direct study of ribosome queuing. In an early study, Zhang et al. studied the effect of clusters of slowly translated codons on the forming of queues across a synthetic transcript, as well as a set of endogenous mRNAs [10]. Mitarai et al. approximated codon translation rate using 3 categories and showed that ribosome collisions in the *E. coli lacZ* operon are more frequent than previously thought, where about 5% of ribosome translation time is wasted because of collisions along the mRNA [11]. In a later study, Mitarai and Pedersen compared the ribosome traffic on 87 wild-type sequences and sequences where the synonymous codons were swapped randomly, and showed that wild-type genes tend to reduce ribosome collisions [12]. Chu et al. performed two studies, accompanied by experiments expressing CFLuc, that analyzed ribosome traffic under different codon sequence variants, and demonstrated the effect of codon usage and queuing of ribosomes in the 5'-end region on translation initiation [13,14]. In recent years, dozens of additional studies based on variants of the TASEP model have been employed to answer fundamental questions in molecular evolution, functional genomics, synthetic biology, and more [2]. However, performing an accurate transcriptome-wide assessment of ribosome queuing remains a challenge due to the need for careful calibration of the model [2].

A number of studies in recent years have suggested that translation is rate-limited by the initiation rate of a gene [15–17]. In this regime, traffic jams are less likely to form due to the low density of ribosomes on the mRNA, and are expected to have a weaker relation with codon usage [18]. However, an ongoing debate remains over the relative contribution of different factors to the control over translation rate, which is also related to the rate limiting nature of initiation vs. elongation [14,19–22].

In recent years, ribosome profiling (Ribo-seq) has been the state-of-the-art technology for monitoring the transcriptome-wide distribution of translating ribosomes at high-resolution [23]. Ribosome profiling is based on deep-sequencing of ribosome-protected mRNA fragments (RPFs) from living cells, such that the sequence of each fragment indicates the position (footprint) of a translating ribosome on the transcript. Ribo-seq has been used to detect positions with significantly enriched reads. Strong peaks in the reads density have been interpreted as ribosome pausing sites and were linked to various possible causes [24–32]. If indeed ribosome pausing is common, we expect queues to form around such pauses.

It should be noted that typically, Ribo-seq protocols filter fragments that are longer than a single ribosome (approximately 30bp), with few exceptions [33]. Thus, adjacent ribosomes and regions where pairs (or more) of ribosomes are frequent, are expected to be under-represented in most Ribo-seq datasets. This may limit the analysis of ribosome stalling and queuing if indeed a significant portion of ribosomes are paired. However, there are currently no comprehensive estimates of the extent of ribosome queuing. In this study, we provide for the first time, quantitative estimates for the portion of pairs of ribosomes and analyze the dynamics of ribosome queuing using experimental and simulated data.

Results

Estimating the fraction of paired ribosome footprints

The probability that two ribosomes are adjacent to one another is affected by a long list of variables such as the number of ribosomes in the cell, the length and number of translated mRNAs, the initiation rates, the distribution of elongation rates along the mRNA, etc. Thus, it is not trivial to estimate this value without a direct experiment.

To estimate the quantity of ribosomes that are adjacent to one another in the transcriptome—which is related to the extent of queuing / traffic jams on transcripts—the first steps of the ribosome profiling protocol in yeast were performed as previously described [23,34] (details in the Materials and Methods section). Ribosome protected footprints (RPFs) of various sizes were generated, followed by sucrose gradient fractionation. The resulting profiles were analyzed, along with previously reported footprint gradients [23,33,35], by employing Gaussian mixture to model the different components within the profile (Fig 1, details in the Materials and Methods section). Previous studies focused mostly on isolating the RPF fraction with footprints protected by a single ribosome (referred to here as mono-RPFs), with some exceptions [33]. Here, we estimated the size of two fractions: mono-RPFs (mRPFs) and di-RPFs (dRPFs), the latter containing footprints protected by two ribosomes (a disome). The area of each fraction in the profile is proportional to the number of ribosomes in each fraction. Thus, the ratio between the dRPFs fraction and the mRPFs fraction can provide an estimation for the ratio between the number of ribosomes that are directly adjacent to one another and the number of isolated ribosomes on the transcript. The ratio obtained ranged from 0.196 to 0.352 in the four datasets, suggesting that at least 1 in 6 ribosomes is adjacent to another and form a pair. When considering the existence of larger clusters of ribosomes, however, in smaller quantities, the probability of observing a pair of adjacent ribosomes is expected to be

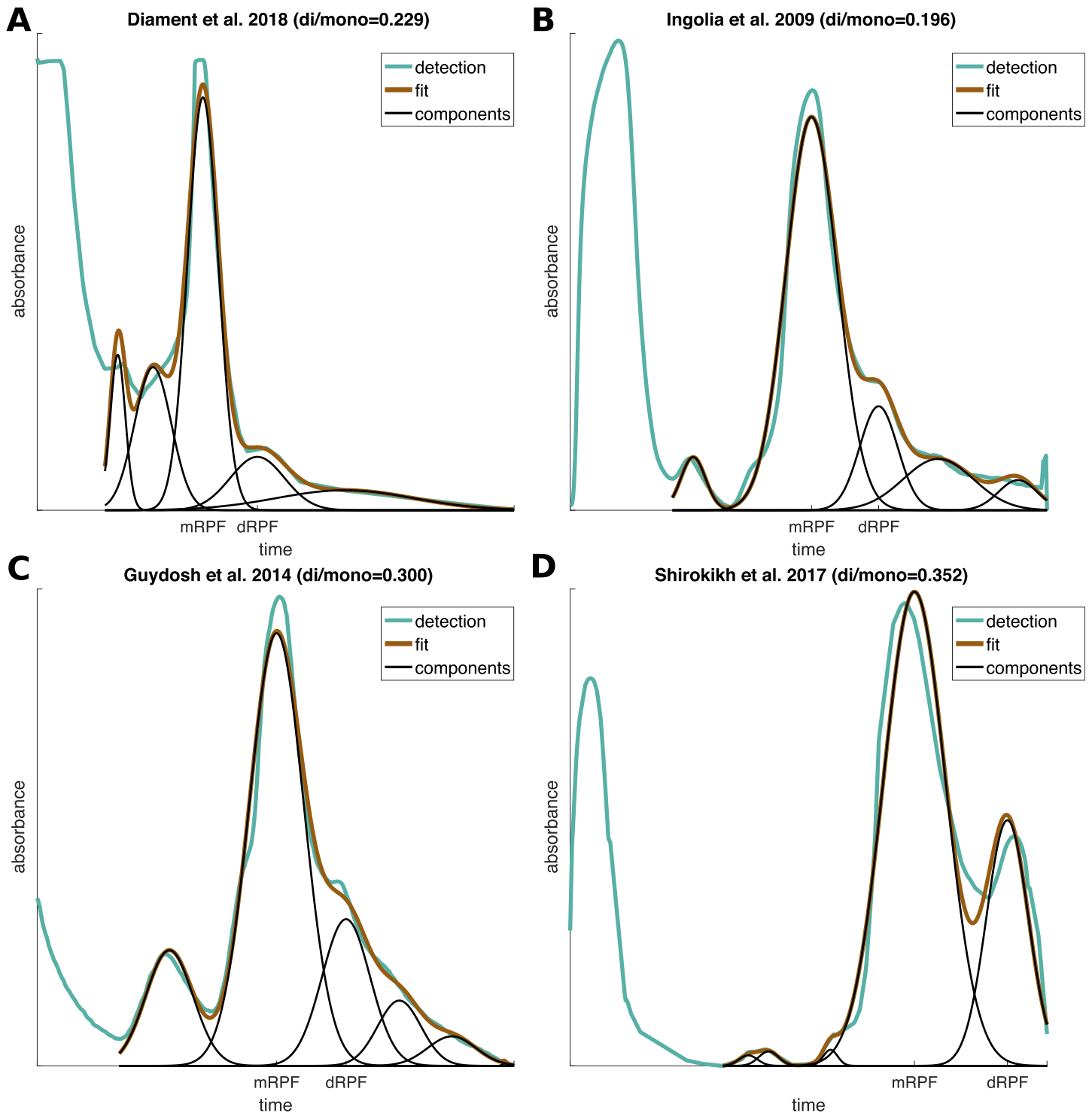


Fig 1. Sucrose gradient fractions. (A) Quantification of ribosome protected footprints (RPFs) according to size. The plot shows the estimated components underlying the observed distribution, by means of Gaussian mixture. Each component reflects a set of footprints with a typical size, originating from a single ribosome (mono-RPF, mRPF), a pair of ribosomes (di-RPF, dRPF), etc. The ratio of dRPFs to mRPFs is reported in the caption. (B) Same for Ingolia 2009 data, profile reproduced from [23]. (C) Same for Guydosh 2014 data, profile reproduced from [33] (Fig 1 in the original paper). (D) Same for Shirokikh 2017 data, profile reproduced from [35] (Figure 5 in the original paper). See also S1 Fig.

<https://doi.org/10.1371/journal.pcbi.1005951.g001>

even higher, at least 1 in 5 as justified by the computational model in the next sections. We obtained similar results using different methods (S1 Fig).

The variance between different studies may reflect the condition in which the experiment was performed, such as the type and amount of RNase used (which may cut dRPFs at some rate and relocate them to the mRPF fraction), but possibly also additional conditions that affect translational efficiency in cells. The high fraction of dRPFs detected in all the experiments, however, suggests that ribosome queuing is more prominent and widespread than typically assumed, and that the distribution and determinants of traffic jams in the transcriptome are important fundamental questions. This may also suggest that due to the large extent of queuing, previous results that attempted to answer these questions by sequencing only mRPFs were based on limited data.

