Comprehensive translational profiling and STE AI uncover rapid control of protein biosynthesis during cell stress

Attila Horvath, Yoshika Janapala, Katrina Woodward, Shafi Mahmud, Alice Cleynen, Elizabeth E Gardiner, Ross D Hannan, Eduardo Eyras, Thomas Preiss, Nikolay E Shirokikh

To cite this version:

Attila Horvath, Yoshika Janapala, Katrina Woodward, Shafi Mahmud, Alice Cleynen, et al.. Comprehensive translational profiling and STE AI uncover rapid control of protein biosynthesis during cell stress. Nucleic Acids Research, 2024, 52 (13), pp.7925-7946. 10.1093/nar/gkae365 . hal-04787783

HAL Id: hal-04787783
https://hal.science/hal-04787783v1

Submitted on 18 Nov 2024

Nucleic Acids Research, 2024, **52**, 7925–7946
https://doi.org/10.1093/nar/gkae365
Advance access publication date: 9 May 2024
RNA and RNA-protein complexes

Comprehensive translational profiling and STE AI uncover rapid control of protein biosynthesis during cell stress

Attila Horvath[1,†], **Yoshika Janapala**[1,†], **Katrina Woodward**[1], **Shafi Mahmud**[1], **Alice Cleynen**[1,2], **Elizabeth E. Gardiner**[3], **Ross D. Hannan**[1,4,5,6,7], **Eduardo Eyras**[1,8,9], **Thomas Preiss**[1,10,*] and **Nikolay E. Shirokikh**[1,*]

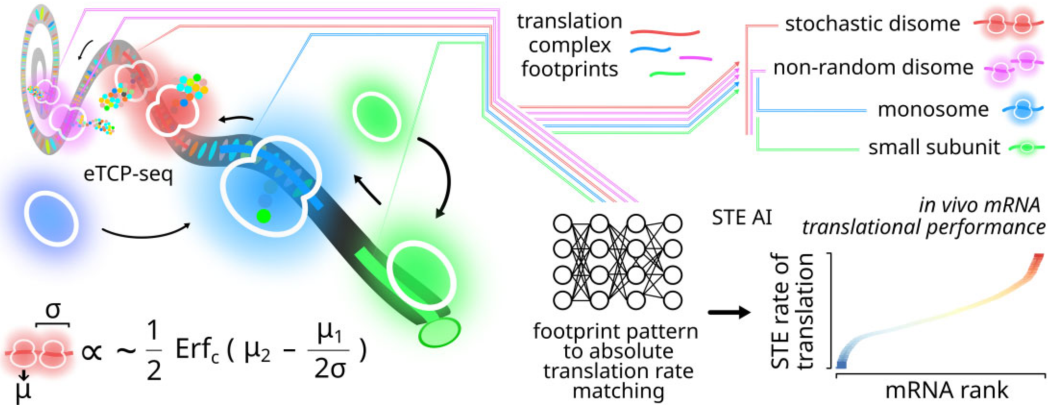
[1]Division of Genome Sciences and Cancer, The John Curtin School of Medical Research, and The Shine-Dalgarno Centre for RNA Innovation, The Australian National University, Canberra, ACT 2601, Australia
[2]Institut Montpelliérain Alexander Grothendieck, Université de Montpellier, CNRS, Montpellier, France
[3]Division of Genome Sciences and Cancer, The John Curtin School of Medical Research, and The National Platelet Research and Referral Centre, The Australian National University, Canberra, ACT 2601, Australia
[4]Department of Biochemistry and Molecular Biology, University of Melbourne, Parkville 3010, Australia
[5]Peter MacCallum Cancer Centre, Melbourne 3000, Australia
[6]Department of Biochemistry and Molecular Biology, Monash University, Clayton 3800, Australia
[7]School of Biomedical Sciences, University of Queensland, St Lucia 4067, Australia
[8]Division of Genome Sciences and Cancer, The John Curtin School of Medical Research, and The Centre for Computational Biomedical Sciences, The Australian National University, Canberra, ACT 2601, Australia
[9]EMBL Australia Partner Laboratory Network at the Australian National University, Canberra, ACT 2601, Australia
[10]Victor Chang Cardiac Research Institute, Darlinghurst, NSW 2010, Australia

[*] To whom correspondence should be addressed. Tel: +61 432847526; Fax: +61 432847526; Email: nikolay.shirokikh@anu.edu.au
[*] Correspondence may also be addressed to Thomas Preiss. Tel: +61 478493849; Fax: +61 478493849; Email: thomas.preiss@anu.edu.au
[†] The first two authors should be regarded as Joint First Authors.

Abstract

Translational control is important in all life, but it remains a challenge to accurately quantify. When ribosomes translate messenger (m)RNA into proteins, they attach to the mRNA in series, forming poly(ribo)somes, and can co-localize. Here, we computationally model new types of co-localized ribosomal complexes on mRNA and identify them using enhanced translation complex profile sequencing (eTCP-seq) based on rapid *in vivo* crosslinking. We detect long disome footprints outside regions of non-random elongation stalls and show these are linked to translation initiation and protein biosynthesis rates. We subject footprints of disomes and other translation complexes to artificial intelligence (AI) analysis and construct a new, accurate and self-normalized measure of translation, termed stochastic translation efficiency (STE). We then apply STE to investigate rapid changes to mRNA translation in yeast undergoing glucose depletion. Importantly, we show that, well beyond tagging elongation stalls, footprints of co-localized ribosomes provide rich insight into translational mechanisms, polysome dynamics and topology. STE AI ranks cellular mRNAs by absolute translation rates under given conditions, can assist in identifying its control elements and will facilitate the development of next-generation synthetic biology designs and mRNA-based therapeutics.

Graphical abstract



Received: January 9, 2024. **Revised:** March 21, 2024. **Editorial Decision:** April 11, 2024. **Accepted:** April 25, 2024

Introduction

Translation of mRNA into proteins is an actively regulated stage of gene expression (1–3). Many rapid responses of eukaryotic cells are largely based on, or are invoked by, translational control, including archetypal processes such as nutrient-induced control, endoplasmic reticulum stress, dynamic cell reprogramming before and during embryonic development, and synaptic plasticity (4–8). Regulation of translation is of heightened importance with the advent of mRNA-based vaccines and therapeutics (9–11). Novel approaches utilize structured regions and code optimality of mRNA to inform next-generation, AI-based tools for design of more efficient mRNA (12–21). The performance of these tools depends on the depth and accuracy of the underlying data and concepts. Yet, it remains a substantial challenge to accurately determine mRNA translation rates transcriptome-wide, which is an obstacle in our ability to understand and employ translational control (22,23).

In yeast, including *Saccharomyces cerevisiae*, translational control is one of the major components of the cell stress response, such as stress caused by nutrient starvation (3,6,8,24). Indeed, the classic example of *GCN4* regulation demonstrates the importance of rapid and specific translational change as the initial nutrient stress response, which then converts into longer-term transcriptional reprogramming (6,25,26). Many studies have highlighted the complexity of glucose starvation responses in yeast (6,7,27,28). Beyond the reduction in polysomes explained by decreased translational engagement of mRNAs, many mRNAs move to granules, phase separate or decay rapidly (29–32); some are newly synthesized (33,34), and on some mRNAs, loss of translation initiation factors such as eIF4A results in translational stalling at the early stages of mRNA engagement (35–38). Ribosomal stalling on mRNA in stressed cells can lead to stacked ribosomes and the induction of ribosome quality control mechanisms (27,39,40). Overall, stress-induced translational control is a highly dynamic process intertwined with gene expression regulation and tightly integrated with transcriptional reprogramming downstream (6,25,38,41,42).

Mechanistic insights into translation are often gained by polysome or ribosome profiling approaches, because of their convenience and transcriptome-wide reach (4,43–47). In polysome profiling, the extent of mRNA association with polysomes is assessed, and comparisons between the order of polysomes are made. In ribosome profiling (ribo-seq), frequencies of footprints from singular ribosomes (monosomes), generated by limited RNase treatment, are measured relative to mRNA abundance. These techniques use high-throughput sequencing to provide deep transcript-wise information, but have several limitations (48,23,49–51). One is the assumption that a higher density of ribosomes present on mRNA is indicative of more intense translation (4). This is not always correct as there are code (52–54), codon (50,52,55–57), peptide (ribosomal interaction and conformation) (58,59), localization (signal peptide) (60–62), co-translational folding (63–65) and quality control-induced mechanisms of ribosomal stalling on mRNA (66–68), including some cases of well-characterized artifacts of translational stabilization with antibiotics and cell stress (4,50,69,70). Thus, the presence of more ribosomes on mRNA can also be a sign of a reduced translational output. Another limitation is that polysome and ribosome profiling experiments employ normalization of the polysomal mRNA or ribosome footprints sequenced by the 'total mRNA' sequencing signal (48,71). This assumes that (a) all mRNA is accessible to translation at the time of measurement, (b) there are no rapid alterations of mRNA abundance that can bias the measurements, and (c) mRNA is represented or degraded uniformly relative to the locations of ribosomes over it. However, nuclear export can be controlled in an mRNA-specific manner (72), mRNAs can undergo phase separation into granules (73–75), and there are mechanisms of mRNA degradation targeting specific regions (27), including translationally-induced decay (57,70). These processes create a biological disparity in mRNA coverage in the context of ribosomal association with mRNA. Furthermore, methods for RNA purification as well as sequencing library preparation and the attendant biases and normalization requirements differ between the total mRNA reference and the polysome or ribosome footprint samples. These methodological limitations mean that the resultant data are disconnected from the absolute mRNA abundances, thus only enabling per-mRNA comparisons between conditions. Therefore, a more accurate assessment of protein biosynthesis enabling the calculation of absolute translational rates would be of critical value.

Here, we employed an enhanced variant of our Translation Complex Profiling sequencing, eTCP-seq, to reveal ribosomal scanning and other phases of translation outside the elongation cycle, in addition to the elongating ribosomes (4,5,76–78). In the eTCP-seq, data is expanded to include a 'disome' fraction resistant to nuclease attack (5). Importantly, eTCP-seq uses rapid *in vivo* formaldehyde fixation to stabilize translational complexes, which is indiscriminate to complex type and results in the most comprehensive available 'snapshot' of translation (4). We find that the 'disomes', or co-localized ribosomes, revealed by eTCP-seq could be of multiple types, in similarity to the elongation, initiation, termination and recycling complexes captured by the method (4,76,78,79). We thus further propose a more complete classification of ribosomal co-localization types, which in addition to the known stalling-induced 'disomes' (or other multisomes) includes co-localization resulting from spatial proximity of the ribosomes other than linear neighboring over the mRNA, and diffusion-mediated stochastic variance of the underlying processes (initiation, elongation) (4,5,76–78). These 'stochastic disomes' are not a result of catching up due to differences in average elongation rate between leading and trailing ribosomes and represent a principally new type of signal.

We then develop an unbiased and comprehensive, unsupervised machine learning pipeline to define the Stochastic Translation Efficiency (STE) measure. Our approach incorporates rich sets of translation complex positional information across the entire translation cycle as measured by eTCP-seq. We compare STE with an RNA-seq-normalized footprint measure (as used in Translation Efficiency; TE) to demonstrate a much higher concordance of STE with absolute protein biosynthesis rates. We use STE to characterize acute translational regulation in yeast cells subjected to 10-minute glucose starvation, revealing translational up-regulation of oxidative stress and sugar metabolism genes. We propose STE as a robust AI-based measure of translational output with the advantages of: (i) inherent single-molecule-event nature of the main measure component (the stochastic co-localization footprint signal); (ii) similar types of data used in all calculations; (iii) independence from the abundance and availability changes of mRNA and (iv) an insensitivity to the differences in library preparation techniques.

Materials and methods

Cell material, fixation regimen and cytosol collection

Cell fixation and harvesting was performed as described before (76,77) with modifications (5). Wild-type (WT) yeast of cell line BY4741 (MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0) was grown in 1 liter of YPDA [1% w/v yeast extract (Merck/Sigma-Aldrich cat. no. 70161), 2% w/v peptone (Merck/Sigma-Aldrich cat. no. 70178), 2% w/v dextrose (Merck/Sigma-Aldrich cat. no. 49139), 40 mg/l adenine sulfate (Amresco 0607-100G)] media until an optical density of 0.7–0.8 AU at 600 nm (OD_{600}) was reached. The cells were immediately snap-chilled by mixing with 25% w/v of crushed ice, and 37% w/v formaldehyde solution (methanol-stabilized solution; 'formalin'; Merck/Sigma-Aldrich cat. no. F11635-500ML) was added immediately under constant mixing to a final concentration of 2.2% (w/v). Cells were incubated for 10 min on ice for fixation and then pelleted by centrifugation at 4°C, 5000 × g for 5 min. The fixed cell pellet was resuspended and washed with 40 ml of buffer A (20 mM HEPES–KOH pH 7.4 at 25°C, 100 mM KCl, 2 mM $MgCl_2$ and 250 mM glycine), followed by centrifugation at 4°C, 5000 × g for 5 min. The supernatant was discarded and the cell pellet was then resuspended in 40 ml of buffer A1 (20 mM HEPES–KOH pH 7.4 at 25°C, 100 mM KCl and 2 mM $MgCl_2$). This quenching step is critical for avoiding irreproducible crosslinking; and it must occur within 20 min of cell harvest. The cells were pelleted again by centrifugation at 4°C, 5000 × g for 5 min, and the buffer A1 wash was repeated two more times. The washed cell pellet was aspirated (∼1 g wet cell mass) and resuspended in 550 μl of buffer A2 (buffer A1 supplemented with 5 mM DTT, 1 U/μl RNaseOUT RNase inhibitor (Thermo Fisher Scientific) and 1 × Complete EDTA-free Mini Protease Inhibitor (Merck). The cell suspension was then flash-frozen by dripping into liquid nitrogen and the pellets stored at −80°C.