Measurement of queues

In order to study the locations where queues form, we sequenced both fractions of mRPFs and dRPFs (details in the Materials and Methods section). Next, the distribution of mRPFs and dRPFs in the transcriptome was analyzed using our data and Guydosh and Green's dataset [33]. RPKM (reads per kilobase per million) values of genes according to both RPF types showed a strong correlation between the two ($r = 0.88$, $p < 10^{-307}$, $n = 6,664$; Spearman's rho and its asymptotic p-value; Fig 2A), however still lower than typically observed between replicates of the mRPF fraction [36]. We expect differences between the distribution of mRPFs and dRPFs to reflect the location of ribosomal traffic jams.

Meta-gene profiles of normalized footprint counts (NFC, i.e. read counts divided by the transcript's mean read count) were calculated for mRPFs and dRPFs (Fig 2, S2 Fig and S3 Fig for the separate datasets). When examining the 5'-end of genes (Fig 2B), an enrichment of dRPFs compared to mRPFs was observed. In addition, it can be seen that the profile of dRPFs decreases into the gene. Genes were divided into two sets of highly and lowly expressed (top / bottom 50% according to their protein abundance). In order to control for the number of reads in the sets, the highly expressed genes were downsampled to have the same total number of reads as the lowly expressed genes. We found that both mRPF and dRPF profiles had higher normalized counts in the set of highly expressed genes compared with lowly expressed genes ($p = 3 \times 10^{-13}$ / 7×10^{-11} , respectively, rank-sum p-value). Both mRPFs and dRPFs had a median normalized count significantly higher than the background (1.24–1.35-fold for mRPFs and 1.58–1.61-fold for dRPFs, $p < 0.01$, empirical p-value based on random sampling of positions within the same set of genes). This may imply that ribosome queuing is frequent in the 5'-end region, and specifically in highly expressed genes.

Peak calling was performed in order to detect regions with significant ribosome pausing, and study its association with the traffic around the peak. Peaks were defined as positions that are at least 5 standard deviations above the ORF's mean (ignoring positions empty of reads). In the vicinity of detected peaks in the mRPF profiles a dRPF peak can be observed, distanced 9 codons upstream (Fig 2C). These di-footprints mark the position of two ribosomes—a ribosome positioned 9 codons upstream to the mRPF peak, and a second one approximately where the mRPF was detected. Unlike in the 5'-end region, mRPF peaks were more extreme in lowly expressed genes ($p = 9 \times 10^{-52}$, 3.44-fold above background compared with 1.92 in top genes). Similarly, the associated dRPF peaks were mostly related to lowly expressed genes ($p = 3 \times 10^{-8}$, 2.05-fold above background compared with 1.28). Correspondingly, in the vicinity of peaks detected in dRPF profiles an mRPF peak was detected 9 codons downstream (Fig 2D). Furthermore, the dRPF peaks were more extreme in lowly expressed genes ($p = 2 \times 10^{-17}$, 6.66-fold above background compared with 4.49), and the associated mRPF reads were

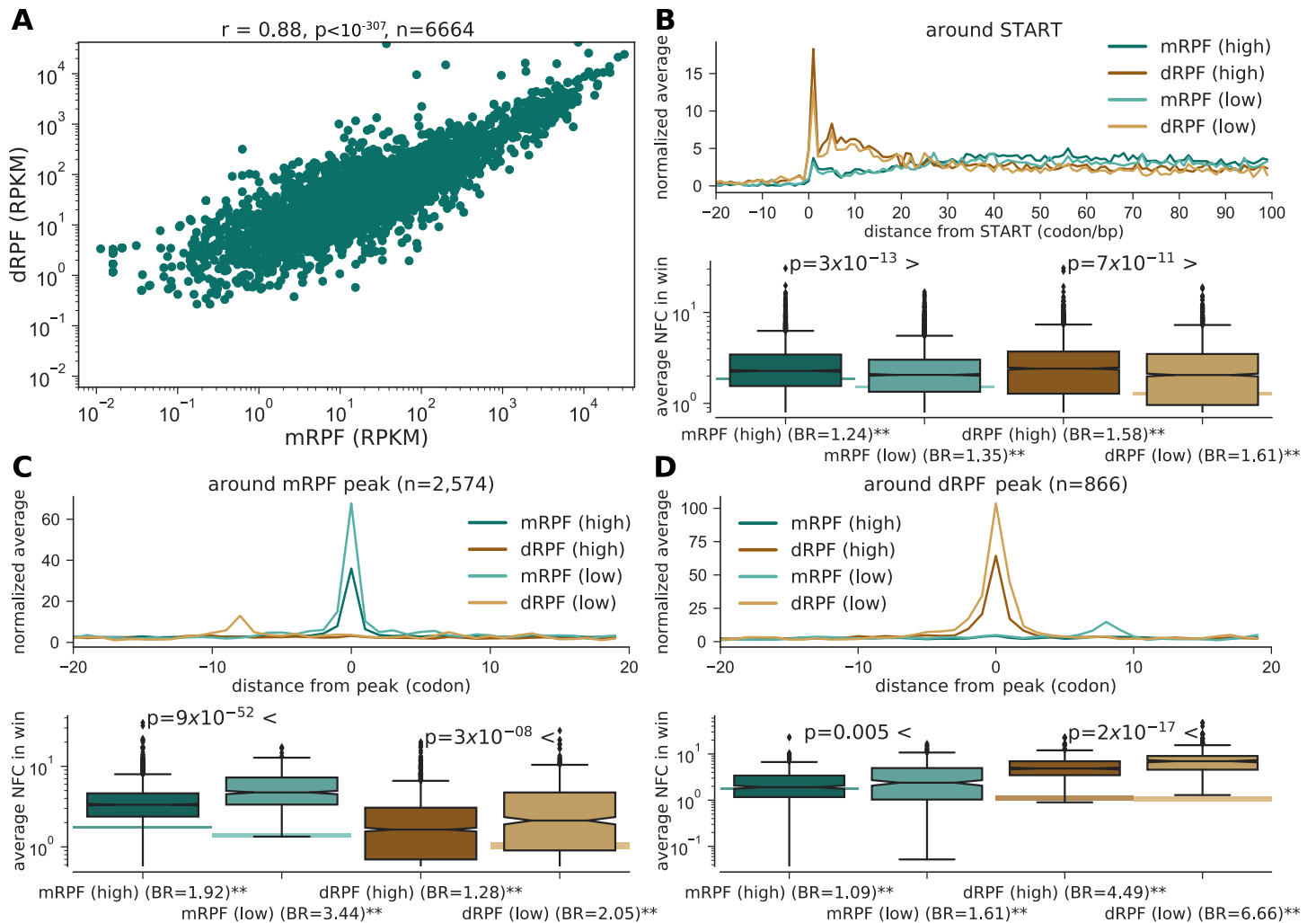


Fig 2. Ribosome profiling. (A) mRPFs vs. dRPFs, RPKM values for each gene. Spearman's rho and an asymptotic p-value are reported in the caption. (B) *Top*: Meta-gene analysis around the beginning of the ORF for mRPFs and dRPFs, plotted for highly and lowly expressed genes. Read counts were normalized per gene according to gene average (NFC), and the average across genes per position is shown. *Bottom*: The distribution for all genes of the mean NFC across the window in the top panel. P-values according to rank-sum test compare highly and lowly expressed genes. The background bands show the resulting medians when sampling random positions across the same genes (95% of the sampled medians fall within this range). Background ratio (BR) of the observed median vs. null is reported below (** denotes empirical p-value < 0.01). (C) Same, around peaks detected in mRPF profiles. (D) Same, around peaks detected in dRPF profiles. See also S2 Fig, S3 Fig.

<https://doi.org/10.1371/journal.pcbi.1005951.g002>

enriched in lowly expressed genes ($p = 0.005$, 1.61-fold above background compared with 1.09). One hypothesis that may explain this result is that when traffic jams form in lowly expressed genes, they tend to be more severe. Traffic jams are expected to form in regions with high ribosomal density—for example, in highly expressed genes with high initiation and translation rates. On the other hands, highly expressed genes are under selective pressure to increase translation efficiency and to reduce traffic jams, for example by developing a ramp in their 5'-end [37,38]. Due to this, lowly expressed genes may still contribute significantly to the observed queues despite being less dense.

Whenever ribosomes form a queue, the traditional ribosome profiling protocol (which sequences mRPFs) is expected to fail to fully capture the present ribosomal density. We tried to estimate the percentage of queued positions and locations of such queues by detecting missing mRPFs in the profiles. The following rule was employed: a position had a significant

number of mRPFs missing if the number of mRPFs in that position was in the bottom 10% of that gene, while the number of dRPFs in that position was in the top 10% of that gene. For each gene with at least 10% dRPF reads coverage, the percentage of positions with missing mRPFs was calculated according to this rule (median: 9%). In addition, the calculation was repeated for 3 regions of the gene: the first 100 codons, the last 100 codons and the rest of the gene (Fig 3A). The percentage of positions with missing mRPFs was significantly higher in the first / last regions ($p = 4 \times 10^{-5} / 0.03$, respectively; rank-sum p-value), compared with the mid-region, suggesting that these regions tend to contain more traffic jams.