To disrupt the cell wall and membrane, a 10 ml stainless steel grinding jar (Retsch) was pre-cooled by partial submerging in liquid nitrogen and then filled with ∼2 g of the frozen cell suspension pellets and two 12 mm stainless steel grinding balls (Retsch). The sealed grinding jars were shaken at 27 Hz for 1 min in MM400 mixer mill (Retsch), re-cooled by partial submerging in liquid nitrogen followed by continued shaking for 1 min more. The resultant powdered grindate was stored in 1.5 ml low protein binding tubes (Eppendorf) at −80°C in ∼100 mg aliquots and used as necessary. Usually, 600 mg of the grindate was used per one experiment comprising polysome sedimentation profile analysis, separation of the cytosol into translated and non-translated fractions, and further separation of the translated fraction into the ribosomal small subunit (SSU), ribosome (RS) and disome (DS) fractions upon RNase digestion.

Separation of fixed (poly)ribosomal complexes away from the non-translated fractions of cytosolic RNA

We followed the procedure established by us (76,77) to enrich for 'total translated' RNA (*tt*) based on its co-sedimentation with (polysomes, with some refinements (5). A modification was made to allow for a direct separation monitoring using absorbance profile readout upon ultracentrifugation and avoid the re-solubilization step that resulted in higher material losses and excessive denaturation and aggregation previously. ∼100 mg of the frozen cell grindate were thawed, supplemented with 150 μl buffer A2 and clarified by centrifugation at 4°C, 13 000 × g for 5 min. The resultant clarified mixture (∼150 μl) was then loaded onto a 10–20% (w/v) 2.5 ml linear sucrose gradient additionally containing 0.5 ml of 50% sucrose cushion at the bottom, made with buffer 1 (25 mM HEPES–KOH pH 7.6, 100 mM KCl, 5 mM $MgCl_2$, 0.1 mM EDTA, 5 mM DTT). The gradients were prepared using the freeze-thaw method (80) in thinwall Ultra-Clear ultracentrifuge tubes (5 ml, 13 × 51 mm; Beckman-Coulter). To create the 50% sucrose cushion, upon the gradients melting and stabilizing overnight at 4°C, 0.5 ml of 50% sucrose in buffer 1 were layered into the bottom of the tubes with a syringe-attached glass capillary. Tubes were then ultracentrifuged in an SW 55 Ti rotor at 4°C, at 55 000 rpm, average *g*-force 287 980 × g (*k*-factor 49), for 1 h 30 min. These conditions were pre-optimized (using post-ultracentrifugation gradient absorbance trace analysis) to retain the 'free' (non-polysomal) SSUs and LSUs in the top (10–20% sucrose) portion of the gradient, while concentrating the polysomal fraction in the bottom (50%) sucrose cushion without pelleting the material.

The polysomal fraction collected at the bottom of the tubes was concentrated to 100 μl using a Ultracel-10 regenerated cellulose membrane with 10 kDa molecular weight cut-off, in Amicon Ultra-0.5 ultrafiltration devices (Merck). To achieve partial removal of the sucrose, the initial concentrate was further diluted 4 times using buffer 1 and concentrated again to 200 μl. Polysomal presence in the resultant mixture was confirmed by absorbance readout of an analytical sucrose gradient ultracentrifugation run. The mixtures were stored frozen at −80°C and used as input material for the 'total translated' (*tt*) RNA-seq library construction, or the RNase digestion step of the eTCP-seq library construction.

RNase digestion of the fixed (poly)ribosomal complexes and separation of digested material into small ribosomal subunit (SSU), mono(ribo)somal (ribosomes, RS) and di(ribo)somal (disomes, DS) fractions

The procedure followed our approach as described before (76,77) with adjusted ultracentrifugation parameters to achieve higher resolution across all collected fractions (5). The *tt* fraction from the previous step was digested using 4.5 U of *E. coli* RNase I (Ambion) per 1 OD_{260} unit of the fraction for 30 min at 23°C. SUPERaseIn RNase inhibitor (Thermo Fisher Scientific) was then added to the mixture to 0.25 U/μl, to inactivate RNase I, and the samples were transferred to ice. The reaction mixtures were loaded onto 12.5 ml linear 10–40% w/v sucrose gradients formed in 13 ml thinwall polypropylene tubes, 14 × 89 mm (Beckman-Coulter) using the freeze-thaw method (80). Due to the different concentration of the SSU, RS and DS fractions and the resolving power of the gradients for each fraction, different loads of the material were used to achieve optimal separation. For the SSU and RS 13–14 AU, and for DS 10–11 AU per gradient were usually taken. Minimally two gradients were used per each DS purification.

The tubes with the loaded gradients were centrifuged in an SW 41 Ti rotor at 4°C, average g-force 178 305 × g (*k*-factor 143.9), for 3 h 30 min. Absorbance profiles of the resultant sucrose gradients were read at 254 nm, 1.5 ml/min, using the Gradient Fractionator instrument (Brandel). Fractions corresponding to the position and mobility of the SSU, RS and DS complexes were identified in real time and isolated, their average absorbance at 254 nm recorded, and the fraction material further stored at −80°C or processed immediately.

Construction of the eTCP-seq SSU, RS and DS footprint RNA-sequencing libraries

We used our approach as described before (76,77). The approach is based on RNA 3′ polyadenylation, oligo(dT)-dependent reverse transcription, cDNA circularization and PDD-based depletion of ribosomal RNA (81,82), with several streamlining modifications and an expanded ribosomal (r)RNA depletion probe set (Supplementary Table S1). ∼3.0 AU at 254 nm of the gradient-separated SSU, RS and DS material were used per each library.

To de-block the crosslinks and isolate the RNA, sucrose gradient fractions (350 μl) were supplemented with 40 μl of 100% stop solution (10% SDS w/v and 100 mM EDTA), Tris–HCl (pH 2 at 25°C) to 10 mM (4 μl 1 M), glycine to 10 mM (1.6 μl 2.5 M) and deionized nuclease-free water to obtain 400 μl as the final volume. Acidic phenol:chloroform:isoamyl alcohol 125:24:1 (pH 4.0–5.0) (Merck/Sigma-Aldrich) was added followed by vigorously shaking the mixtures using a vortex mixer set to the maximum speed for 2 min, and continuing shaking at 65°C, 1400 rpm for 30 min in a thermomixer (Eppendorf). Phase separation was facilitated by centrifuging the mixture at 12 000 × g for 10 min at room temperature. The aqueous phases were then collected and transferred to fresh 1.5 ml Eppendorf tubes. RNA was precipitated by adding 0.1 volumes of 3 M sodium acetate (Invitrogen/Thermo Fisher Scientific), 20 μg of glycogen (Invitrogen/Thermo Fisher Scientific) and 2.5 volumes of absolute ethanol (Merck/Sigma-Aldrich). The tubes were vortexed for 1 min and incubated at −20°C for at least 2 h. RNA was pelleted by centrifugation at 12 000 × g for 30 min at room temperature. The supernatant was discarded, and the pellet was washed twice with 80% v/v ethanol by centrifugation at 12 000 × g for 30 min at room temperature. The RNA pellets were dried at 45°C for 10 min, and then dissolved in 20 μl of 1 × HE buffer (final concentration 10 mM HEPES–KOH, pH 7.6 at 25°C and 0.25 mM EDTA, pH 8.0 at 25°C). RNA concentration was estimated using a Nanodrop spectrophotometer (Thermo Fisher Scientific) and RNA quality was further assessed using the RNA 6000 Pico Kit and Bioanalyzer 2100 (Agilent).

To reverse transcribe the RNA and introduce strand-specific identifiers, ∼8 pmol (calculated using the average length of ∼300 nt and amount of ∼1 μg) of the RNA per each library was taken in 20 μl of 1 × HE buffer solution. The RNA solution was transferred into a low DNA binding 1.5 ml tube (Eppendorf), heated at 70°C for 2 min and immediately transferred to ice for 5 min. The RNA was then end-repaired with 20 U of 3′-phosphatase-positive bacteriophage T4 polynucleotide kinase (T4 PNK; New England Biolabs), using conditions recommended by the supplier (1 × PNK buffer without ATP, 2 U/μl RNaseOUT Recombinant Ribonuclease Inhibitor (Thermo Fisher Scientific), incubation at 37°C for 2 h). The reaction was stopped and PNK inactivated by the addition of 17 μl of deionized water and heating of the reaction mixture for 20 min at 65°C. The end-repaired RNA was next 3′ polyadenylated using 0.4 U/μl of *E. coli* poly(A) polymerase (EPAP; New England Biolabs), generally according to the suppliers' recommendations (1 × EPAP buffer, 1 mM ATP, 1 mM DTT and 0.8 U/μl RNaseOUT Recombinant Ribonuclease Inhibitor (Thermo Fisher Scientific), incubation at 37°C for 1 h). The reaction was stopped by the addition of 15 μl of 100% stop solution and the resulting polyadenylated RNA was ethanol-precipitated, washed, dried as described before, and dissolved in 25 μl of 1 × HE. The RNA was next reverse transcribed using the oligo(dT) primer (5′-phosphate-GA TCG TCG GAC TGT AGA ACT CTG AAC G /9-carbon spacer/ G TGA CTG GAG TTC CTT GGC ACC CGA GAA TTC CAT TTT TTT TTT TTT TTT TTT TVN-3′) with SuperScript IV reverse transcriptase (Invitrogen/Thermo Fisher Scientific), generally as recommended by the supplier. The template RNA (up to 4 pmol) was first mixed with 20 pmol of the split adapter primer, 0.5 mM (each) dNTPs mixture, 1 × SuperScript IV buffer and annealed by heating to 75°C for 3 min, then cooling to 65°C, and slow ramping (3°C/s) to 55°C. The reaction mixture was then supplemented with 5 mM DTT, 2 U/μl RNasin Plus (Promega) and 10 U/μl SuperScript IV Reverse Transcriptase enzyme while heated, slow-ramped to 50°C and incubated at 50°C for further 30 min.

To purify the resultant cDNA away from the excess of the RNA and the split adapter primer, and create a strand-specific amplifiable template, the reaction mixture was snap-cooled to 37°C, supplemented with 20 U of *E. coli* exonuclease I (New England Biolabs) and incubated at 37°C for 20 min. The reaction was then stopped by the addition of 15 μl of 100% stop solution. Samples were then subjected to extraction with phenol:chloroform:isoamyl alcohol 25:24:1, pH 7.7–8.8 (Merck/Sigma-Aldrich). Aqueous phase separation and nucleic acid ethanol precipitation were performed as described earlier. The resultant purified cDNA was dissolved in 9 μl of 1 × HE and liberated from RNA contaminants by heating at 85°C for 3 min, snap-chilling on ice for 5 min, supplementing with 1 μl of RNase A/T mix (2 mg/ml of RNase A and 5000 U/ml of RNase T1, Thermo Fisher Scientific), and incubating the resulting mixtures for 20 min at 37°C.

80 pmol of empty split adapter circularization blocker (5′-TTN BAA AAA AAA AAA AAA AAA AAA /iSuper-dT/GG AAT TCT CGG GTG CGT GTG T/3BioTEG/-3′) and 2 μl of 10 × Annealing Buffer (250 mM HEPES–KOH pH 7.6 at 25°C, 1 M NaCl and 50 mM $MgCl_2$) were then added to a final volume of 20 μl and annealing performed by heating the mixtures at 80°C for 2 min, ramping up to 95°C for 1 min, cooling down to 50°C for 1 min and slow-ramping (3°C/min) to 40°C. The RNA-depleted cDNA was then circularized at 40°C according to the manufacturer's instructions, by adding 4 μl of 1 × CircLigase II Buffer (Epicentre/Illumina), 2 μl of 50 mM $MnCl_2$, 8 μl of 500 mM betaine, and 1 μl of 2.5 U of CircLigase II ssDNA Ligase (Lucigen), and incubating the resultant mixtures at 40°C for 3 h. The reaction was stopped by addition of 10 μl of 100% stop solution. 1 volume of neutral phenol:chloroform:isoamyl alcohol (pH 7.7–8.8) (Sigma-Aldrich/Merck) was then added and the mixtures intensively vortexed for 2 min. Separation of the aqueous and phenol phases was facilitated by centrifugation at room temperature, 12 000 × g, for 10 min. Extracted circularized cDNA in the water phase (∼50 μl) was further gel-filtered

using 1 × HE-equilibrated Illustra MicroSpin G-25 Columns (Sigma-Aldrich/Merck) according to the manufacturer's recommendations. To remove the empty split adapter circularization blocker, 23 μl (out of ∼50 μl) of the gel-filtered material were mixed with 40 pmol of the block removal oligonucleotide (5′-ACA CAC GCA CCC GAG AAT TCC ATT TTT TTT TTT TTT TTT T/iSuper-dT/T VNA A/3BioTEG/-3′), and supplemented with 3 μl of 10 × Annealing Buffer. The mixtures were heated at 80°C for 2 min, 95°C for 1 min, 80°C for 10 s, followed by slow-ramping to 50°C, further incubated at 50°C for 1 min and cooled to 40°C. The reaction mixtures were transferred into new 1.5 ml low DNA binding tubes (Eppendorf) containing hydrophilic streptavidin magnetic beads (equaling to 260 μl of the original bead suspension; New England Biolabs), pre-equilibrated with 1 × Annealing Buffer. The tubes with reaction mixtures and beads were incubated at room temperature for 10 min with gentle tube flicking to keep the beads suspended, followed by an incubation at 35°C for 2 min, immediate separation of the beads on a magnetic rack and collection of the supernatant containing the unbound circularized cDNA.

To specifically deplete cDNA representing fragments with rRNA sequences, the circularized cDNA was supplemented with 1.9 μM of each of the depletion DNA probes (Supplementary Table S1) and heated at 95°C for 3 min, cooled to 75°C (3°C/min), supplemented with 1 × Duplex-Specific Nuclease (DSN) buffer (Evrogen), slow-ramped (3°C/min) to 60°C and further supplemented with 0.5 U DSN enzyme while at 60°C. The reaction mixtures were then slow-ramped (3°C/min) to the hybridization temperature of 48°C and further incubated at 48°C for 20 min. The reactions were stopped by the addition of 20 mM EDTA and 55 μl of 1 × HE buffer, the mixture extracted with an equal volume of neutral phenol:chloroform:isoamyl alcohol 25:24:1 (∼75 μl; pH 7.7–8.8; Merck/Sigma-Aldrich), as described earlier, and gel-purified using HE-equilibrated MicroSpin G-25 Columns (Merck/Sigma-Aldrich) according to the manufacturer's recommendations.