Finally, the correlation between the percentage of positions with missing mRPFs in a gene and various features related to its expression and translational efficiency was computed, while controlling for dependencies between the features using partial Spearman correlations. Significant negative correlations were observed with the gene's adaptation of the coding sequence to the tRNA pool as measured via tAI [39] ($r = -0.21$, $p = 1.4 \times 10^{-6}$), and its protein abundance (PA, $r = -0.39$, $p = 3 \times 10^{-55}$). Significant positive correlations were observed with the initiation rate of the gene ($r = 0.10$, $p = 10^{-4}$), the PA per RNA copy ($r = 0.33$, $p = 8 \times 10^{-39}$), and ORF length ($r = 0.15$, $p = 2 \times 10^{-9}$; Fig 3B). Initiation rates for the aforementioned test were inferred using a simulation of ribosome traffic, as discussed in the next section and the Materials and Methods, with ribosomal densities and RNA copies estimated from Ribo-seq and RNA-seq data. Similar results were observed using different thresholds for detection of missing mRPF and appear in S4 Fig. The test was repeated for the dRPF-to-mRPF ratio with similar results (Fig 3B). We further discuss these results below.

Translation simulation

Simulating ribosome movement enables a study of traffic on the mRNA in greater detail. A simulation of translation in 5,450 yeast genes was performed based on the totally asymmetric simple exclusion process (TASEP) model, with parameters calibrated using Arava et al.'s experimental measurements of the number of ribosomes on each transcript as well as RNA copy numbers [40] (details in the Material and Methods section, simulation results reported in S1 Table). Translation initiation rates of genes were fitted to achieve the measured ribosomal density. In addition, RNA and ribosome numbers were scaled globally to achieve the total quantities of transcripts and translating ribosomes that are generally accepted in the literature (details in the Materials and Methods section). The resulting initiation rates (median: 0.09 s^{-1} , with rates ranging from 0.019 s^{-1} in the 10th percentile to 0.29 s^{-1} in the 90th percentile) were in good agreement with previous results ($r = 0.70$, $p < 10^{-307}$, $n = 5,162$; asymptotic p-value for Spearman's rho) [41]. The correlation between the resulting translation rates and PA per RNA copy was moderate ($r = 0.32$, $p = 2.2 \times 10^{-16}$, $n = 5,132$), and similar to their correlation with the ratio of Ribo-seq RPKM to RNA-seq RPKM ($r = 0.37$, $p = 1.3 \times 10^{-170}$, $n = 5,384$). To test the simulated profiles at the codon level, density profiles of the 1,000 densest genes in a high-resolution ribosome profiling dataset were utilized [42]. A median correlation of 0.10 was observed between simulated and measured ribosome density profiles of genes, smoothed using a 10-codon wide sliding window. This value is comparable to the correlation between two mRPF profiles from different experiments ($r = 0.16$ for Brar et al. vs. Gudyosh et al.; this is similar to previous reports [36]).

We generated a simulated distribution of the fractions of RPFs in the simulation (ribosomes were taken to be 10 codons long), counting the number of isolated, pairs, triplets of ribosomes etc. throughout the transcriptome (S4 Fig). First, only directly adjacent ribosomes were considered, with no codons between them, resulting in a dRPF-to-mRPF ratio of 0.11, which is lower than reported in the section above. However, when considering simulated RPFs

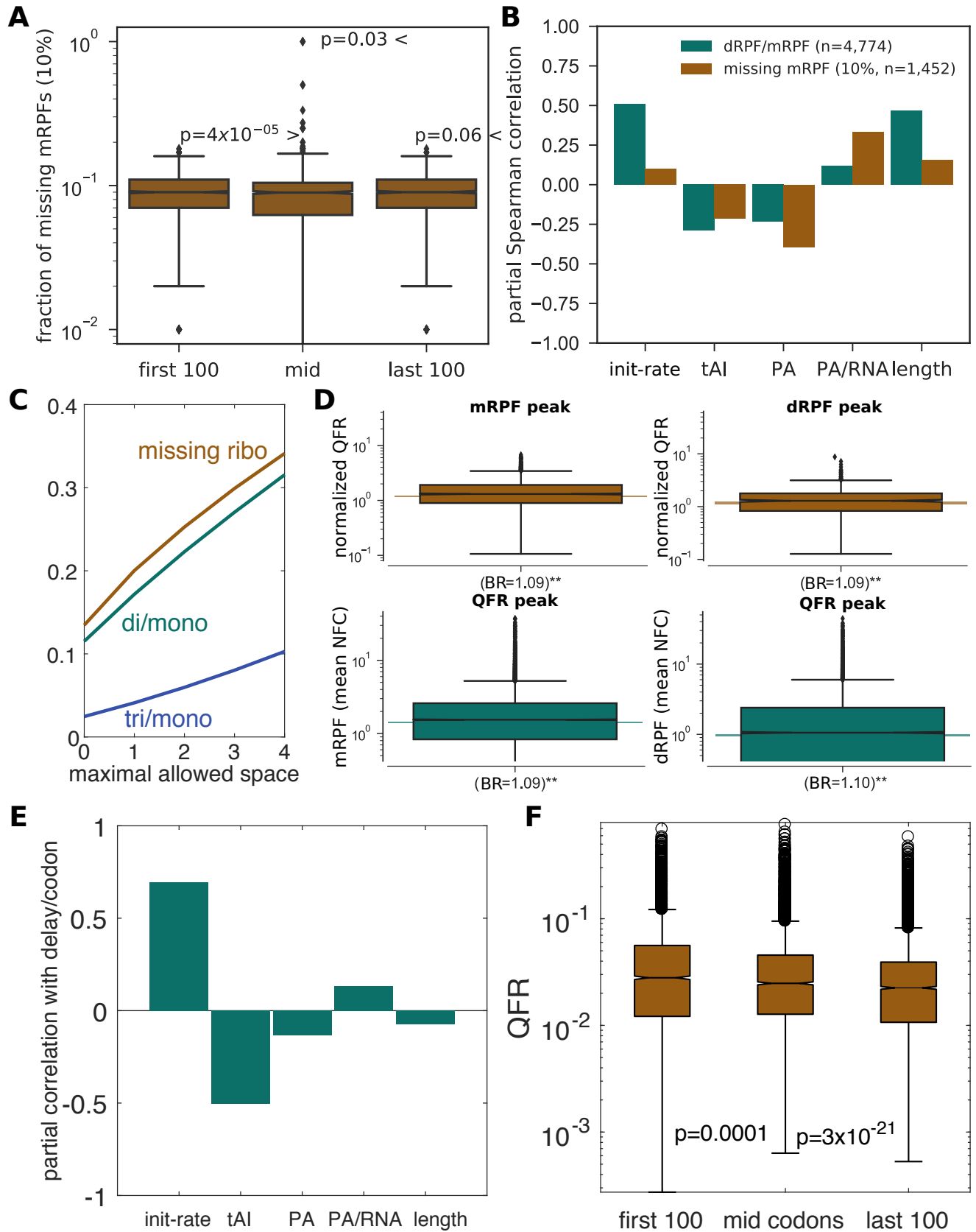


Fig 3. Queue analyses. (A) Fraction of positions with detected missing mRPFs in genes, divided into 3 regions: first 100 codons, last 100 codons and the rest of the gene. mRPF profiles were downsampled to have the same number of reads as the dRPF profiles. Rank-sum p-values reported. (B) Partial Spearman correlations between the dRPF-to-mRPF ratio per gene, and the fraction of missing mRPFs (across the entire ORF) vs. various features (initiation rate, tRNA adaptation index, protein abundance-PA, PA per RNA copy and ORF length), while controlling for their dependencies. mRPF profiles were downsampled to have the same number of reads as the dRPF profiles. *n* indicates the number of genes, and the threshold for missing mRPF detection is indicated in percentage. See also S4 Fig. (C) The ratio between simulated dRPFs and mRPFs, as well as the estimated fraction of missing ribosomes when sequencing mRPFs, for different values of the maximal allowed space between ribosomes on a footprint (see also S4 Fig). (D) *Top*: Enriched simulated queued ribosomes in windows surrounding measured mRPF and dRPF peaks (empirical p-value < 0.01). Y-axis shows the average QFR (normalized similarly to NFC). *Bottom*: Enriched mRPFs and dRPFs in windows surrounding peaks in simulated profiled of the fraction of queued ribosomes per position (empirical p-value < 0.01). (E) Partial Spearman correlations between ribosome delay per codon of a gene, and various features while controlling for their dependencies. (F) Distribution of the queued fraction of ribosomes (QFR) in 3 regions of the gene. Sign-rank p-value between box-plots.

<https://doi.org/10.1371/journal.pcbi.1005951.g003>

comprising of ribosomes up to 1–4 codons apart on the footprint, the observed ratio increased to 0.17–0.32 –results that are in good agreement with the ones above (Fig 3C). Thus, it is possible that some of the RPFs that are typically collected in a ribosome profiling experiment contain ribosomes with 1–4 free codons between them. Other parameters that may affect this ratio are the growth conditions, and the existence and frequency of ribosome pausing, which was not simulated here [25]. Similarly, the simulation was utilized to estimate the size of the fraction of tri-RPFs (containing 3 ribosomes), which was 3- to 5-times smaller than the dRPF fraction, suggesting that dRPFs represent the majority of queued ribosomes (Fig 3C, S4 Fig). Moreover, the model enables to extrapolate and estimate the fraction of total ribosomes that are not isolated since they are adjacent an additional ribosome, and therefore are unlikely to be detected via mRPF sequencing. We estimate that this fraction ranges from 14% (0 codon separation between neighboring ribosomes) to 34% (up to 4 codon separation) (Fig 3C).