To amplify the rRNA-depleted cDNA and perform library size-selection, the resultant purified PDD-treated cDNA was thermocycled with Platinum SuperFi DNA polymerase (Thermo Fisher Scientific), generally according to the manufacturer's instructions and using custom primer pairs compatible with TrueSeq Small RNA Sample Preparation Kit (Illumina), bearing a unique tag for each library. The forward primers were 5′-CAA GCA GAA GAC GGC ATA CGA GAT XXX XXX GTG ACT GGA GTT CCT TGG CAC CCG AGA ATT CCA-3′ (in which XXX XXX represents Illumina's indexing hexanucleotide sequences), and the reverse primer was 5′-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACG TTC AGA GTT CTA CAG TCC GA-3′. The amplification reaction included the 25 μl of purified PDD-treated cDNA, 0.2 mM each of dNTPs, 5 × SuperFi DNA polymerase buffer, 0.5 μM each of primers, 4.2 ng/μl extreme thermostable single-stranded DNA binding protein (New England Biolabs) and 0.02 U/μl SuperFi DNA polymerase (Thermo Fisher Scientific). Typically, 21 cycles were used, with a thermal profile 98°C for 5 min followed by 98°C for 30 s melting, annealing at 62°C for 30 s and extension was performed at 72°C for 45 s, this was followed for the first two cycles. For the third and subsequent cycles, melting was performed at 98°C for 30 s, annealing temperature was increased to 76°C for 30 s followed by extension at 72°C for 45 s; this was repeated for ∼18 cycles. The final extension was performed at 72°C for 1 min. The amount of the amplified DNA samples was equalized per barcode using band intensity measurements obtained by imaging a respective native agarose gel test run with the samples loaded separately for each barcode and pre-stained with 6 × GRGreen loading buffer (Excellgen). Libraries were then pooled together (typically by 4 per barcode), electrophoretically separated in a native agarose gel using 6 × GRGreen loading buffer (Excellgen), and selected for the insert size of 10–250 nt by cutting out the respective region of the gel. The libraries were eluted from the gel by freezing the gel pieces at −20°C for 30 min in Freeze 'N Squeeze DNA gel extraction spin columns (Bio-Rad) and recovering the solution by immediate centrifugation of the columns at room temperature, 13 000 × g, for 3 min. The recovered DNA solution was subjected to neutral phenol:chloroform:isoamyl alcohol 25:24:1 (pH 7.7–8.8, Merck/Sigma-Aldrich) extraction and ethanol precipitation, as described earlier. Dried pellets were dissolved in ∼20 μl 1 × HE buffer, quality-controlled with capillary electrophoresis (using High Sensitivity DNA chips run in Agilent Bioanalyzer 2100), and directed to the high-throughput sequencing input.

Construction of the long-read nanopore direct RNA sequencing (DRS) libraries

Approximately 200 mg of fixed, frozen cell grindate derived exactly as described in 'Cell material, fixation regimen and cytosol collection' was thawed and clarified by centrifugation at 4°C, 13 000 × g for 5 min to remove cell debris. The resulting supernatant (∼300 μl, ∼6.0 AU) was spiked with *in vitro* transcripts *i1* and *i2* (108 ng each; Supplementary Table S1), to control the representation of non-translated RNA, and loaded onto 12.5 ml linear 15–45% w/v sucrose gradients formed in thinwall polypropylene tubes (13 ml, 14 × 89 mm, Beckman-Coulter). The tubes were then centrifuged in an SW 41 Ti rotor at 4°C, average g-force 178 305 × g (*k*-factor 143.9) for 2 h. Absorbance profiles of the sucrose gradients were read at 254 nm (1.5 ml/min) using the Gradient Fractionator instrument (Brandel). Fractions corresponding to the RS, DS, Trisomes ('TS') and polysomes (tetrasomes and above, 'PS') complexes were identified and isolated in real time based on their position and mobility. The average absorbance at 254 nm of each fraction was recorded, and the fraction material was either stored at −80°C or processed immediately.

For the DRS, RS, DS and TS were combined proportionally to their volume (yielding 'ribosome-disome-trisome' fraction or 'RDT'), along with the PS (Supplementary Figure S11A). The RDT and PS fractions were subjected to reverse-crosslinking at 65°C for 30 min, followed by RNA extraction using Solid-Phase Reversible Immobilization (SPRI) beads. The volume of SPRI beads was used at an 1:1 ratio with the fractions. The mixture of RNA and SPRI beads was incubated at room temperature for 5 min on a tube rotator (minimum speed) for 8 min. The tubes were transferred to a magnetic rack and incubated for 2 min, and the supernatant was collected for troubleshooting purposes. The beads containing bound RNA were washed with freshly made 80% ethanol and then air-dried for 8 min at room temperature. The RNA was eluted by adding 32 μl of 1 × HE buffer and incubating for 5 min with the intermittent gentle flicking of the tube at room temperature. The tubes were then transferred to a magnetic rack for 2 min to collect the supernatant containing the extracted RNA in a new low-bind DNA tube. The RNA was quantified using Nanodrop spectrophotometer, and ∼2 μg of RNA was used to prepare DRS libraries.

Quantitative reverse transcription and amplification

Equal volumes of TS and PS were taken and spiked with *in vitro* transcripts *i3* and *i4* (10 ng each; Supplementary Table S1), providing an internal reference for RNA extraction, reverse transcription and amplification. The RNA extraction process was carried out as described in the 'Construction of the long-read nanopore direct RNA sequencing (DRS) libraries'. Upon extraction, a 50 μl reverse transcription reaction was set up, using ∼1 μg of the input RNA, 1 × TURBO DNase buffer (Thermo Fisher Scientific), 0.8 U/μl RNasin Plus (Promega) and 0.04 U/μl TURBO DNase (Thermo Fisher Scientific). The reaction mixture was incubated at 37°C for 25 min, reaction was stopped with 5.5 μl of DNase inactivation resin and the remaining volume recovered into a new tube as recommended by the manufacturer (Thermo Fisher Scientific).

For the reverse transcription, a protocol recommended by the reverse transcriptase manufacturer (Thermo Fisher Scientific) was followed. Briefly, the DNase-treated RNA, 1 μl of 2 μM primer mix (Supplementary Table S1) and 1 μl of 10 mM dNTPs were mixed in a PCR tube; the mixture was briefly spun down and incubated at 65°C for 5 min, followed by rapid cooling through sticking in ice for at least 1 min, to denature the RNA. A master mix was prepared containing 4 μl 5 × SSIV reaction buffer, 1 μl 100 mM DTT, 1 μl 40 U/μl RNasin Plus (Promega) and 1 μl 10 U/μl SuperScript IV Reverse Transcriptase (RT; Thermo Fisher Scientific). A separate mix with the reverse transcriptase enzyme replaced with deionised water was prepared for the 'RT minus' control. The annealed RNA sample from was divided in halves, with one portion used for the reaction and the other for the RT minus mix. The reaction mixtures were then incubated at 50°C for 10 min, followed by incubation at 80°C for 10 min to inactivate the RT. The resulting cDNA can be utilized for qPCR immediately or stored at –20°C.

For the qPCR, the cDNA was diluted 5 × with 1 × HE buffer. A master mix for qPCR was prepared, consisting of 5 μl 2 × SYBR Green PCR Master Mix (Thermo Fisher Scientific), 0.375 μl of each 20 μM forward and reverse primers and 0.250 μl of deionized water for each reaction (yeast targets *SSC1* (YJR045C), *RPS20* (YHL015W), *ATG40* (YOR152C), *GUS1* (YGL245W), *GCN2* (YDR283C), *RPL26A* (YLR344W), *SSE1* (YPL106C) and the spike-in IVTs *i1*, *i3*). qPCR reactions were set using technical triplicates where 6 μl of the master mix were combined with 4 μl of the diluted cDNA or RT- control reactions in 384-well plates. The plates were then sealed, contents mixed and spun down, and thermocycling run on QuantStudio 12K Flex (Thermo Fisher Scientific) with standard protocol for SYBR-based detection, whereby the initial denaturation was carried out at 95°C for 60 seconds, cyclic denaturation at 95°C for 15 s, and cyclic annealing/extension at 60°C for 30 s.

The $2^{-\Delta\Delta Ct}$ method was employed for assessing relative abundance changes. In this method, *i3* was utilized as the reference control. Here, ΔC_t refers to the difference between the C_t value of the target gene and the C_t value of the reference. Furthermore, $\Delta\Delta C_t$ represents the difference between the ΔC_t value of a specific sample of interest (starved for 10 min) and the ΔC_t value of the corresponding reference sample (non-starved). A sample size of $n = 3$ was employed.

Construction of the total (*T*) and total translated (*tt*) RNA-sequencing libraries

RNA-seq libraries were made using the *T* and *tt* fractions. For *T*, ∼200 mg of the frozen cell grindate was subjected to reverse crosslinking followed by RNA extraction generally as described above. The frozen cell grindate was thawed and clarified by centrifugation to remove cell debris at 4°C, 13 000 × g for 5 min. The resultant supernatant (∼150 μl, ∼3.0 AU) was then subjected to reverse crosslinking followed by RNA extraction as described above in the 'Construction of the eTCP-seq SSU, RS and DS footprint RNA-sequencing libraries' section. For *tt*, 200 μl (∼5.0 AU) of the filtered and sucrose-depleted (poly)ribosomal fraction from the 'Separation of the fixed (poly)ribosomal complexes away from the non-translated fractions of cytosolic RNA' section was used, and the reverse crosslinking performed as for *T*. ∼1 μg of the RNA obtained from each, *T* and *tt*, was directed into making the rRNA-depleted RNA-seq libraries employing GENEWIZ services. Generally, the rRNA was depleted in the RNA fractions and VAHTS Total RNA-seq (HMR) Library Prep Kit for Illumina-compatible sequencing were used for library preparation, during which the rRNA-depleted RNA was fragmented and reverse-transcribed. First strand cDNA was synthesized using ProtoScript II Reverse Transcriptase with random primers and actinomycin D. The cDNA second strand was synthesized using Second Strand Synthesis Enzyme Mix, which included dACG-TP/dUTP. The double-stranded cDNA was SPRI bead-purified and then treated with End Prep Enzyme Mix to repair both ends and add 3′ oligo(dA) in one reaction, followed by a T-A ligation to add amplification adapters to both ends. Size selection of the adapter-ligated DNA was then performed using SPRI beads, and fragments up to ∼400 bp (with the approximate insert size of 100–300 bp) were recovered. The dUTP-labeled second strand was digested with Uracil-Specific Excision Reagent enzyme. Each sample was then amplified by PCR using Illumina's P5 and P7 index-containing primers. The PCR products were cleaned up using SPRI beads and quantified by Nano quality control (Nano QC), to estimate library concentration. Next-generation sequencing libraries were then multiplexed and paired-end sequenced to 150 nt read length, to obtain at least 100 M reads per each barcode.

High-throughput sequencing and mapping of the reads

Sequencing was performed on a HiSeq 2500 (Illumina by GENEWIZ) using settings compatible with TrueSeq Small RNA Sample Preparation Kit (Illumina) and paired-end reads of 150 nt. Barcoded libraries were mixed in equimolar or otherwise desired proportion, and sequenced with 400 M reads per lane mode; additional lanes were invoked when necessary.

The 150 nt Illumina paired reads were subjected to quality control using Trimmomatic (SLIDINGWINDOW: 7 nt; phred quality cut-off: 24) followed by adapter trimming including the (A)20 tract from the reverse transcription primer. The reads containing no 3′ poly(A) or shorter than 17 nt were discarded. To assign reads to the genome a stepwise alignment strategy was performed as follows. Reads were first filtered to keep only those not mapping to any rRNA species (inferred from the locus: chromosome XII:450 000–491 000). The remaining reads were then mapped to a custom tRNA sequence set based on GtRNAdb (PMC4702915). Unmapped reads from this alignment were then aligned with spliced RNA sequences containing 'misc_RNA', 'ncRNA', 'tRNA' or 'snoRNA' primary tags in SacCer3. Finally, the remaining unmapped sequences were aligned to a genomic mRNA reference consisting of all protein-coding gene regions with an up- and downstream 1000 bp flanking genomic sequences, allowing multiple mappings.

Nanopore direct RNA sequencing and mapping of the reads

DRS libraries were prepared and run as described previously (83), using SQK-RNA002 kit and MinION Mk1B (ONT) equipped with R9.4.1 flowcells under control of MinKNOW v5.7.2. Raw FAST5 were collected and then basecalled into FASTQ using Guppy v5.0.11. The sequencing reads were aligned against the reference genome (Saccharomyces_cerevisiae.R64-1-1.108) from ENSEMBL using minimap2 with '-ax map-ont -k14' parameters. The mapped reads were converted into bam files with samtools. Featurecounts tool was used to count the reads based on the annotation provided by the ENSEMBL GTF file for'Saccharomyces_cerevisiae.R64-1-1.108'. The full R code, commands, parameters and software packages used in the subsequent processing to normalize and interpret the data are available upon request. Briefly, the sequencing passes were combined into biological replicates and variance stabilizing transformations (vst) were used for PCA plot generation. Raw read counts were normalized by the total read count and the scaling factor calculated as part of the DESeq2 package. Differentially abundant mRNAs across the fractions (RDT *versus* PS) and conditions (non-starved *vs.* 10-minute glucose starved) were determined based on alignment with a negative binomial distribution (P-value < 0.05, FC > 2) and cumulative read count >20 calculated using two biological replicates.