Simulated queue properties

We analyzed the distribution across the transcriptome of queued ribosomes and their waiting times in the simulation, that is, ribosomes that cannot complete an elongation cycle due to the next codon being covered by another ribosome. The probability of observing a queue in a gene was 0.21 (median over genes, with probabilities ranging from 0.024 in the 10th percentile to 0.58 in the 90th percentile). Typically, 3.67% of the translating ribosomes on an mRNA were delayed due to traffic jams (median over genes, with the 10th and 90th percentiles being 0.9% and 14.2%), and 5.9% of all translating ribosomes in the cell were delayed at any given time. This fraction is related to the number of observed dRPFs in the previous section, however it is expected to be smaller than the latter since some of the ribosomes that are within a footprint's range from one another may not interfere with another ribosome's progress at all times (only at time intervals when the two are directly adjacent and the downstream ribosome is elongating more slowly than the upstream ribosome). The total delay time of a ribosome during the synthesis of a complete gene was 3.5s (median over genes, with delay times ranging from 1.07s in the 10th percentile to 11.65s in the 90th percentile).

Next, the meta-gene analyses in the above section were repeated for simulated profiles. In windows surrounding observed Ribo-seq peaks (the ones detected in Fig 2), a significant enrichment was observed in the queued fraction of (simulated) ribosomes (QFR) (1.09-fold compared with background around mRPF and dRPF peaks, Fig 3D). Similarly, peaks were detected in the profiles of the QFR values per position, with significant enrichment of mRPF and dRPF reads observed in their vicinity (Fig 3D). Thus, the simulated traffic has a significant relation with the measured RPFs. Differences between the experimental data and simulation can be partially explained by the high level of inter-experiment variance, and by the fact that our model does not include long translational pauses, which may affect a significant portion of ribosomal traffic [25].

We tested the relation between ribosome delay times per codon and various sequence/expression-related features of the gene, including the inferred initiation rate, tRNA Adaptation Index (tAI) score, protein abundance (PA), PA per RNA copy, and ORF length. Partial Spearman correlations were computed between ribosome delay times per codon and each feature given the other 4 features (Fig 3E). A strong positive correlation with initiation rates ($r = 0.69$; $p < 10^{-307}$; $n = 5,080$; asymptotic p-value) and a strong negative correlation with tAI scores ($r = -0.50$, $p < 10^{-307}$, $n = 5,080$) were observed, implying that genes that contain codons recognized by less abundant tRNA species tend to exhibit ribosome congestion and traffic jams across their sequence. These results are similar to the ones reported above based on Ribo-seq data (Fig 3B). Furthermore, weak yet significant correlations were observed with PA ($r = -0.13$; $p = 1.4 \times 10^{-21}$; $n = 5080$), PA/RNA ($r = 0.13$; $p = 3.2 \times 10^{-32}$; $n = 5,080$) and gene length ($r = -0.07$; $p = 10^{-7}$; $n = 5,080$). Next, QFR was computed in 3 different regions of the ORF—the first 100 codons (median: 2.8%), the last 100 codons (2.25%) and the rest of the sequence (2.48%)—for 4,204 genes with appropriate length (Fig 3F). On average, the percentage of queued ribosomes in the first region was 1.9/2.1-fold higher in the first region compared with the middle/last region (respectively) in that gene; this is with agreement with the 5'-end enrichment observed in Ribo-seq data.

The simulation was repeated for 100 random genomes where the order of synonymous codons within each gene was permuted, as well as for 100 random genomes where the global rates of codons were shuffled (without altering the gene sequence, details in the Material and Methods section), and employed them to calculate two empirical p-values: p_1 controlling for codon order, and p_2 controlling for the general decoding rates of the different codons. We found that traffic jams were less prevalent in the real genome according to the various properties discussed above, including: ribosome delay time ($p_1 < 0.01$, $p_2 < 0.01$), queue probability ($p_1 < 0.01$, $p_2 = 0.01$) and QFR ($p_1 < 0.01$, $p_2 < 0.01$), as well as the ratio of simulated dRPF to mRPF ($p_1 < 0.01$, $p_2 < 0.01$). Interestingly, the percentage of queued ribosomes in the first 100 codons was higher than in permuted sequences ($p_1 < 0.01$), while in other regions of the gene it was lower than in permutations ($p_1 < 0.01$). This is in agreement with previous suggestions that selection acts on codon order to decrease congestion downstream by delaying ribosomes when entering the ORF [38].

P_2 values were also utilized to test the simulation calibration. The simulation results were in better agreement with experimental data for the selected set of codon decoding rates than for random assignments of the same 61 rates. For example, the correlation between simulated and measured footprint density profiles was significantly higher for the selected parameters ($p_2 < 0.01$), as was the correlation between simulated translation rates and PA/RNA ($p_2 = 0.01$), and the correlation between simulated translation rates and Ribo-seq/RNA-seq RPKM ratio ($p_2 = 0.01$).

Functional gene groups with extreme traffic conditions

We studied gene sets that exhibited extreme traffic conditions, either with respect to the rest of the genes or with respect to randomized variants of these genes. The resulting z-scores from comparing an observed property of a gene (e.g., its translation rate) with the distribution of that property in variants with shuffled synonymous codons and the same ribosomal load on the transcript can be regarded as an optimality score for the sequence (details in the Materials and Methods section). The set of genes with the highest translation rate compared to its respective random set (top 5%) has a significant subset of common genes with the set of genes with the lowest fraction of queued ribosomes compared to random (bottom 5%, $p = 5 \times 10^{-50}$, hypergeometric p-value), and vice versa ($p = 10^{-88}$, Fig 4A). This is expected, as genes with an

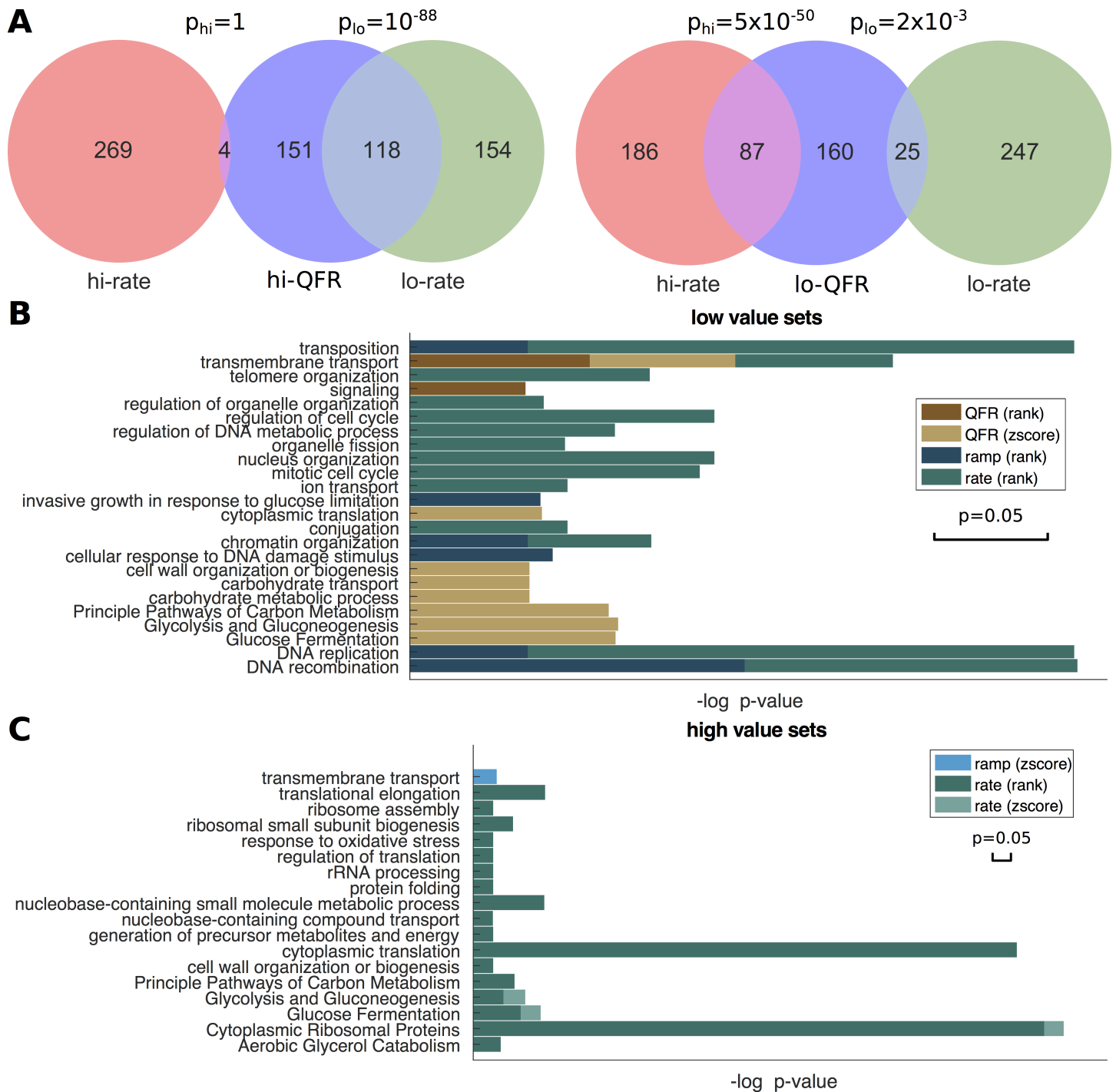


Fig 4. Functional enrichment. (A) Sets of genes with extreme properties relative to randomizations of their synonymous codon order. *Left:* Intersection between the set of genes with the highest QFR compared to random, and genes with the highest/lowest translation rate compared to random. *Right:* The same for the lowest QFR compared to random. (B)–(C) Functional enrichment in different gene sets with extreme properties, according to their rank among genes, or according to their z-scores vs. codon randomizations. Each of the stacked bars represents the enrichment p-value of the corresponding term (longer bars for smaller p-values) in the gene set (denoted by color, see legend). The ramp score was defined by the ratio of QFR in the first 100 codons and the middle of the gene (excluding the first/last 100 codons). The scale of the 0.05-threshold is shown below the legend.

<https://doi.org/10.1371/journal.pcbi.1005951.g004>

optimized codon order (and reduced traffic jams) can reach a higher translation rate while occupying the same average number of ribosomes (see for example [37]). In addition, a significant negative correlation was observed between the z-scores of QFR and the inferred initiation rates of genes reported in the sections above ($r = -0.13$, $p = 6.2 \times 10^{-21}$, $n = 5,450$). This result is in agreement with the hypothesis that highly expressed genes are expected in many cases to be under selective pressure to optimize their sequence as to minimize queuing (see for example [37]).

Functional enrichment was performed for the above sets of genes (Fig 4B and 4C, sets appear in S2 Table), detecting a number of biological processes and pathways over-represented in genes with a low QFR z-scores, such as cytoplasmic translation ($p = 0.0305$, hypergeometric p-value adjusted for FDR, light brown in Fig 4B), and glucose fermentation ($p = 0.004$, light brown in Fig 4B). The largest subset of QFR-optimized genes was associated with transmembrane transport ($p = 0.0215$, light brown in Fig 4B). Interestingly, the same process was also enriched in genes with high z-scores of dense queuing in the first 100 codons compared with the middle codons of the genes ($p = 0.0216$, light blue in Fig 4C). Many of these genes are not highly expressed, as discussed below, and may be selected for these properties for reasons other than translational efficiency, such as co-translational folding [43].

Next, functional enrichment was performed for sets of highly and lowly ranked genes, according to their properties (sets appear in S2 Table). A number of pathways were enriched both in highly translated genes (top 5%, dark green in Fig 4C), as well as in rate-optimized genes (top 5% of z-scores, light green in Fig 4C) according to translation rate, namely cytoplasmic ribosomal proteins ($p_{\text{rank}} = 2.8 \times 10^{-41}$, $p_{\text{zs}} = 0.0422$ for the highly ranked and highly optimized sets, respectively), glycolysis ($p_{\text{rank}} = 7.1 \times 10^{-3}$, $p_{\text{zs}} = 0.0285$) and glucose fermentation ($p_{\text{rank}} = 4.2 \times 10^{-4}$, $p_{\text{zs}} = 0.0382$). A number of pathways were enriched in highly translated genes (dark green in Fig 4C), as well as in QFR-optimized genes (light brown in Fig 4B), namely cytoplasmic translation ($p_{\text{rank}} = 2.5 \times 10^{-39}$, $p_{\text{zs}} = 0.0305$), carbon metabolism ($p_{\text{rank}} = 1.2 \times 10^{-3}$, $p_{\text{zs}} = 5.2 \times 10^{-3}$), cell wall organization ($p_{\text{rank}} = 0.0385$, $p_{\text{zs}} = 0.0424$), glycolysis ($p_{\text{zs}} = 4.1 \times 10^{-3}$) and glucose fermentation ($p_{\text{zs}} = 4.4 \times 10^{-3}$). This is expected, as highly translating genes also tend to have higher ribosomal densities and therefore their codon usage may be under selection to decrease traffic jams.

Discussion

Ribosome traffic jam is a phenomenon that has been mentioned many times in the context of genome evolution, biophysics of translation, translation modeling, biotechnology, and more. To the best of our knowledge, this study provides the first comprehensive experimental quantification of traffic jams during translation. We estimated that at least 20% of the ribosomes in the budding yeast *S. cerevisiae* are positioned close enough to another ribosome on the transcript to be collected as a single footprint in a ribosome profiling experiment. This is a direct indication for the level of queuing in transcripts. Our estimates using sucrose gradient analyses have been consistent across experiments, and with simulated ribosome traffic based on empirical parameters. The observed variance in results obtained from different datasets may have been related to the experiment's particular conditions and protocol. The current experimental resolution limits the accuracy of analysis to footprints containing a single (mRPF) or a pair (dRPF) of ribosomes, but future Ribo-seq experiments may enable the analysis of larger footprints.

The results reported here suggest that queuing is not a negligible phenomenon. It appears to have important effects on organism evolution and fitness. Since translation is the process that consumes most of the energy in the cell [44], an increase of x percentage in ribosome

density should be related to a similar increase in growth rate or fitness (as it is directly related to the energy consumed by the cell). It is clear that even a change of 1% or even 0.1% (as was estimated in Fig 3D) in the fitness of a micro-organism should have a very strong effect on its evolution (for example, in the case of *S. cerevisiae*, an organism with a doubling time of 90min, a mutant with 0.1% increase in fitness is expected to practically be 97% of the population after 8 months). Moreover, increase in traffic jams may have an even stronger impact when considering its potential effect on ribosome abortion, that can be related again to energy waste, as well as to the production of nonfunctional and even toxic truncated proteins. Thus, from a biological/evolutionary point of view these values are potentially significant. We believe that further comprehensive experimental and computational efforts should be spent in the future to better understand and quantify the effect of local increase/decrease in ribosome densities/queues on organism fitness.

Due to these implications, ribosome queuing should be considered when designing Ribo-seq experiments (or using any other technology for studying translation), when analyzing mRNA translation, or when modeling mRNA translation. Specifically, the results reported here support the usage of the dozens of TASEP based models in studies that have been published in recent years [2,9,45]. These results also support the idea that the genomes of various organisms are under selection for generally minimizing queuing (see, for example, [10,38,41,46]).

Our approach has enabled the detection of transcript properties that are associated with high/low levels of queuing; for example, we showed that higher initiation may tend to increase queuing while higher adaptation to the tRNA pool may tend to decrease queuing. Thus, our methodology can be used for developing better models of translation and for designing synthetic genes with increased/decreased queuing levels.

In summary, we believe that the computational-experimental approach demonstrated here will enable the analysis of translation at a higher resolution. It is essential to perform similar studies in various organisms and in different conditions to understand the significance of queuing across the tree of life. In the current study, we focused on translation under standard growth medium for yeast. It is important to emphasize that it is generally believed that under stress condition there are more traffic jams due to non-typical expression of tRNA species [47]; thus, it is possible that the estimates reported here serve as a lower bound for the ribosome traffic levels observed in the environment and evolution of yeast and other micro-organisms.

Materials and methods

Ribosome profiling

Ribosome profiling was performed as previously described [23,34], with minor modifications. rRNA was depleted using Ribo-Zero by EpiCentre. The small RNA sequencing kit from New England BioLabs [48] was utilized. Sucrose gradient fractionation was employed for filtering reads containing one or two ribosomes. Additional filtering of DNA was performed with Blue-Pippin to select the relevant length prior to sequencing.

Ribosome profiling analyses

A Gaussian mixture model was fitted to the sucrose gradient profiles of ribosome footprints. To this end, an iterative expectation-maximization algorithm was used (implemented in MATLAB's `fitgmdist` function). As input to the algorithm, 500 equally spaced points were sampled from the X-axis of the profile (which is related to particle size), with their representation in the input set (the number of times each point appears) proportional to their absorbance value (the point of maximum absorbance arbitrarily set to appear 100 times in the input set).

Model fitting was run multiple times using a varying number of components (4 to 8), using random seeding of the algorithm, as well as seeding with components centered around peaks in the profile (detected using MATLAB's findpeaks), until reaching an optimal solution. The quality of models was assessed based on the Bayesian information criterion and manual inspection of candidate solutions with scores distanced at most 0.5% from the optimum.

Transcript sequences of *Saccharomyces cerevisiae* were obtained from Ensembl [49] (R64 release 87), and UTR lengths were based on previously reported major transcript isoforms (mTIF, selecting the longest one available for each transcript, and using a default length of 500bp for missing UTRs) [50]. Adaptors (AGATCGGAAGAGCACACGTCT in our dataset) were trimmed from reads using Cutadapt [51] (version 1.12). Bowtie [52] (version 1.2) was employed to map them to the transcriptome. In the first phase, Ribo-seq reads that mapped to rRNA and tRNA sequences were discarded using Bowtie parameters '-n 2--seedlen 21 -k 1--norc'. In the second phase, the remaining Ribo-seq reads, as well as RNA-seq reads, were mapped to the transcriptome with Bowtie parameters '-v 2 -a--strata--best--norc -m 200'. When the 3' adaptor contained polyA alignments were extended to their maximal length by comparing the polyA with the aligned transcript until reaching the maximal allowed error (2 mismatches). Unique alignments were first assigned to the ribosome occupancy profiles. For multiple alignments, the best alignments in terms of number of mismatches were kept. Then, multiple aligned reads were distributed between locations according to the distribution of unique ribosomal reads in the respective surrounding regions. To this end, a 100nt window was used to compute the read count density RCD_i of unique reads in vicinity of the M multiple aligned positions in the transcriptome, and the fraction of a read assigned to each position was $RCD_i / \sum_{j=1}^M RCD_j$. The location of the A-site (the site of the upstream ribosome in the case of dRPFs) was estimated for each read length by performing meta-gene analysis of the distribution of 5'-end of reads upstream to the start codon of ORFs. If a global maximum was detected in the range of 5-20bp upstream to the start codon (no larger peak observed down-/upstream from this range), this distance was selected as the shift to the P-site for this read length, otherwise these reads were discarded. For 2 datasets where peaks were hard to detect, a default shift of 15bp was employed. Mapping statistics and accession numbers for all datasets used appear in S3 Table and S5 Fig. An aggregated dataset of our experimental data and Guydosh et al.'s was analyzed in Figs 2 and 3, and each dataset was analyzed independently in S2 Fig and S3 Fig. Protein abundance data for dividing genes into highly / lowly expressed sets was obtained from PaxDB (GPM_2012_09) [53]. Ribo-seq reads were analyzed in Python 3.5 using custom scripts and Biopython 1.6.8, SciPy 0.19.0, and NumPy 1.12.1.