Footprint end definition and mapping

After quality filtering and adapter removal, the 3′ ends of the reads were used as reference points for the ribosome-protected footprints. In cases where the 3′ end position of the trimmed read sequences was aligned immediately upstream of adenine nucleotides, it was impossible to resolve the exact location of the 3′ end of the corresponding original RNA fragment before polyadenylation. With these reads, the 3′ end position was randomly assigned with equal probability between all possible locations. Assignment of start codon and ORF regions was performed as described earlier (76).

Length-based footprint classification and quantification

Footprints were defined by dissection of the footprint length distribution of the RS and DS eTCP-seq libraries with the range of 27–38 nt ('short') and 56–82 nt ('long'). Reads Per Kilobase per Million Mapped reads (RPKM) values were calculated for both paired RNA-seq fragments and ribosome-protected footprints.

Detection of non-random footprint peaks and calculation of the peak-free (stochastic 'base') footprint density

Putative stalling sites were detected by an in-house hidden Markov model-based algorithm using the R package depmixS4 (84). Both the original and the peak-segregated (*BASE* and *PEAKS*) signals of the eTCP-seq RS and DS were then used for downstream analyses. Please refer to the supplied source code for more details.

Calculation of the translation efficiency (TE)

'Translation efficiency' measures were calculated as the ratio of the normalized occupancy (RPKM) of eTCP-seq RS fraction and the RNA-seq from the poly(ribo)some co-sedimenting (T; TE^{T}) and the total cell lysate (tt; TE^{tt}) pools along the ORF regions using a 500-bp window both up- and downstream.

Definition and quantification of the features used for stochastic Translation Efficiency (STE) AI

STE measures were determined using the eTCP-seq SSU, RS and DS profiles. 'DS/RS' refers to the unfiltered, unprocessed and otherwise unmodified ratio of the normalized occupancies of RS (using only short footprint; 27–38 nt) and DS (using only long footprint; 56–82 nt). RS_{BASE} or DS_{BASE} denotes the average signal after peak removal from the ribosome or disome profile while RS_{PEAKS} or DS_{PEAKS} refers to the average signal of the predicted stalling sites. After peak removal, gaps were substituted by the average signal of the remaining codons. 5′ UTR scanning obstruction ratio was calculated as the number of SSU footprints within the 5′ UTR divided by the number of footprints within the 5′ UTR and over the annotated start codon. Initiation efficiency was derived from the ratio of the normalized SSU footprints divided by the normalized RS signal over the ORF region. Finally, elongation commencement efficiency was calculated as the ratio of the normalized RS signal around the start codon (the 5′ end of the footprint falls into the ± 7 codon window relative to the annotated start codon) divided by the average RS signal over the ORF region.

Ensemble machine learning and STE AI

An ensemble predictor for the real protein synthesis rate per mRNA per second, taken from Riba *et al.* (85), was built in two tiers. In the first tier, a Support Vector Machine and a Conditional Forest model were trained on two non-overlapping segments of the input data. The following features were derived and used: normalized SSU signal over the annotated 5′ UTR, normalized RS signal (only short, monosomal footprints), normalized DS signal (only long, disomal footprints), DS/RS ratio with and without footprint length filtering on the DS signal, and with or without removing the predicted stalling sites from the RS and DS fractions, 5′ UTR scanning obstruction ratio calculated as library normalized SSU signal within the 5′ UTR divided by SSU signal within the 5′ UTR and start codon-associated footprints, initiation efficiency ratio calculated as normalized RS quantified on the ORF region divided by the normalized SSU signal measured on the start codon ($\mathrm{RS}_{ORF}/\mathrm{SSU}_{START}$), elongation commencement ratio calculated as normalized RS quantified on the ORF region divided by normalized RS signal measured on the annotated start codon (RS_{ORF}/RS_{START}), and the annotated ORF length. In the second tier, another Conditional Forest 'governor' was trained using the above-mentioned features and the predictions from the tier 1 models. All models were 10-fold cross-validated and tested with different train-test ratio (0.5, 0.6, 0.7 and 0.8) and with different 5′ UTR length cutoffs (10, 20 and 50 nt). Contribution plots with the 'mean decrease in accuracy' were generated from the Conditional Forest models from tier 1 and tier 2 (governor model) using the varimp from the partykit R package (86).

Calculation of the differential $^{\delta}$STE, STE clustering and gene ontology (GO) term analysis

Differential $^{\delta}$STE (changes in $\log_2$(STE) values across conditions for the same gene or transcript) were calculated as $^{\delta}{}_1STE = STE_A/(STE_A + STE_B)$, $^{\delta}{}_2STE = 2 \times ((STE_A/(STE_A + STE_B)) - 0.5)$ and $^{\delta}{}_3STE = (STE_A - STE_B)/STE_B$. Heatmaps and *k*-means clustering were generated using the R packages ggplots2 and pheatmap (87). Gene Ontology and were performed using the R package ClusterProfiler (88). To construct sliding window Gene Ontology analysis, an in-house algorithm was used (please refer to the respective source code).

Calculation of the mRNA polysome abundance, polysome sedimentation factor and its difference

Library normalization was performed by excluding genes with less than 20 read counts cumulatively across the conditions and fractions, normalizing all non-zero counts by the library size followed by inputting all zero counts with 0.1 (gene polysome abundance; PA). The pairwise Pearson's correlations of the normalized library within replicates were calculated in R with Benjamini-Hochberg methodmethod for p-value adjustment and alpha set to 0.05. Polysome sedimentation factor was calculated as $PSF = PA_{PS}/(PA_{PS} + PA_{RDT})$, where PA_{PS} and PA_{RDT} were relative gene polysome abundances calculated as indicated above. The differential $^{\delta}$PSF is defined as $^{\delta}PSF_{S10} = PSF_{S10}/(PSF_{S10} + PSF_{NS})$. The correlations between the PSF across the conditions and $^{\delta}$PSF to $^{\delta}$STE (model 1 and model 2) were calculated by Pearson.

Mathematical modeling

The developed mathematical model of stochastic disome formation was verified by generating positions of ribosomes over 10 million iterations. Then the position of individual ribosomes was modeled using Gaussian distributions with identical standard deviations ($\sigma = 1$) and their relation was assessed in R using the stats package. The resulting probabilities and distributions were also verified in Mathematica (89). Line and boxplots were plotted using the ggplot2 R package.

Construction of *in silico* translatome and simulation of translation supported by *in vivo*-derived parameters

To generate the distribution of elongation rates over codons, the ribosome profiling-derived data (90) was fitted by log-normal distribution. A log-normal distribution modeling was also employed for initiation rate reconstruction using available experimental data (91). Maximum values for the initiation rate (number of initiating ribosomes per unit time) were restricted to 60 initiation events per min for each individual model mRNA. The *in silico* translatomes were modeled with a prototypical ORF consisting of 333 codons, using the fitted elongation speed profile. Simulations were run for 36000 model seconds (10 model hours) with a range of fitted initiation (μ = 1, 10, 20, 30, 40, 50, 60) and elongation rate distributions. 'Slow' and 'fast' segments were taken from the lower and upper quartiles of the elongation rate distribution, respectively, while the trimodal patterns of fast-slow-fast and slow-fast-slow were constructed with a middle segment of 7 codons (21 nt) surrounded by either a fast or a slow segment. Line, trace and box plots were generated using the R package ggplot2.

Published data used in the study

To generate a spectrum of elongation rates over codons, ribosome profiling-derived data were used (90). Initiation rate distribution was reconstructed using a maximum frequency-limited log-normal distribution fitted to available experimental data (91). To train and benchmark the developed ensemble models, protein synthesis rates were taken from a study that employed metabolic labeling and mass-spectrometry (85).

Results

Co-localization of ribosomes on mRNA as a result of diffusion-defined molecular motion

'Collided ribosomes' were described early on (57,62) and further characterized since, yet such nuclease-resistant co-localized ribosome occurrence (mostly in the form of disomes) is commonly only attributed to discontinuous elongation rates, as in ribosomal stalls or slowdowns on mRNA (39,60,63,92–97). Attempting to generalize this problem, we asked what a minimal prerequisite of ribosomal co-localization on mRNA might be (Figure 1A). Using a theoretical mRNA with infinite ORF length and featuring only two ribosomes, in our model a co-localization takes place when the trailing ribosome would attempt to elongate further on the message than the leading ribosome. Instead of predicating that ribosomal positions are deterministically defined, we used a more realistic assumption that diffusional heterogeneities of the elongation rate create a probabilistic function of ribosome localization, if two or more 'independent' ribosomes are considered (98–102). Simplifying this distribution to normal and identical in standard deviation for all ribosomes (with μ representing distance between ribosomes over mRNA defined as the inverse initiation frequency), this model results in a normal-like stochastic disome occurrence (see more in Supplementary Materials; Figure 1B and Supplementary Figure S1B, top lane). The model was further verified by averaging 10 million randomly-generated pairs of ribosome positions and identifying the frequency of corresponding disome positions (Supplementary Figure S1B, bottom lane). Intuitively, stochastic disome probability maximum does not exceed 0.5 for two ribosomes. The disome probability also increases monotonously albeit non-linearly with ribosome density (*i.e.* initiation frequency), and is dependent on the localization variance (σ), which cumulatively embodies diffusional variances of all contributing processes (Figure 1C, Supplementary Figure S1A). Interestingly, increasing the number of ribosomes does not principally change these characteristics, although it results in higher maximal stochastic disome frequency (Supplementary Figure S1C).

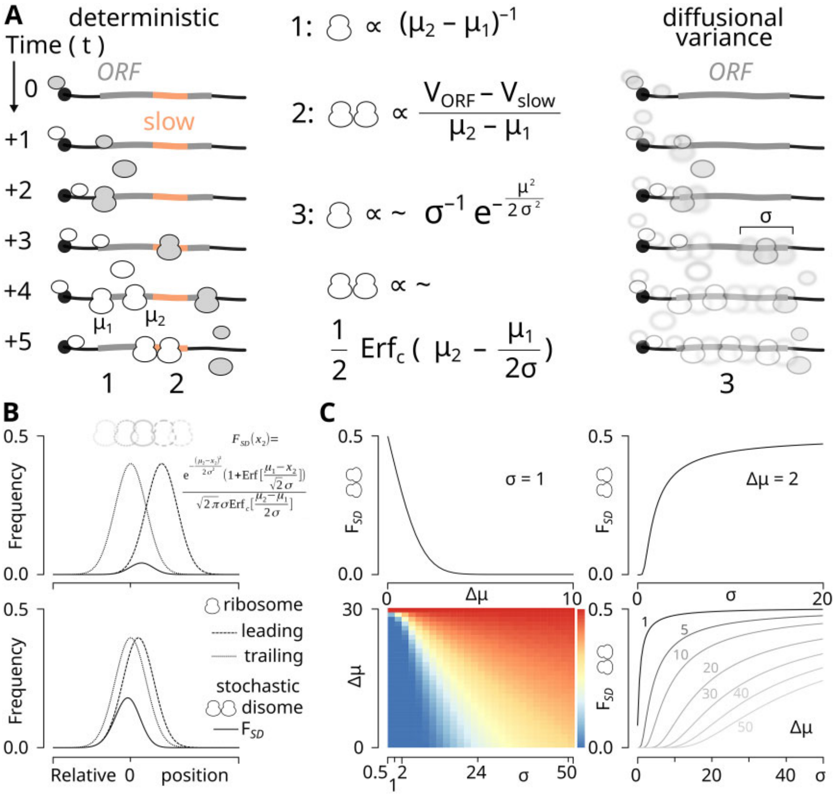


Figure 1. Stochastic co-localized ribosomes (stochastic disomes, DS) result from diffusional initiation and elongation rate variances. (**A**) Comparison of the common 'deterministic' view (*left*) of ribosomal (RS) progression on mRNA with a diffusion-aware stochastic model (*right*). Note that in the deterministic model, co-localized ribosomes can only occur at certain sites with slow elongation rate or arrested translation. μ represents the distance between ribosomes along mRNA, σ represents the localization variance and $\mathrm{Erf_c}$ is the complementary Gaussian error function (see text for more explanations). V_{ORF} and V_{slow} indicate the elongation rate (velocity) for the whole or a designated part of the ORF, respectively. (**B**) Generalized mathematical model of stochastic disome occurrence (stochastic disome frequency, F_{SD}) for an infinitely long model mRNA containing two RS (see more under 'Mathematical modeling' in Materials and Methods). (**C**) F_{SD} for case as in (B) monotonously depends on initiation frequency across all variance values.

The model can be extended to accommodate multiple ribosomes in two ways. First, we can examine the probability of disome occurrence when multiple ribosomes elongate on the same mRNA (Supplementary Figure S1C). This analysis reveals that when more than two ribosomes are present, the probability of disome appearance is substantially higher, consequential to the high translation initiation and elongation rates. The apparent disome frequency dependence on the initiation frequency does become more complex (Supplementary Figure S1D; compare graphs with just two ribosomes, green, and all five ribosomes, magenta), but remains monotonous. Second, we can examine the appearance of multisomes such as trisomes, tetrasomes and pentasomes (Supplementary Figure S1D), when *e.g.* up to five ribosomes are simultaneously present on the mRNA. As revealed by this analysis, although multisomes do occur when multiple ribosomes elongate on the same mRNA, the probability of their occurrence is relatively low compared to the disomes, even under favorable rate and variance combinations. In contrast, stochastic disomes are quite frequent and, normalized by the average number of monosomes, can provide a proxy of translation rate, primarily driven by the translation initiation frequency. An interesting additional feature revealed by this analysis is the 'unimodal' disome occurrence in the case of tightly packed (rapidly initiated) ribosomes and high localization variance-to-initiation frequency ratio, which transitions into a broader disome occurrence distribution, as the packing becomes more loose (Supplementary Figure S1D) and the variances shrink, and further transforms into a multi-modal disome frequency, where modes are attached to the nearby ribosomal frequency peaks. Thus, we can conclude that stochastic disome formation may take place even under the least favorable combinations of initiation rate and cumulative diffusional variance. In this latter case, ribosomes that happened to initiate in a more rapid succession would keep traveling on mRNA as a group and stochastically co-localize, still providing a time-averaged 'base' disome signal.

While we modeled the stochastic-aware ribosomal behavior along an mRNA ORF using a continuous probability distribution function (for simplicity; which nonetheless can be numerically binned into a Poisson-like distribution), it can be useful to apply some constraints to this model. For example, μ_i can be limited as more or equal to the $\mu_{i(min)}$ of, *e.g.* $2\sigma_i$, to reflect a steric inability of the SSUs to occupy the same space on mRNA and not to be loaded until after the previous SSU sufficiently clears the mRNA 5′ end. To further incorporate the effects of translation initiation, the model can be expanded by unfolding the probability of the SSU start codon occurrence through a more complex function defined by the SSU 5′ cap attachment rate and its variance, SSU 5′ UTR scanning speed and its variance, start codon recognition speed and efficiency and their variances, and incorporating an effect of the possible 5′ UTR length difference. These changes to the model are outside the scope of the work, and while theoretically intriguing, will not principally affect our conclusions, similar to the effect of a possible crosslinking delay-induced 'leakage' of the complexes into the later stages. Another limitation stems from the steric clash phenomenon and a finite length of mRNA, which is generally within a few orders of magnitude of the ribosome-occupied fragment lengths. For example, ORFs, excluding uORFs, have a median length of 1000 nt, based on the current yeast genome annotation (SacCer3; https://www.ncbi.nlm.nih.gov/assembly/GCF_000146045.2/). Arguably, with the commonly accepted highest possible ribosomal density of ∼0.1 codon^{-1} (indirectly confirmed by footprinting assays performed by us and others (62,76,103–105), as well as by translation rate estimates and modeling (85,106,107) and electron microscopy (108)), these ORFs could maximally simultaneously contain up to thirty-three ribosomes. An intriguing conclusion is that with the finite density of ribosomes over a given mRNA, the DS probability function has a distinctive shape reflecting both the density and the absolute number of the ribosomes over that mRNA. Therefore, theoretically, the distribution of DS probability of occurrence can be used to infer variations in the absolute load of an mRNA with ribosomes at steady-state, identifying how many ribosomes are present on that mRNA at a given moment in time, and further, deconvoluting fractions of the given mRNA by their absolute occupancy with the ribosomes. Undoubtedly, including these finer effects into the modeling would make it more realistic, but the model would also lose its generality and more complex distributions would have been introduced. To overcome these limitations, we further employed a simulation-based modeling framework using experimentally available initiation and elongation rate distributions.

Stochastic-aware simulation of ribosomal dynamics predicts an elongation rate-independent disome population and links disome occurrence with translation initiation rate

To attest to the potential of stochastic co-occurrence of ribosomes in a more lifelike scenario, we constructed a parameterized model of protein synthesis, somewhat inspired by the classical solutions to road traffic, Monte Carlo simulations of translation, M/M/1-queues and totally asymmetric simple exclusion process (TASEP; and its derivatives) (Figure 2A) (57,85,101,102,107,109–112). We wished to test if our model can result in ribosome co-occurrence within the parameters of key steps relatable to the actual biological processes of translation. Recent single-molecule studies provided quantification of both, the initiation event frequency and elongation speed, and gave an estimate of their variance (85,90,113–119). Precise measurement of single-molecule events during, *e.g.* certain steps of translation initiation and elongation, using FRET approaches, provide a direct visualization of the stochasticity of these events, including the direct estimation of translation initiation variance (114,117). Recently, it has been shown that stochasticity during translation initiation can alter subsequent elongation rate and result in cell fate switching (120). Other types of experiments are well-suited for *en masse* observations, such as in multi-tagged fluorescent systems combined with super-resolution microscopy (91,113,121), or the analysis of deeply sequenced ribosome footprint libraries (90). From these data, it is possible to obtain a probability density function reflective of the cumulative *in vivo* translation initiation and elongation rate variances, which also comprises the diversity of the tRNAs (122–124) and their kinetics as well as diffusional variances.

Average translation initiation rates have been shown to vary greatly, with observations of as low as < 0.1 ribosomes per minute up to ∼15 ribosomes per minute (thus maximally allowing each ribosome to attach every 4 seconds), with the median of about 1 ribosome per minute and the bulk between 0.5 and 2 ribosomes per minute at 37°C (91). Similarly, elongation rate variance expressed as the variance of the calculated ribosomal dwell time over each codon based on ribosome profiling data (90) demonstrates a skewed normal-like distribution with a modal value around 200 ms per codon (∼5 codons/second) at 30°C. As each ribosome covers a space on mRNA of ∼10 codons, it would take ∼2 s for it to clear the occupied space at the modal speed. These constrains would be difficult to precisely incorporate within a generalized representation of translation but are sufficient to construct a reasonably-performing numerical model.

To generate the distribution of elongation rates over coding triplets per second, ribosome profiling-derived data was used (90) and modeled by log-normal distribution (Figure 2A, Supplementary Figure S2A). A maximum frequency-limited log-normal distribution modeling was further employed for initiation rate reconstruction using available experimental data (91) (Supplementary Figure S2B), for the range of expected values ($\mu_i = 1, 10, 20, 30, 40, 50, 60$; Supplementary Figure S2B), allowing a maximum of 60 initiating ribosome attachments per minute per each mRNA and including non-deterministic diffusional variance component for each of the set average frequencies.

We next created a translation model roughly based on M/M/1 queue principles (125), using a prototypical ORF consisting of 333 codons and initially with a homogeneous elongation speed profile taken from the reconstructed distribution of elongation rates (Supplementary Figure S2A). We used a feedback-independent model with equal and unrestricted availability of initiating ribosomes, no specific delay in conversion of the initiating to elongating ribosome and elongating to terminating or recycled ribosome, and employed a second-based codon-wise iteration starting from the 'empty' mRNA. The system was allowed to cycle over 36 000 model seconds (10 model 'hours'). We then varied the modal SSU attachment (initiation) and its distribution function respectively, by scaling the entire distribution towards faster or slower initiation, and investigated the initiation rate effects using spatio-temporal ribosome positioning diagrams termed trace plots.

We first created scenarios with an unbiased uniform (Figure 2B, $\mu_i = 1, 10, 60$; more in Supplementary Figure S2C) elongation speed (codon) pattern, and next with more complex codon patterns including 'fast', 'slow', 'fast-slow', 'slow-fast', 'fast-slow-fast' and 'slow-fast-slow' code stretch transitions across the ORF (Figure 2C, 'fast-slow-fast', 'slow-fast-slow'; more in Supplementary Figure S3A, top lane). The 'slow' and 'fast' segments were sampled from the lower and upper quartiles of the elongation rate distribution, respectively (Supplementary Figure S2A). The bimodal ORFs had their 'slow' and 'fast' segments split at the ORF midpoint, while the trimodal ORFs had their middle segment spanning 7 codons (21 nt) positioned centrally. This analysis revealed that the disome (DS) to ribosome (RS) ratio strongly correlates with the number of initiating ribosomes, independent of the elongation rate structure over the ORF (Figure 2B, bottom lane).

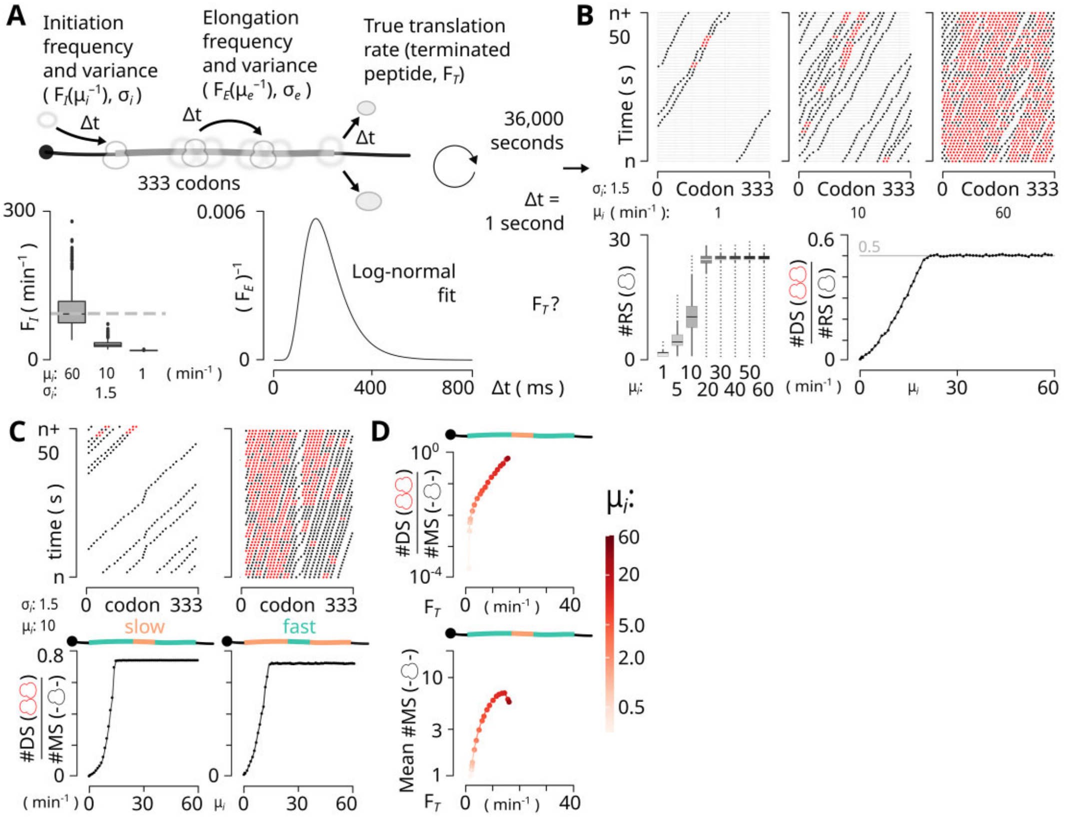


Figure 2. Stochastic co-localized ribosomes (stochastic disomes, DS) are monotonously linked to translation initiation rate. (**A**) Computational diffusion-aware model of mRNA translation accounting for the experimentally-determined distributions of initiation and elongation rates and variances (frequency of initiation F_I and frequency of elongation F_E, respectively, see more under 'Construction of *in silico* translatome and simulation of translation supported by *in vivo*-derived parameters' in Materials and Methods). (**B**) Model from (A) realized across different median initiation rates and using a homogeneous F_E log-normal distribution of codon dwell times (elongation dynamics) for the entire simulated Open Reading Frame (ORF) of 333 codons. Temporal (Y-axis) traces of individual ribosome positions on mRNA ('trace plots') for slow, medium and fast initiation rate cases are shown across the consecutive mRNA codons (X-axis) (*top*). Note that even a slow initiation rate with sparsely situated RS results in stochastic co-localizations. Steady-state number of RS over the model 333-codon ORF is dependent on the initiation rate, until saturation is reached (*bottom left*). Disome to monosome (DS to MS; whereby MS is a singular ribosome) ratio at steady state unambiguously depends on the initiation frequency (*bottom right*). (**C**) Model realizations as in (B), but for mRNAs with heterogeneous F_E, resulting in various patterns of average codon dwell times, one with a first 'fast' segment (sampled from the higher quartile of the elongation rate distribution), a 'slow' segment (sampled from the higher quartile of the elongation rate distribution) in the middle representing a 'stalling' or slowdown site, and a third 'fast' segment (*left*), and one with the inverse pattern (*right*). (**D**) Translation rate (protein yield as determined by the terminated peptide frequency F_T) retains monotonous dependency on DS to RS ratio (*top*), as opposed to the mean monosome (MS) density on mRNA (equivalent of the normalization of RS to RNA signals in experimental ribosomal profiling data) (*bottom*).

Finally, we systematically investigated how the average number of singular ribosomes (ribosomes that are not part of a di- or multisome), which we refer to as monosomes (MS), as well as DS, and the DS to MS ratio (Figure 2D, 'fast-slow-fast'; more in Supplementary Figure S3A and B) depend on the overall translation rate (the number of terminated ribosomes/second, F_T). This analysis revealed that mean monosomal frequency (MS frequency; a modeling equivalent of the ribosomal footprint signal widely used in calculations of the Translation Efficiency or TE) shows ambiguous mapping for the multiple cases of elongation modalities, prominently exemplified by the fast-slow, fast-slow-fast transitions and the generally slow ORFs, where a given TE value could be mapped to more than one protein biosynthesis rate F_T (Supplementary Figure S3A and B). The DS to MS ratio, in contrast, demonstrated a strong correlation with the protein biosynthesis rate, provided near-linear response across substantial range of biosynthesis rates for all tested variations of the elongation modalities, and was generally comparable in performance (although somewhat different) to the ideal case of theoretical ribosomal frequency over ORFs measured as mean RS (Supplementary Figure S3A and B).

In summary, our simulation and its evaluation suggest that the DS to MS ratio can be used as an important contributor in the measurement of translation rates, and that the stochastic DS_s to MS ratio is unambiguously correlated with the translation initiation frequency. Critically, unlike in other currently employed techniques, the disome to monosome ratio is based on values of a similar nature, which avoids biases deriving from methodological and instrumentational differences and, importantly, provides a capability to directly quantify absolute protein biosynthesis rates.

Mechanisms of ribosome co-localization on mRNA

Having confirmed the theoretical plausibility of stochastic ribosome co-localization on mRNA, we further considered other possible reasons for ribosome co-localization on mRNA in actual biological systems. Multiple datasets now exist that describe the occurrence of co-localized ribosomes (most commonly, in the form of disomes, sometimes multiple 'stacked' ribosomes) on mRNA (39,40,60,63,92,126–128). Many of these datasets are resulting from some form of limited digestion of the ribosomal fraction and the occurrence of the longer-than-monosomal nuclease-protected fragments of mRNA, although electron microscopy observations are also available (129). The most frequently quoted and intuitive mechanism of disome occurrence is the one associated with a localized unevenness of the translation elongation rate, a phenomenon of multiple possible origins. For example, codon usage and aminoacyl-tRNA availability, specifics of the kinetics of each of the codon:tRNA interactions on the ribosome, local nucleotide structure, nascent polypeptide properties and the nascent protein domain structure have been implicated as factors leading to non-uniform elongation rates (39,63). Thus, these factors explain the possibility of a non-random DS_{nr} occurrence at specific positions of elongation rate slow-down. It can be speculated that within non-stabilized polysomes, and polysomes stabilized with specific translation elongation inhibitors, detectable disomes (and any other co-localized ribosomes) are indeed, to a large extent, resulting from such elongation slow-downs (4).