Translation simulation

Translation was modelled stochastically using a totally asymmetric simple exclusion process (TASEP) model, as performed in previous studies [54–56]. An mRNA with N codons is modeled as a lattice of N sites. At any time t , each site can be occupied by a single ribosome. Each ribosome translates a single site, as well as covers L sites in total, so that other ribosomes are excluded from these sites (cannot occupy or translate them).

The model dynamics consist of 3 possible events: (1) A free ribosome will attach to site $i = 1$ with rate λ , provided that the first $L/2$ codons on the mRNA are empty; (2) An attached ribosome located at site i will move to the next site $i + 1$ with rate λ_i , provided codon $i + L/2$ is not covered by another ribosome (otherwise we refer to this ribosome as “queued” or “delayed”); (3) In the case of $i = N$ the ribosome movement is in fact a termination step. It is assumed that the time between initiation attempts is distributed exponentially with rate λ . In addition, it is assumed that the time between jump attempts from site i to $i + 1$ is exponentially distributed

with translation rate λ_i , which is inversely proportional to the mean translation time t_i of the codon. It follows, that the time between any consecutive events is exponentially distributed (the minimum of exponentially distributed random variables) with the rate μ , given by:

$\mu(\{n_i\}_{i=1}^N) = \lambda + \sum_{i=1}^N n_i \lambda_i$, where $n_i \in \{0,1\}$ denotes the translation state of site i (equal to 1 while being translated, and 0 otherwise). The probability that a specific event was an initiation attempt is given by $\lambda/\mu(\{n_i\})$, while the probability that a specific event was a jump attempt (or termination event) from site i to site $i + 1$ is given by $n_i \lambda_i / \mu(\{n_i\})$.

In each simulation step, an event and the time that passed between events are sampled according to the above distributions. Next, state variables are updated according to the rules for each event. A pseudo-code of the simulation is given in [S1 Text](#). In order to bring the system to a steady state, the simulation was initiated with an mRNA empty of ribosomes, followed by simulation steps until reaching 100 terminations. The system was then simulated for additional 10,000 terminations while keeping track of various properties of the system (details in the Monitored features section below). For example, the steady state rate of protein production was determined by dividing the number of termination events by the total time that has passed. For this number of simulation steps, the difference between two simulation runs in the translation rate for a gene was small, typically 0.6% (median, ranging from 0.1% in the 10th percentile of differences between genes to 2% in the 90th percentile). The simulation was implemented and analyzed in MATLAB R2016b.

Monitored features

During simulation, steady state distributions were generated for a number of variables related to the translation process. The state variables $\{n_i\}$ were integrated over the simulation time, and divided by total simulation time to get a vector of ribosomal densities for each site on the transcript (the fraction of time a site was occupied by a ribosome). The sum of which is equal to the mean number of ribosomes on the transcript in steady state. Similarly, density profiles were generated for simulated footprints of any feasible size k (the number of ribosomes in the footprint) by assigning, at each simulation step, every ribosome to a footprint set according to its distance from neighboring ribosomes. If the distance between two ribosomes was smaller than the maximal allowed spacing, the pair was assigned to the same footprint. After completion of assignments, state variables $\{m_i^k\}$ were set to 1 (translated) / 0 (otherwise) according to the position of the upstream ribosome in each footprint of size k . The simulated footprint density profiles were obtained by integrating $\{m_i^k\}$ over time. Finally, queuing and delay of ribosomes were monitored during the simulation. First, a state variable was defined per site $\{d_i\}$, which was set to 1 once a ribosome in site i failed at a jump event due to the next codon being covered by another ribosome, and was set back to 0 once the ribosome successfully jumped. The density of queued ribosomes was obtained by integration similarly to above; the sum of which gives the average number of queued ribosomes on the transcript (or the queued fraction of ribosomes, QFR, when normalized by the number of ribosomes in the gene/region). Second, the total time that each ribosome spent translating the RNA, as well as the total time spent in queued state, were averaged over the simulated terminations to obtain the average translation and delay times. Finally, a state variable q was defined for the transcript, which was set to 1 as long as any ribosome was delayed across the transcript, and utilized it to obtain the probability for a queue.

In order to keep track of the quantities of ribosomes / RPFs in the cell, the RNA copy numbers measured by Arava et al. [40] were employed, while taking into account the estimated occupied (actively translated) RNA copies and the simulated ribosomal / RPF densities described above. For the purpose of computing the correlation between the simulated profiles

and Ribo-seq read count profiles (as reported in [S5 Fig](#)), reads were sampled from this simulated RPF distribution until reaching an equal number of reads to those in the Ribo-seq experiment.

Simulation calibration

Codon translation times were obtained from [\[57\]](#), where they were fitted using a whole-cell translation simulation model with finite cellular resources and Ribo-seq data ([S4 Table](#)). Initiation rates were selected to optimize the fit of the number of ribosomes on each transcript in steady state to the experimental data of Arava et al. [\[40\]](#). In this study, the authors measured a transcriptome-wide profile of ribosome association with mRNAs, including the number of ribosomes per-transcript (a total of 51,819 ribosomes on 5,700 transcripts, [S3 Table](#) in the original paper), their RNA copy numbers (a total of 12,534 molecules) and an estimation of the fraction of actively translated RNA copies of each type. As estimations for the total number of ribosomes ($187,000 \pm 56,000$, of which 70–85% are engaged in active translation) and RNAs (ranging from 12,500 to 60,000) in a yeast cell vary in the literature [\[40,58,59\]](#), we considered 3 major scenarios: (1) A “balanced” scenario, taking the number of RNAs to be in the mid-range of the above estimates (26,000 [\[60\]](#)), as well as taking the number of ribosomes to be in the mid-range (187,000, of which 85% actively translating); (2) A “dense” scenario, taking the number of RNAs to be in the lower range (12,500) and the number of ribosomes to be in the upper range (243,000, of which 85% actively translating); (3) A “sparse” scenario, taking the number of RNAs to be in the upper range (60,000) and the number of ribosomes to be in the lower range (131,000, of which 70% actively translating). In the first scenario, all but 194 genes converged to feasible initiation rates (that successfully induced the number of ribosomes in steady state as reported by Arava et al., after scaling), and the median initiation rate was identical to a previous study (0.09s^{-1} [\[41\]](#)). In the second and third scenarios 2,341 and 4 genes failed, respectively; the median initiation rate was 0.13s^{-1} and 0.02s^{-1} ; and the fit of the model to experimental data was generally less accurate ([S5B Fig](#)). The model was tested with varying termination times ranging from 1-fold to 500-fold the typical translation time of the fastest codon, and did not find significant improvement for longer termination times and therefore set it to be as fast as the fastest codon ([S5C Fig](#)). Ribosome size was set to be 10 codons, based on the typical footprint length of mRPFs. Results for various choices of ribosome size appear in [S5D Fig](#). It can be seen that results do not change considerably in the range of 9–15 codons, with marginally better fitting with experimental data in the range of 10–12 codons. Finally, the model was tested with different footprint spacing parameters (allowing up to 0–4 codons between ribosomes on a single footprint), and found a slight increase (but no clear maximum) in correlation with Ribo-seq double footprints (dRPFs) and single footprints (mRPFs) when allowing a higher number of free codons between ribosomes within a footprint ([S5E Fig](#)).

Initiation rates were also fitted to Ribo-seq and RNA-seq data using the TASEP model, for the purpose of analyzing their relation with missing mRPFs and the dRPF-to-mRPF ratio ([Fig 3B](#)). To this end, the number of ribosomes on each transcript was estimated using the ratio of Ribo-seq (dRPF) to RNA-seq RPKM values; the number of copies of each transcript was estimated using the RNA-seq RPKM values. Ribosome and RNA quantities were then scaled to have the same total numbers as described in the first scenario above.

Randomization

We generated simulations for 100 genomes according to each of two models. In the first model, we permuted the rates of codons globally, i.e. by changing the translation rate of all positions of a codon in the transcriptome at once. The table with permuted codon rates was

not constrained by codon synonymy, hence this model may serve as a validation for the calibration of the simulation. In the second model, we kept the original codon rates, but shuffled the order of synonymous codons across the sequence of a gene. Thus, this model controls for codon order in each transcript. In both randomizations, we selected initiation rates resulting in the same average number of ribosomes on the transcript as measured in real genes.

Functional enrichment

We utilized annotations of GO terms (accessed 04/09/15) obtained from SGD [61] and mapped them to generic slim definitions of the biological process ontology [62]. In addition, we obtained pathways from Wiki Pathways (accessed 29 April 2016) [63].

Supporting information

S1 Text. Simulation pseudo-code.

(PDF)

S1 Table. Simulated traffic properties. The table reports for each gene the initiation rate, mean number of ribosomes on the transcript, translation rate, translation time, delay time, the queued fraction of ribosomes (QFR), and a measure of ramp strength based on the ratio between QFR in the first 100 codons and the middle of the gene (excluding first/last 100 codons).

(XLSX)

S2 Table. Gene sets with extreme traffic conditions. Top/bottom 5% of genes according to their rank and z-scores (based on shuffling of synonymous codon order).

(XLSX)

S3 Table. Datasets and mapping statistics.