Based on our theoretical considerations as outlined earlier, and the possibilities of additional complex ribosome-to-ribosome contact stabilization by *in vivo* rapid unbiased crosslinking (4,5,78), we propose the existence of at least three other classes of co-localized ribosomes (mostly, disomes DS; Figure 3B), in addition to the DS_{nr}. One class of stochastic disomes (DS_s) might derive from the variances of the underlying chemical reactions, based on the diffusion-derived variances and the probabilistic nature of the chemical reactions themselves. Further, disomes can exist that, despite relating to non-contiguous regions of the mRNA, become crosslinked due to their spatial proximity. These spatial proximity disomes could either reflect specific non-random spatial arrangements within the polysome (DS_{sp}), or derive from stochastic events in the polysome as a result of random (diffusional) coincidences (DS_{rsp}) (Figure 3B). For example, flexing of the polyribosomes resulting in random contacts between distant individual ribosomes or random *trans*-polysomal contacts between ribosomes in crowded environments could lead to the DS_{rsp} formation. Each of these classes of co-localized ribosomes carry useful information about the translation rate, spatial organization and locality of the polysomes. While the DS_{sp} and DS_{rsp} classes of co-localized disomes remain outside the scope of this work, it is important to note that both are composed of ribosomes separated by substantial gaps along the mRNA, and thus can be identified and filtered out by their monosomal footprint length (Figure 3B).

Footprint signatures of stochastic co-localization in translation complex profiling data

To experimentally address the possible occurrence of diverse types of 'disomes' *in vivo*, we employed an enhanced version of translation complex profile sequencing (eTCP-seq) (5,76–78). We used rapidly-crosslinked translational complexes separated into total RNA (*T*), total translated RNA (*tt*), polysome-engaged ribosomal small subunit (SSU), ribosome (RS) and disome (DS) fractions (Figure 3A), based on their distinct sedimentation profiles (Supplementary Figure S4B).

In the polysomal material stabilized *in vivo* by rapid crosslinking, we can detect a fraction of polysomes that are highly resistant to nuclease treatment, similar to prior observations (5,39,60,62) (Supplementary Figure S4A and B). This fraction yields mostly disomes, but also contains some higher-order polysomes such as trisomes (Supplementary Figure S4A). Bioinformatic analysis (Supplementary Figure S4C) revealed that the DS fraction prominently featured longer footprints (Supplementary Figure S4D), whereas SSU and ribosome fractions mapped in accordance with the previously-published results (Supplementary Figure S4E) (76,78).

In-line with the principles outlined above for the possible different 'disome' classes, our bioinformatics analysis revealed a complex footprint length spectrum in the nuclease-resistant disome data (Supplementary Figure S4D and E). Disome footprint data contained a considerable fraction of monosomal-length footprints, both short (A-site accessible/'rotated', ∼21 nt) and normal (∼32 nt) monosomal footprint versions (4) that are reflective of the different sub-steps of the translation elongation cycle (130). These footprints can belong to the DS_{sp}, DS_{rsp} classes, as well as to the disome classes that protect a longer footprint, which has been subsequently cleaved while preserving the disome arrangement through other chemical crosslinks between the ribosomes. Disome data also contains long footprint (∼64 nt) versions that can only be a result of protection from consecutively and compactly localized (also often referred to as 'stacked', sometimes 'collided') pairs of ribosomes on mRNA. The long footprints can result from the non-random DS_{nr} and the stochastic DS_s classes, in which the possible configurations of the ribosomes can be also diverse, resulting in all possible combinations, short-short (∼42 nt), short-long/long-short (∼52 nt) and long-long (∼64 nt) of the footprints (Supplementary Figure S4D). To ensure unambiguous attribution for the downstream analyses, we kept only monosomal footprints (∼32 nt) for ribosomes and long footprints (∼64 nt) for disomes (Supplementary Figure S4D and Supplementary Figure S5A and B). It is noteworthy that upon such filtering, the DS to RS correlation decreased, pointing towards a complex relationship between the DS and RS footprint versions reflecting a narrower and more precisely defined translation complex spectrum (compare Supplementary Figure S5A and B).

A characteristic and visually prominent feature of the well-covered ORFs upon filtering short footprints away is a combination of peaks of long footprint coverage with some 'background' coverage that is relatively evenly distributed across the ORF body (Figure 3C, Supplementary Figure S6). We interpret the local peaks DS_{PEAKS} of the coverage profiles as representing the DS_{nr} population, as it has been suggested before (60,92). To systematically identify these peaks in a background-considerate fashion, we used a hidden Markov model splitter algorithm to detect and remove putative stalling sites from both RS and DS profile (see Materials and Methods for more details; Figure 3B and C, Supplementary Figure S5C and D for the representative results of such transformation). We then posited that the peak-excluded, long disome footprint-length 'background' coverage DS_{BASE} represents the DS_s population (Supplementary Figure S4B–D). The DS_s coverage is variable across different mRNAs, likely reflective of their differential translation activity (Figure 3B, Supplementary Figure S6). Furthermore, the peak-excluded DS_{BASE}/(short RS) ratio is also variable across different mRNAs, bolstering our modeling conclusion about the non-linear dependence of DS_s on the translation initiation rate (Supplementary Figure S5E and F). Further, comparison of the

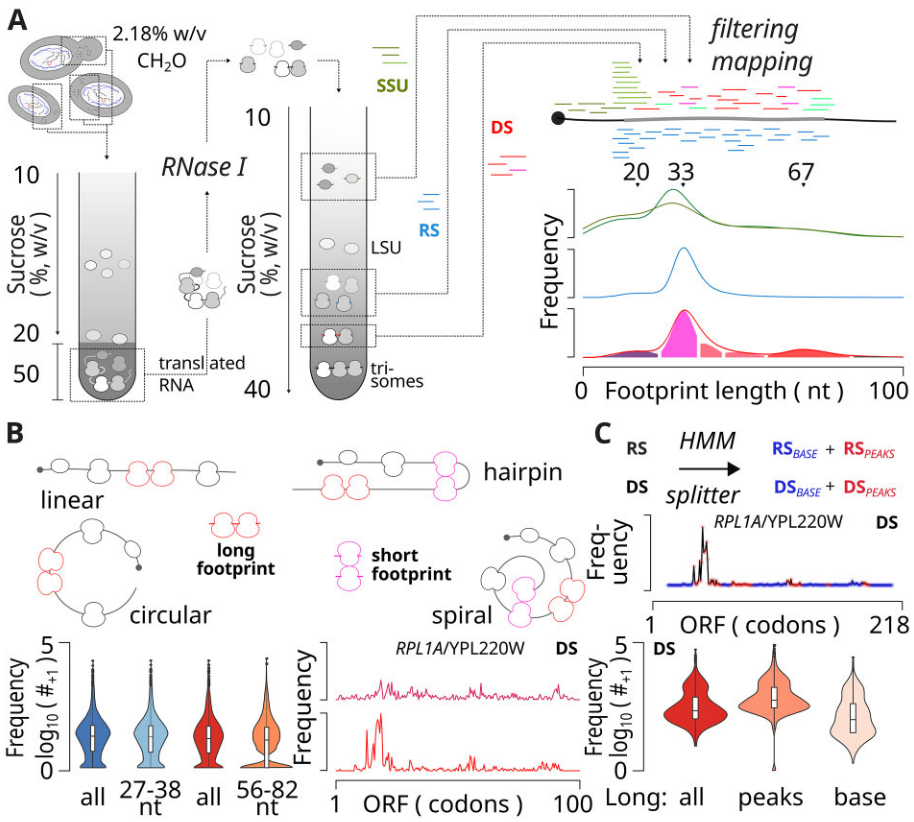


Figure 3. Enhanced translation complex profile sequencing (eTCP-seq) identifies a comprehensive range of translational intermediates, including stochastic and non-random ribosomal co-localizations on mRNA. (**A**) Outline of the enhanced eTCP-seq approach underpinning STE calculations. Schematic illustrating the sedimentation-based separation of total translated RNA (RNA^{tt}), and its RNase I cleavage and fractionation to generate small ribosomal subunit (SSU), ribosome (RS) and disome (DS) footprints (*left*). Footprint mapping from different fractions and the respective fragment length distributions are shown (*right*). Note the presence of an additional peak with ∼67 nt mode in the DS fraction. (**B**) Features and theorized origins of the DS footprints. Schematic showing plausible polysome arrangements and different potential types of contacts between RS, either distal or proximal to each other along the mRNA. The former generate localized 'short' footprints associated with the DS fraction, while the latter yield 'long' DS footprints resulting from either stochastic or non-random proximal co-localization, as well as 'short' footprints due to their partial cleavage in halves (*top*). Many transcripts have RS (blue) and DS (red) fractions with substantial frequency of footprints in the characteristic 27–38 nt ('short') and 56–82 nt ('long') length ranges, respectively (*bottom left*). DS footprints filtered into 'short' (purple, top) and 'long' (red, bottom) groups for a well-covered *RLP1A* mRNA ORF demonstrate partially concordant coverage, with some peaks appearing as unique to either group (*bottom right*). (**C**) Use of a hidden Markov model-based (HMM) splitter to detect (and isolate) peak (non-random) and base (stochastic) signal components (*top right*). Many transcripts demonstrate the presence of a large stochastic component in the long DS footprint signal (*bottom right*). Note that the length-normalized peak signal (as shown) is inherently higher than the base signal due to ribosome 'concentration' at the specific stall or slowdown sites over a short region of mRNA.

footprint-filtered DS_{BASE} to short RS ratio using metabolic labeling and mass-spectrometry (85), revealed strong positive correlation to the protein output for mRNAs where there is a detectable disome signal, directly confirming our theoretical assumptions (Supplementary Figure S5G; upper group).

Validation of the stochastic co-localization signal component in *in vivo* footprinting data

We next identified the stochastic component of the disome footprints, to use it for protein output modeling as theorized above. We first isolated 'long' (56–81 nt) footprints for the disomal and 'short' (27–38 nt) footprints for the ribosomal fractions (Supplementary Figure S5A and B). Next, a hidden Markov model (HMM) splitter was used to identify and remove disome and ribosome 'spikes' (RS_{PEAKS}, DS_{PEAKS}), which likely descend from localized stalls or slow-downs and thus could be a compounding factor for stochastic disome signal fitment due to their non-randomness (Supplementary Figure S5C and D). Of note, ribosomes and disomes found outside peak regions (RS_{BASE}, DS_{BASE}) constitute a considerable proportion of all signals (83% for the RS fraction and 23% for the DS fraction), and the correlation between RS and DS signal gradually decreases with every next filtering step demonstrating that DS_{BASE} signal contains a different type of information (Supplementary Figure S5A–D). As the long disome footprints are virtually absent in the RS data (Figure 3A), the remaining signal directly supports the stochastic appearance of ribosomal co-incidence and initiation rate-accidental tight packing (Figure 3C; see more in the Supplementary Materials).

Upon all filtering, we inferred the disome to ribosome (DS/RS) signal (Supplementary Figure S5E and F) and compared it against known accurate measurements of protein synthesis rates performed in *S. cerevisiae* using metabolic labeling and mass-spectrometry (85). Importantly, length-filtered DS_{BASE}/RS values returned strong positive correlation to the protein output for the mRNAs with a robust stochastic disome signal, directly confirming our theoretical conclusions (Supplementary Figure S5G and H).

Ensemble machine learning featuring stochastic disome signals accurately predicts translation rates *in vivo*

We then employed an ensemble learning, featuring a cross-validated Support-Vector Machine (SVM) and Conditional Forest models, which used inputs from ORF length-normalized footprint densities of different fractions, and various measures of translation including DS/RS ratios, initiation and elongation efficiencies (see more in Supplementary Materials), and the ORF length. These two 'first tier' models were trained on non-overlapping segments of the data set. We next trained another (second tier) Conditional Forest model termed 'governor', using the predictions of the first-tier models. The governor returned ∼0.6–0.7 correlation with the mass-spectrometry-based measurements benchmarked with an independent data subset not used in the training process (30% of the whole data set). Surprisingly, we could not achieve a meaningfully-positive correlation when using ribosomal footprint counts normalized to either total or total translated RNA-seq signals (TE^T, TE^{tt}; Figure 4A). Remarkably, of all components used, the length-filtered DS_{BASE}/RS ratio, and the normalized ribosome signal, showed the highest contribution to the predictive power of the model (Supplementary Figure S7C), while the inclusion of ORF length demonstrated only moderate improvements (not shown).

To confirm the generalization capability of our approach, we tested models with different training:test ratios (0.6 or 0.7) and minimum 5′ UTR length parameters (5′ UTR > 10 or 5′ UTR > 20 nt) (Supplementary Figure S8E). Both governor models resulted in similar accuracy in predicting protein synthesis rates and defined mRNA (gene) clusters with similar biological/molecular functions. Because the DS_{BASE}/RS diffusion variance-derived component is unique to our modeling, we deemed our measure as 'Stochastic Translation Efficiency', or STE. Importantly, STE provides a direct estimate of the absolute protein output per mRNA per unit of time, and thus can be used to rank mRNAs and respective UTRs by their power to compete for initiating ribosomes in any given condition.

Stochastic translation efficiency (STE) reveals details of rapid re-prioritizing of cellular translation in stress

Applying STE to a protoptypical scenario of nutrient starvation (10-minute glucose depletion) in *S. cerevisiae* (Supplementary Data S1 and S2), we observed gene clustering by response type and magnitude with finer details of gene expression control revealed. In the non-starved cells, 72 genes appeared in the top 25% and 4446 genes – in the bottom 25% of STE values, with maximal F_T values approaching 0.055 p/s, (model #1; Figure 4B and Supplementary Data S1). mRNAs of ribosome biosynthesis (RiBi), sugar and amino acid metabolism and biosynthesis of secondary metabolites genes were enriched in high STE, whereas those of endocytosis appeared translationally suppressed (Figure 4C and Supplementary Data S1 and S2). In the starved cells, RiBi transcripts were translated less efficiently (variably), amino acid and secondary metabolism transport and stress granule-related transcripts were substantially inhibited, while endoplasmic reticulum (ER) membrane-related mRNAs accelerated their translation (Figure 4C and Supplementary Figure S9A and B, also Supplementary Figure S10A and B).

Genes involved in oxidative metabolism balance, such as re-oxidation of intracellular NADH under anaerobic conditions, glycolytic breakdown of carbohydrates into pyruvate, glycolysis and gluconeogenesis (*OSM1*/YJR051W, *TPI1*/YDR050C, *PGK1*/YCR012W), as well as ER-associated protein degradation (ERAD) and SSU production (*DFM*/YDR411C, *RPS21A*/YKR057W), were highly translationally up-regulated, underpinning the importance of cytoplasmic gene control in nutrient response. A signature of accelerated transcriptome restructuring was observed, where RNA metabolism genes (rRNA maturation, transcription and gene expression) were up-regulated, whereas general translation, nucleotide and ribose phosphate/carboxylic acid metabolism were down-regulated (Figure 4C and Supplementary Figure S9A and B, also Supplementary Figure S10A and B).

To enhance the robustness of our conclusions and enable comparisons with alternative interrogation methods like polysome profiling, we incorporated nanopore direct long read sequencing (DRS) and qPCR analyses. This allowed us to evaluate the identical conditions of exponential growth and a 10-minute glucose starvation (Supplementary Figure S11A). For the qPCR, we selected targets with various $^1STE_{NS}$ that are not implicated in ribosomal pausing/stalling and demonstrate a simple, unobstructed scanning pattern (76,79). DRS offers exact matching of read counts to the mRNA molecule counts and is devoid of RNA coverage artifacts or amplification biases, while spike-in-normalized qPCR allows comparisons of absolute transcript levels. We included total clarified cell lysate (CCL) fraction for the qPCR analysis and further employed one of the traditional comparisons between frequencies of mRNA (polysome abundance, 'PA') in the heavier polysomal fraction (tetra-, pentasomes and further polysomes, 'PS' fraction) and light fractions (mono-, di- and trisomes, and their combined 'RDT' fraction), which removes a concern for the accessibility of mRNA to translation (Supplementary Figure S11A).

qPCR results confirmed our expectations towards a generally reduced abundance of mRNA for most of the genes tested, signifying transcript destabilization in the glucose-starved condition (Supplementary Figure S11B). Importantly, these results also confirm expectations of differential destabilization: while mRNAs such as *SSC1* showed only a minor abundance reduction, *GUS1* and *GCN2* were > 50-fold depleted (Supplementary Figure S11B). Interestingly, *ATG40* mRNA exhibited robustly higher PS to TS change in the starved condition, confirming translational acceleration in glucose starvation as measured by STE (and DRS; see Supplementary Data S1–S6).

DRS data exhibited excellent correlation (Supplementary Figure S11D) and highly similar gene count saturation histograms across biological replicates (Supplementary Figure S11E), replicate grouping and condition/fraction separation in the PCA (Supplementary Figure S11F). Nonetheless, upon calculation of the Polysome Sedimentation Factor (PSF) expressed through the PA_{PS} and PA_{RDT} ratio ($PSF = PA_{PS}/(PA_{PS} + PA_{RDT})$) as a proxy of translational engagement (similar to previous studies (35)), PSF exhibited high dispersion (Supplementary Figure S11G) and thus confirms our conclusions that frequency normalization-based techniques lead to less robust observations. While the overall reduction of the heavy polysome occupancy was apparent in the starved cells' PSF_{S10} against the non-starved counterparts

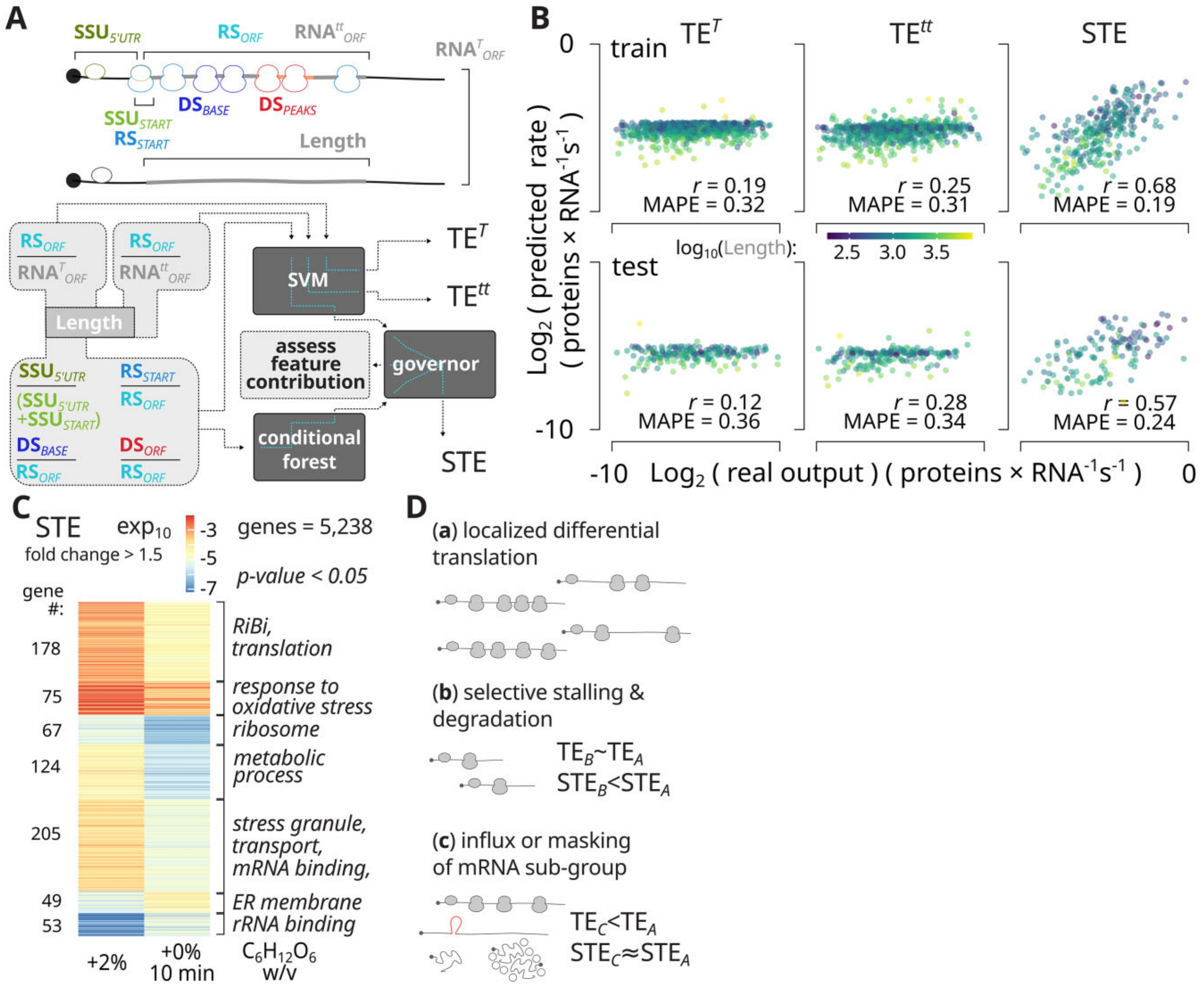


Figure 4. Stochastic translation efficiency (STE) measure maps to experimentally-determined absolute protein biosynthesis rates and dissects classes of mRNAs with different power in translation. (**A**) Schematic of ensemble learning approach featuring a Support Vector Machine (SVM) and a Conditional Forest (CF) model (first tier), separately trained and cross-validated on two, non-overlapping subsets of the training data. These two models are then used by another Conditional Forest model termed 'governor' (second tier), which provides conditional scoring and votes for the results of the first tier (see 'Ensemble machine learning and modeling of STE' in Materials and Methods for further details). All features used to quantify STE are like-to-like normalized values, as opposed to the translation efficiency-like TE^T and TE^{tt} calculated by normalizing RS ORF signal to total (*T*) or total translated (*tt*) RNA-seq signals, respectively. Note, all models include ORF length as an input. (**B**) STE demonstrates high correlation with experimentally-defined protein synthesis rates (metabolic labeling-based quantitative mass spectrometry data (85)). (**C**) Classes of translationally up- or down-regulated mRNAs during yeast response to a 10-minute glucose depletion as determined by STE. The classes were inferred using *k*-means clustering of mRNAs with STE fold-change of at least 1.5 between the non-starved and starved states. (**D**) Schematic illustrating robustness of the STE measure of translational power in a background of rapid transcription-, degradation- or masking- and phase separation-induced alterations of mRNA abundance and availability to translation, allowing to isolate mRNA-specific, UTR-driven translational control *in vivo*.

(PSF_{NS}; Supplementary Figure S11H), PSF, unlike STE, did not correlate well with the isotope-measured protein biosynthesis rates, and exhibited a substantial open reading frame (ORF) length bias, as can be expected (Supplementary Figure S11I). To reduce the ORF length bias, PSF can be compared transcript-wise across conditions by introducing the relative contribution $^{\delta}PSF_{S10} = PSF_{S10}/(PSF_{S10} + PSF_{NS})$. $^{\delta}PSF_{S10}$ exhibited strong correlation between replicates (Pearson correlation coefficient 0.76; Supplementary Figure S11J) and indeed showed a positive correlation with similarly defined $^{\delta}STE$, albeit with groups of genes separating away from the regression line (see Supplementary Figure S12 and Discussion for further dissection of these groups). Altogether, these results provide a validation and highlight the advantages of the STE approach.

Discussion

In this study, we identified new types of co-localized ribosomes on mRNA, stochastic and spatial disomes, in addition to the commonly-appreciated, 'collided' disomes resulting from stalling or local elongation slow-downs (92). The stochastic disome signal, together with the other measurements derived from eTCP-seq, link the absolute translation initiation and protein biosynthesis rates, which we project as an AI-based measure. STE AI is based on a comprehensive snapshot of positional information for all *in vivo* translational complexes, presented to ensemble machine learning using biophysical and bioinformatic segregation of the types of complexes and phases of translation cycle of mRNA, whilst retaining diversity of the profiling data. This robust measure can be used to define translation initiation-controlled protein biosynthesis rates for each mRNA comparatively across conditions as well as absolutely in a single condition, independent of RNA abundance and variations. Only like-for-like footprint signals are used in STE AI, canceling internal biases associated with normalization to different signal types. With STE, we demonstrate that yeast translational control invoked by glucose depletion is more complex than previously thought and includes translational acceleration as well as RNA metabolism adjustments to accommodate new protein biosynthesis demands.

Importantly, STE is based on fundamental principles of translation complex distribution and does not need to account for specific sequences and structures of mRNA, enabling universal application. STE AI, and the underlying eTCP-seq data also revealing mechanistic features of translation of individual mRNAs, can be used to inform synthetic biology and mRNA therapeutic platforms with accurate translational rate predictions to facilitate new mRNA designs.

STE as a new robust measure of protein biosynthesis rates based on ‘randomly’ formed disomes

We propose STE as a new measure of translation efficiency, which is largely devoid of biases regularly encountered in classical translation efficiency (TE) measures and can provide accurate transcript-wise information about the protein biosynthesis rates within the translated fraction of mRNA. STE ranks different mRNAs by their ‘translational power’ within a single experiment, as it can predict the absolute specific protein output. Further, STE can determine changes in the translation of mRNAs across multiple conditions comparatively, such as in its $^{\delta}$STE format. $^{\delta}$STE highlights gene expression alterations enacted purely through translational control. For example, $^{\delta}$STE identifies clusters of translationally-controlled mRNAs upon a 10-minute glucose starvation response (Figure 4C and D, Supplementary Figures S8C, S10 and Supplementary Data S1, S2). Importantly, STE includes weighting components related to the initiation efficiency, such as those derived from the 5′ UTR- and/or start codon-localized eTCP-seq SSU footprinting data, thus encompassing the differential input of initiation, and specifically scanning, into the translational power of a given mRNA.

An intriguing speculation is that those mRNAs in which the high DS_s/RS is not supported by evidence of efficient ongoing initiation, while they had formed relatively dense polysomes through once active initiation, have since been translationally suppressed without their disassembly. The suppression could be specific, such as in the cases of ribosomal quality control or nonsense-mediated control mechanisms (56,66,96,131–134), or unspecific, such as through too tight ribosomal packing (135). These transiently inactive or less active polysomes are less evident using other measures (Supplementary Figure S7C and D). Indeed, the scanning control, and the start codon recognition dynamics (expressed as the inverse RS_{ORF}/RS_{START}) components derived from the eTCP-seq data provided a substantial contribution to the STE accuracy in both model implementations (Supplementary Figure S7C and S8E), and thus appear indispensable to realistic footprint-based protein synthesis rate measurements.