(XLSX)

S4 Table. Codon translation times. Obtained from: Zur H, Tuller T. Tracking the evolution of mRNA translation via whole cell simulation. [57].

(XLSX)

S1 Fig. Sucrose gradient fractions. (A) Quantification of ribosome protected footprints (RPFs) according to size. The plot shows the area under the curve (AUC) for mRPF and dRPF fractions. The ratio of dRPFs to mRPFs is reported in the caption. Peaks were detected using MATLAB's findpeaks function, and their width was determined by the positions where the curve decreases to half the height of the peak, or where the 2nd derivative turns positive. (B) Same for Ingolia 2009 data, profile reproduced from [23]. (C) Same for Gydosh 2014 data, profile reproduced from [33]. (D) Same for Shirokikh 2017 data, profile reproduced from [35] (Figure 5 in the original paper). See also Fig 1.

(EPS)

S2 Fig. Fig 2 repeated for the Diamant et al. dataset.

(EPS)

S3 Fig. Fig 2 repeated for the Gydosh et al. dataset.

(EPS)

S4 Fig. (A) Partial correlations between missing mRPFs and transcript features, as reported in Fig 3B, for different detection thresholds. (B) A histogram of simulated RPFs, *i.e.* the number of ribosomes in the transcriptome that are isolated or part of a pair, triplet etc. when no space

is allowed between ribosomes on the same footprint. The simulated dRPF to mRPF ratio is reported. **(C)** Same, where the space of ribosomes in a footprint does not exceed 1 codons. **(D)** Same, where the space of ribosomes in a footprint does not exceed 2 codons. **(E)** Same, where the space of ribosomes in a footprint does not exceed 3 codons. **(F)** Same, where the space of ribosomes in a footprint does not exceed 4 codons. (EPS)

S5 Fig. Methods. **(A)** Distributions of mapped double ribosome protected footprints (dRPFs) in the datasets used in this study. **(B)** Calibration of simulation parameters according to 3 scenarios for the number of ribosomes and RNA molecules in a cell (details in the Materials and Methods section). Simulated mRPFs and dRPFs were sampled to have the same number of reads as in the respective Ribo-seq dataset (an aggregation of our experiment and Guydosh et al.'s). Translational efficiency (TE) was defined as the ratio between RPKM values of Ribo-seq and RNA-seq reads, and compared to simulated translation rates. Initiation rates were inferred as described in the methods and compared to data from Ciandrini et al. **(C)** Calibration of the termination rate relative to the fastest codon. **(D)** Calibration of ribosome size. **(E)** Results for the simulated RPFs for different allowed spacing within a ribosome footprint. (EPS)

Acknowledgments

We thank Hadas Zur for providing the codon translation times. A.D. is grateful to the Azrieli Foundation for the award of an Azrieli Fellowship. This study was supported in part by a fellowship from the Edmond J. Safra Center for Bioinformatics at Tel-Aviv University.

Author Contributions

Conceptualization: Tamir Tuller.

Data curation: Alon Diamant, Elisheva Schochet, Tamir Tuller.

Formal analysis: Alon Diamant, Elisheva Schochet, Tamir Tuller.

Methodology: Anna Feldman, Martin Kupiec, Yoav Arava, Tamir Tuller.

Supervision: Tamir Tuller.

Validation: Alon Diamant, Tamir Tuller.

Visualization: Alon Diamant, Tamir Tuller.

Writing – original draft: Alon Diamant, Tamir Tuller.

Writing – review & editing: Alon Diamant, Martin Kupiec, Yoav Arava, Tamir Tuller.

References

1. Ingolia NT. Ribosome profiling: new views of translation, from single codons to genome scale. *Nature Reviews Genetics*; London. 2014; 15: 205–13. <https://doi.org/10.1038/nrg3645> PMID: 24468696
2. Zur H, Tuller T. Predictive biophysical modeling and understanding of the dynamics of mRNA translation and its evolution. *Nucleic Acids Res.* 2016; 44: 9031–9049. <https://doi.org/10.1093/nar/gkw764> PMID: 27591251
3. Warner JR, Rich A, Hall CE. Electron Microscope Studies of Ribosomal Clusters Synthesizing Hemoglobin. *Science.* 1962; 138: 1399–1403. <https://doi.org/10.1126/science.138.3548.1399> PMID: 17753859

4. Simha R, Zimmerman JM, Moacanin J. Polymerization Kinetics of Biological Macromolecules on Templates. *The Journal of Chemical Physics*. 1963; 39: 1239–1246. <https://doi.org/10.1063/1.1734422> PMID: 5878586
5. Zimmerman JM, Simha R. The kinetics of multicenter macromolecule growth along a template. *Journal of Theoretical Biology*. 1965; 9: 156–185. [https://doi.org/10.1016/0022-5193\(65\)90106-2](https://doi.org/10.1016/0022-5193(65)90106-2) PMID: 5875315
6. MacDonald CT, Gibbs JH, Pipkin AC. Kinetics of biopolymerization on nucleic acid templates. *Biopolymers*. 1968; 6: 1–25. <https://doi.org/10.1002/bip.1968.360060102> PMID: 5641411
7. MacDonald CT, Gibbs JH. Concerning the kinetics of polypeptide synthesis on polyribosomes. *Biopolymers*. 1969; 7: 707–725. <https://doi.org/10.1002/bip.1969.360070508>
8. Gordon R. Polyribosome dynamics at steady state. *Journal of Theoretical Biology*. 1969; 22: 515–532. [https://doi.org/10.1016/0022-5193\(69\)90018-6](https://doi.org/10.1016/0022-5193(69)90018-6) PMID: 5803335
9. Zia RKP, Dong JJ, Schmittmann B. Modeling Translation in Protein Synthesis with TASEP: A Tutorial and Recent Developments. *J Stat Phys*. 2011; 144: 405–428. <https://doi.org/10.1007/s10955-011-0183-1>
10. Zhang S, Goldman E, Zubay G. Clustering of low usage codons and ribosome movement. *J Theor Biol*. 1994; 170: 339–354. <https://doi.org/10.1006/jtbi.1994.1196> PMID: 7996861
11. Mitarai N, Sneppen K, Pedersen S. Ribosome collisions and translation efficiency: optimization by codon usage and mRNA destabilization. *J Mol Biol*. 2008; 382: 236–245. <https://doi.org/10.1016/j.jmb.2008.06.068> PMID: 18619977
12. Mitarai N, Pedersen S. Control of ribosome traffic by position-dependent choice of synonymous codons. *Phys Biol*. 2013; 10: 056011. <https://doi.org/10.1088/1478-3975/10/5/056011> PMID: 24104350
13. Chu D, Thompson J, von der Haar T. Charting the dynamics of translation. *Biosystems*. 2014; 119: 1–9. <https://doi.org/10.1016/j.biosystems.2014.02.005> PMID: 24631460
14. Chu D, Kazana E, Bellanger N, Singh T, Tuite MF, von der Haar T. Translation elongation can control translation initiation on eukaryotic mRNAs. *EMBO J*. 2014; 33: 21–34. <https://doi.org/10.1002/embj.201385651> PMID: 24357599
15. Kudla G, Murray AW, Tollervey D, Plotkin JB. Coding-Sequence Determinants of Gene Expression in *Escherichia coli*. *Science*. 2009; 324: 255–258. <https://doi.org/10.1126/science.1170160> PMID: 19359587
16. McCarthy JEG. Posttranscriptional Control of Gene Expression in Yeast. *Microbiol Mol Biol Rev*. 1998; 62: 1492–1553. PMID: 9841679
17. Shah P, Ding Y, Niemczyk M, Kudla G, Plotkin JB. Rate-Limiting Steps in Yeast Protein Translation. *Cell*. 2013; 153: 1589–1601. <https://doi.org/10.1016/j.cell.2013.05.049> PMID: 23791185
18. Plotkin JB, Kudla G. Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet*. 2011; 12: 32–42. <https://doi.org/10.1038/nrg2899> PMID: 21102527
19. Firczuk H, Kannambath S, Pahle J, Claydon A, Beynon R, Duncan J, et al. An in vivo control map for the eukaryotic mRNA translation machinery. *Mol Syst Biol*. 2013; 9: 635. <https://doi.org/10.1038/msb.2012.73> PMID: 23340841
20. Supek F, Šmuc T. On Relevance of Codon Usage to Expression of Synthetic and Natural Genes in *Escherichia coli*. *Genetics*. 2010; 185: 1129–1134. <https://doi.org/10.1534/genetics.110.115477> PMID: 20421604
21. Ben-Yehzekel T, Atar S, Zur H, Diamant A, Goz E, Marx T, et al. Rationally designed, heterologous *S. cerevisiae* transcripts expose novel expression determinants. *RNA Biology*. 2015; 12: 972–984. <https://doi.org/10.1080/15476286.2015.1071762> PMID: 26176266
22. Gamble CE, Brule CE, Dean KM, Fields S, Grayhack EJ. Adjacent Codons Act in Concert to Modulate Translation Efficiency in Yeast. *Cell*. 2016; 166: 679–690. <https://doi.org/10.1016/j.cell.2016.05.070> PMID: 27374328
23. Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science*. 2009; 324: 218–223. <https://doi.org/10.1126/science.1168978> PMID: 19213877
24. Artieri CG, Fraser HB. Accounting for biases in riboprofiling data indicates a major role for proline in stalling translation. *Genome Res*. 2014; 24: 2011–2021. <https://doi.org/10.1101/gr.175893.114> PMID: 25294246
25. Ingolia NT, Lareau LF, Weissman JS. Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes. *Cell*. 2011; 147: 789–802. <https://doi.org/10.1016/j.cell.2011.10.002> PMID: 22056041

26. Ishimura R, Nagy G, Dotu I, Zhou H, Yang X-L, Schimmel P, et al. Ribosome stalling induced by mutation of a CNS-specific tRNA causes neurodegeneration. *Science*. 2014; 345: 455–459. <https://doi.org/10.1126/science.1249749> PMID: 25061210
27. Li G-W, Oh E, Weissman JS. The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature*. 2012; 484: 538–541. <https://doi.org/10.1038/nature10965> PMID: 22456704
28. Martens AT, Taylor J, Hilser VJ. Ribosome A and P sites revealed by length analysis of ribosome profiling data. *Nucl Acids Res*. 2015; 43: 3680–3687. <https://doi.org/10.1093/nar/gkv200> PMID: 25805170
29. Mohammad F, Woolstenhulme CJ, Green R, Buskirk AR. Clarifying the Translational Pausing Landscape in Bacteria by Ribosome Profiling. *Cell Reports*. 2016; 14: 686–694. <https://doi.org/10.1016/j.celrep.2015.12.073> PMID: 26776510
30. Sabi R, Tuller T. A comparative genomics study on the effect of individual amino acids on ribosome stalling. *BMC Genomics*. 2015; 16: S5. <https://doi.org/10.1186/1471-2164-16-S10-S5> PMID: 26449596
31. Sabi R, Tuller T. Computational analysis of nascent peptides that induce ribosome stalling and their proteomic distribution in *Saccharomyces cerevisiae*. *RNA*. 2017; 23: 983–994. <https://doi.org/10.1261/rna.059188.116> PMID: 28363900
32. Woolstenhulme CJ, Guydosh NR, Green R, Buskirk AR. High-Precision Analysis of Translational Pausing by Ribosome Profiling in Bacteria Lacking EFP. *Cell Reports*. 2015; 11: 13–21. <https://doi.org/10.1016/j.celrep.2015.03.014> PMID: 25843707
33. Guydosh NR, Green R. Dom34 Rescues Ribosomes in 3' Untranslated Regions. *Cell*. 2014; 156: 950–962. <https://doi.org/10.1016/j.cell.2014.02.006> PMID: 24581494
34. Ingolia NT, Brar GA, Rouskin S, McGeachy AM, Weissman JS. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat Protocols*. 2012; 7: 1534–1550. <https://doi.org/10.1038/nprot.2012.086> PMID: 22836135
35. Shirokikh NE, Archer SK, Beilharz TH, Powell D, Preiss T. Translation complex profile sequencing to study the in vivo dynamics of mRNA-ribosome interactions during translation initiation, elongation and termination. *Nat Protocols*. 2017; 12: 697–731. <https://doi.org/10.1038/nprot.2016.189> PMID: 28253237
36. Diamant A, Tuller T. Estimation of ribosome profiling performance and reproducibility at various levels of resolution. *Biology Direct*. 2016; 11: 24. <https://doi.org/10.1186/s13062-016-0127-4> PMID: 27160013
37. Tuller T, Zur H. Multiple roles of the coding sequence 5' end in gene expression regulation. *Nucl Acids Res*. 2015; 43: 13–28. <https://doi.org/10.1093/nar/gku1313> PMID: 25505165
38. Tuller T, Carmi A, Vestsigian K, Navon S, Dorfan Y, Zaborske J, et al. An Evolutionarily Conserved Mechanism for Controlling the Efficiency of Protein Translation. *Cell*. 2010; 141: 344–354. <https://doi.org/10.1016/j.cell.2010.03.031> PMID: 20403328
39. Reis M dos, Savva R, Wernisch L. Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res*. 2004; 32: 5036–5044. <https://doi.org/10.1093/nar/gkh834> PMID: 15448185
40. Arava Y, Wang Y, Storey JD, Liu CL, Brown PO, Herschlag D. Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*. *PNAS*. 2003; 100: 3889–3894. <https://doi.org/10.1073/pnas.0635171100> PMID: 12660367
41. Ciandrini L, Stansfield I, Romano MC. Ribosome Traffic on mRNAs Maps to Gene Ontology: Genome-wide Quantification of Translation Initiation Rates and Polysome Size Regulation. *PLoS Comput Biol*. 2013; 9: e1002866. <https://doi.org/10.1371/journal.pcbi.1002866> PMID: 23382661
42. Brar GA, Yassour M, Friedman N, Regev A, Ingolia NT, Weissman JS. High-Resolution View of the Yeast Meiotic Program Revealed by Ribosome Profiling. *Science*. 2012; 335: 552–557. <https://doi.org/10.1126/science.1215110> PMID: 22194413
43. Gloge F, Becker AH, Kramer G, Bukau B. Co-translational mechanisms of protein maturation. *Current Opinion in Structural Biology*. 2014; 24: 24–33. <https://doi.org/10.1016/j.sbi.2013.11.004> PMID: 24721450
44. Buttgerit F, Brand MD. A hierarchy of ATP-consuming processes in mammalian cells. *Biochemical Journal*. 1995; 312: 163–167. <https://doi.org/10.1042/bj3120163> PMID: 7492307
45. von der Haar T. MATHEMATICAL AND COMPUTATIONAL MODELLING OF RIBOSOMAL MOVEMENT AND PROTEIN SYNTHESIS: AN OVERVIEW. *Computational and Structural Biotechnology Journal*. 2012; 1: e201204002. <https://doi.org/10.5936/csbj.201204002> PMID: 24688632
46. Navon S, Pilpel Y. The role of codon selection in regulation of translation efficiency deduced from synthetic libraries. *Genome Biology*. 2011; 12: R12. <https://doi.org/10.1186/gb-2011-12-2-r12> PMID: 21284851

47. Quax TEF, Claassens NJ, Söll D, van der Oost J. Codon Bias as a Means to Fine-Tune Gene Expression. *Molecular Cell*. 2015; 59: 149–161. <https://doi.org/10.1016/j.molcel.2015.05.035> PMID: 26186290
48. Latif H, Szubin R, Tan J, Brunk E, Lechner A, Zengler K, et al. A streamlined ribosome profiling protocol for the characterization of microorganisms. *BioTechniques*. 58. Available: http://www.academia.edu/23544215/A_streamlined_ribosome_profiling_protocol_for_the_characterization_of_microorganisms
49. Yates A, Akanni W, Amode MR, Barrell D, Billis K, Carvalho-Silva D, et al. Ensembl 2016. *Nucl Acids Res*. 2016; 44: D710–D716. <https://doi.org/10.1093/nar/gkv1157> PMID: 26687719
50. Pelechano V, Wei W, Steinmetz LM. Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature*. 2013; 497: 127–131. <https://doi.org/10.1038/nature12121> PMID: 23615609
51. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*. 2011; 17: 10–12. <https://doi.org/10.14806/ej.17.1.200>
52. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*. 2009; 10: R25. <https://doi.org/10.1186/gb-2009-10-3-r25> PMID: 19261174
53. Wang M, Weiss M, Simonovic M, Haertinger G, Schrimpf SP, Hengartner MO, et al. PaxDb, a database of protein abundance averages across all three domains of life. *Mol Cell Proteomics*. 2012; 11: 492–500. <https://doi.org/10.1074/mcp.O111.014704> PMID: 22535208
54. Dana A, Tuller T. Determinants of Translation Elongation Speed and Ribosomal Profiling Biases in Mouse Embryonic Stem Cells. *PLoS Comput Biol*. 2012; 8: e1002755. <https://doi.org/10.1371/journal.pcbi.1002755> PMID: 23133360
55. Dana A, Tuller T. The effect of tRNA levels on decoding times of mRNA codons. *Nucl Acids Res*. 2014; 42: 9171–9181. <https://doi.org/10.1093/nar/gku646> PMID: 25056313
56. Tuller T, Veksler-Lublinsky I, Gazit N, Kupiec M, Ruppin E, Ziv-Ukelson M. Composite effects of gene determinants on the translation speed and density of ribosomes. *Genome Biology*. 2011; 12: R110. <https://doi.org/10.1186/gb-2011-12-11-r110> PMID: 22050731
57. Zur H, Tuller T. Tracking the Evolution of mRNA Translation via Whole Cell Simulation. submitted. 2017;
58. von der Haar T. A quantitative estimation of the global translational activity in logarithmically growing yeast cells. *BMC Systems Biology*. 2008; 2: 87. <https://doi.org/10.1186/1752-0509-2-87> PMID: 18925958
59. Zenklusen D, Larson DR, Singer RH. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat Struct Mol Biol*. 2008; 15: 1263–1271. <https://doi.org/10.1038/nsmb.1514> PMID: 19011635
60. Pelechano V, Chávez S, Pérez-Ortín JE. A complete set of nascent transcription rates for yeast genes. *PLoS ONE*. 2010; 5: e15442. <https://doi.org/10.1371/journal.pone.0015442> PMID: 21103382
61. Cherry JM, Adler C, Ball C, Chervitz SA, Dwight SS, Hester ET, et al. SGD: Saccharomyces Genome Database. *Nucl Acids Res*. 1998; 26: 73–79. <https://doi.org/10.1093/nar/26.1.73> PMID: 9399804
62. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. *Nature Genetics*. 2000; 25: 25–29. <https://doi.org/10.1038/75556> PMID: 10802651
63. Kelder T, Iersel MP van, Hanspers K, Kutmon M, Conklin BR, Evelo CT, et al. WikiPathways: building research communities on biological pathways. *Nucl Acids Res*. 2012; 40: D1301–D1307. <https://doi.org/10.1093/nar/gkr1074> PMID: 22096230