Stochastic and non-random disome signal components of STE

Highlighting the importance of the DS_s signal, among all other eTCP-seq-derived parameters it alone allowed for a clear discrimination between the types of translational behavior of mRNAs transcriptome-wide. When analyzed against experimentally-defined protein biosynthesis rates, the DS_{BASE}/RS (a close approximation of DS_s/RS) clearly identified three types of translational behavior (Supplementary Figure S5G). In the first type, the DS_{BASE}/RS value retained its high frequency (compared to the initial, unfiltered DS value) and was positively correlated with the protein biosynthesis rate. Notably, the positive correlation cannot be fully explained by ORF length differences between respective mRNAs. These mRNAs mostly contained those ranked high by STE and included highly translated transcripts with GO terms ‘cytosolic part’ and ‘ribosomes/ribosomal subunits’ (Supplementary Figure S5H). The second group of transcripts demonstrated a partially decreased DS_{BASE}/RS value while retaining moderate to high protein biosynthesis rate, thus being relatively well-occupied by the RS. This situation is only possible if the polysomal organization itself actively prevents the formation of stochastic DS_s. Such effects can occur in the polysomes that remain constitutively attached to cellular structures, such as the endoplasmic reticulum (ER) (136). GO analysis of this group indeed demonstrates enrichment for the ER and Golgi-associated and exported proteins (Supplementary Figure S5H). Strikingly, these results are recapitulated by the $^{\delta}PSF_{S10}$ measure, which shows a negative correlation of change with $^{\delta}STE_{S10}$ for the ER genes, but retains a positive correlation with all other genes (Supplementary Figure S12A,B).

The third group of mRNAs demonstrates the lowest DS_{BASE}/RS value, likely representing transcripts of low abundance and insufficient signal, low protein output as confirmed by the low STE and reduced protein biosynthesis rates, or potentially functional pausing of the elongating ribosomes over mRNA (137). Indeed, certain elongation rate transitions (such as ‘fast-slow-fast’; Supplementary Figure S3, top row) generate disproportionally low incidents of DS_s, compared to other scenarios. Thus, among the third group we can expect transcripts with non-random RS slowdown or stalling. Remarkably, GO analysis of this group shows enrichment of the components of proteasome assembly (Supplementary Figure S5H), which have been demonstrated before to exhibit specific folding-dependent ribosomal pausing that is relieved upon co-translational assembly of the respective complexes (60,66,95,138). Confirming these observations, further dissecting $^{\delta}PSF_{S10}$ to $^{\delta}STE_{S10}$ by deviation from the regression line into genes where glucose-starved translational involvement is under-estimated and over-estimated by $^{\delta}PSF_{S10}$ reveals a complex mix of ORF length and ribosome stalling/slowdown biases. The extent of ribosome stalling or slowdown can be directly estimated by an eTCP-seq ‘spikiness’ (TCP-S), expressed as per-transcript fraction of long peak disome footprints (TCP-S long) or maximum peak footprints (TCP-S max) of all ribosome and disome footprint types, and is distinctively different in the deviating $^{\delta}PSF_{S10}$ genes (Supplementary Figure S12E).

It remains uncertain what are the biological roles of stochastically co-localized ribosomes, and what are the functional implications of this process on translation and related molecular activities. It would be of interest to evaluate functional, transient ribosomal co-localization and the impact on accessibility and configuration of the inter-ribosomal surfaces, which in turn can favor certain elongation cycle configurations. For example in cryoelectron microscopy data, transiently-formed, actively-translating disomes in a bacterial system have bL9 protein of the leading ribosome in a stretched-out conformation, preventing too tight contact with the leading ribosome and possibly blocking access of some of the translation factors to the trailing ribosome (130). These disome interactions were shown to favor certain ribosomal rotational configurations and possibly help prevent frameshifting and aberrant translocation in transient collisions (130). Kinetically, stochastic co-localization non-linearly affects the performance of the trailing ribosome, stopping it more frequently as it approaches the leading ribosome. As such, this 'gentle throttling' can help to keep some distance between the translating ribosomes and prevent longer pauses, even if one of the ribosomes is consistently performing less well in elongation.

Overall, the above examples demonstrate that the inclusion of DS_s in calculations of translation rates provides a deep connection with the underlying biophysical processes and justifies why using the distinct stochastic and non-random eTCP-seq signals can create a solid foundation for the footprint-based definition of protein biosynthesis rates.

STE provides a refined measurement of static translational control

Critically, STE reveals both static and dynamic control of cellular translation. In *naïve* cells, STE exposed translation rates varying up to ∼100-fold between mRNAs, generally consistent with previous observations (85,139–142). As determined by STE, mRNAs encoding alcohol dehydrogenase, fatty acid transferase, heat shock proteins and other metabolic enzymes were efficiently translated, whereas ribosome assembly, RNA biogenesis, and RNA polymerase-related genes were less so. It is interesting to speculate why the cells would evolve this type of control in steady conditions, when a finely-tuned transcriptional program should not require any additional corrections. A parsimonious explanation is that the perfect conditions of exponential growth in high glucose are not what *Saccharomyces cerevisiae* was exposed to during much of its evolution. Thus, we observe yeast in an evolutionarily atypical environment whereby additional adjustments to gene expression are required. In this case, we indeed can expect to observe the effects of translational buffering and perhaps dosage compensation (140,143–147), which would otherwise be difficult to explain for a fast-evolving species. Another explanation could be that we detect an effect characteristic to only a subfraction of the cells, as our experiments were performed in a non-synchronized culture. Indeed, the stochasticity of *GCN4* initiation rates results in different yeast re-programming pathways (120). The more complex behavior of higher eukaryotic cells will warrant detailed investigations of translational control at peak and off-duty cell cycle times (148–151).

Accurate capture of translational control during glucose starvation in yeast by STE

Our major findings for the glucose stress-induced gene expression dynamics revealed by STE align with published work, but provide highly resolved clusters of mRNAs with discrete gene ontology. Confirming prior observations, the most translated mRNAs and those arguably contributing the most to polysome formation are indeed suppressed in the starved cells' STE, as can be judged by the riddance of the non-starved STE top quarter values from the starved cells' STE distribution (377 genes in the non-starved versus 103 genes in the starved against non-starved STE distribution of model #1; Supplementary Figures S7B, S8C; Supplementary Data S1). These observations are confirmed by the PSF measure, where clipping of the higher rates of translational engagement is also exhibited by the starved cells (Supplementary Figure S11H).

After 10 min, yeast substantially accelerate translation of mRNAs encoding antioxidant activity and oxidative stress response functions (observed with both models; Figure 4C, Supplementary Figures S9A and S10A). This is a true 'emergency' measure to cope with the elevated reactive oxygen species emanating from the highly active respiratory chain in the absence of glucose (28,152–154). The slightly less acute STE up-regulation is related to the mannosylation, sugar transferase activity and cell wall metabolism (Figure 4C, Supplementary Figures S9A and S10A), which may result from translation-based compensation for the RNA abundance-level shutdown of these processes, in order to maintain a required multiplication rate, although links between certain mannosylation reactions to overcoming oxidative stress and acute translational repression are also known (155). The upregulation of exopeptidases in this group could be a direct adaptive response to the low availability of nutrients (Figure 4C, Supplementary Figures S9A and S10A). The mRNA groups with prominent STE down-regulation associate particularly with RNA polymerase I, translation initiation suppression, rRNA and ribosomogenesis genes, similar to yeast responses to other acute nutrient stresses (156). The urgency of the response in these cases is dictated by the high energy consumption of the respective processes, which is vitally threatening in the background of glucose absence. Why a translation-level response is required in this case is unclear but may relate to an inability to quickly remove the respective RNAs, or a necessity to preserve them for future use and shut down the corresponding pathways non-destructively. Interestingly, there is a mild translational de-repression in the starved cells for the group of transcripts that have low specific protein output in the non-starved cells (Supplementary Figures S7A and S8B). We speculate that this can be a consequence of different competitive pressures in translation of mRNA in starved cells, with less mRNA types and initiation factors but more ribosomes and SSUs available as observed in response to viral infections (148–151). Thus, STE allows us to uncover novel peculiarities of protein biosynthesis regulation.

Specific highlights of the STE-identified cases of acute translational regulation include increased translation of the YGL256W/*ADH4* (alcohol dehydrogenase; which in the absence of glucose converts ethanol to acetaldehyde producing nicotinamide adenine dinucleotide (NAD) + hydrogen (H)), YCL040W/*GLK1* (glucokinase; which is a glucose sensor), YFL014W/*HSP12* (plasma membrane protein involved in membrane organization during stress conditions including starvation), YAL038W/*CDC19* (pyruvate kinase functions important for tricarboxylic acid (TCA) cycle and glucose fermentation), YDL010W/*GRX6* (involved in oxidative stress response), YER067W/*RGI1* (involved in energy metabolism under respiration), YDR459C/*PFA5* (protein fatty acyltransferase) and others hinting towards metabolic rearrangements set in motion 10 min into starvation, alongside the synthesis of heat shock proteins as a general stress response mechanism found in yeast (Supplementary Data S1, S2). Conversely, YOR224C/*RPB8* (RNA polymerase subunit common to Pol I, II and III), YJL148W/*RPA34* (RNA Pol I essential for nucleolar assembly), YJL177W/*RPL17B*, YPL079W/*RPL21B* (ribosomal proteins important for LSU assembly), YLR221C/*RSA3* (protein crucial for ribosome maturation), YCR047C/*BUD23* (ribosome biogenesis factor required for rRNA processing and nuclear export) and other mRNAs were acutely down-regulated, confirming severe down-regulation of processes with high energy expense (Supplementary Data S1, S2).

Outlook

Footprints from co-localized ribosomes detected by eTPC-seq carry information about single-molecule events such as diffusion- and spatial configuration-driven transient contacts between ribosomes, as they form polysomes with translated mRNA *in vivo*, and will be useful in interpreting polysomal configurations per-mRNA type (130,157,158). We suggest that similar high-throughput data containing information about random individual molecular co-incidence can be utilized across a variety of biological molecular processes to provide robust measurements in dynamic situations.

STE AI can aid investigations of translational dynamics beyond technical reproducibility. For example, mRNAs may be unequally distributed and selectively translated over different sub-cellular locations. Under such circumstances, any selective translational stalling and degradation of mRNAs can cause disproportionate and functionally irrelevant ribosomal density on mRNA, if it is calculated based on the relative ribosome footprint to the relative RNA-seq signal ratio (Figure 4D). A scenario even more detrimental to the regular approaches is that of a rapid influx of newly-produced mRNA (*e.g.* intense nuclear export, rapid release from condensates) or, inversely, its rapid masking into some form of structurally-segregated foci, such as processing bodies or condensates (Figure 4D). In these cases, common under many stress responses of eukaryotic cells, reliance on the mRNA concentration or relative abundance can be misleading. By avoiding the use of mRNA abundance measurements, STE alleviates these limitations.

We see utility for STE in scenarios involving rapid reorganization of RNA functional states. STE AI can be used when comparison of the absolute 'translational power' between mRNAs is important, such as in cross-species measurements of adaptive responses, mRNA therapeutic vaccine design or synthetic biology applications in search of universal mRNA building blocks that provide well-defined, cell-type-specific translational output (13,159,160).

Data availability

All short-read sequencing data from this study have been uploaded to NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) with GSE200091 accession. All direct nanopore long-read sequencing data from this study have been uploaded to the NCBI Sequence Read Archive (SRA) as BioProject PRJNA1022817. Code used in the work is available on Figshare: https://doi.org/10.6084/m9.figshare.24996602.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

The authors are grateful to the members of the Division of Genome Sciences and Cancer at The John Curtin School of Medical Research (JCSMR), The Australian National University (ANU). The authors acknowledge The Biomolecular Resource Facility of JCSMR, ANU and Microscopy Australia facilities in the Centre for Advanced Microscopy, ANU, supported by funding from ANU and the Australian Government.

Funding

National Health and Medical Research Council of Australia (NHMRC) Investigator Grants [GNT1175388 to N.E.S. and GNT2018363 to T.P.]; Bootes Foundation Grant 2022 (to N.E.S.); Australian Research Council Discovery Grant [DP180100111 to T.P., N.E.S.]; NHMRC Research Fellowship [GNT1135928 to T.P.]; European Union's Horizon 2020 – Research and innovation program Marie Sklodowska-Curie grant [agreement No 890462 to A.C.]. Funding for open access charge: National Health and Medical Research Council (NHMRC) Investigator Grant [GNT1175388, GNT2018363]; Bootes Foundation grant 2022; European Union's Horizon 2020 – Research and innovation program Marie Sklodowska-Curie grant [890462].

Conflict of interest statement

None declared.

References

1. Hershey,J.W.B., Sonenberg,N. and Mathews,M.B. (2019) Principles of translational control. *Cold Spring Harb. Perspect. Biol.*, **11**, a032607.
2. Shirokikh,N.E. and Preiss,T. (2018) Translation initiation by cap-dependent ribosome recruitment: recent insights and open questions. *Wiley Interdiscip. Rev. RNA*, **9**, e1473.
3. Hinnebusch,A.G., Ivanov,I.P. and Sonenberg,N. (2016) Translational control by 5′-untranslated regions of eukaryotic mRNAs. *Science*, **352**, 1413–1416.
4. Shirokikh,N.E. (2021) Translation complex stabilization on messenger RNA and footprint profiling to study the RNA responses and dynamics of protein biosynthesis in the cells. *Crit. Rev. Biochem. Mol. Biol.*, **57**, 1–44.
5. Janapala,Y., Woodward,K., Lee,J., Rug,M., Preiss,T. and Shirokikh,N.E. (2021) Rapid in vivo fixation and isolation of translational complexes from eukaryotic cells. *J. Visual. Exp.*, **178**, e62639.
6. Janapala,Y., Preiss,T. and Shirokikh,N.E. (2019) Control of translation at the initiation phase during glucose starvation in yeast. *Int. J. Mol. Sci.*, **20**, 4043.
7. Ashe,M.P., De Long,S.K. and Sachs,A.B. (2000) Glucose depletion rapidly inhibits translation initiation in yeast. *Mol. Biol. Cell*, **11**, 833–848.
8. Houston,R., Sekine,S. and Sekine,Y. (2020) The coupling of translational control and stress responses. *J. Biochem.*, **168**, 93–102.
9. Ouranidis,A., Vavilis,T., Mandala,E., Davidopoulou,C., Stamoula,E., Markopoulou,C.K., Karagianni,A. and Kachrimanis,K. (2021) mRNA therapeutic modalities design,

formulation and manufacturing under Pharma 4.0 principles. *Biomedicines*, **10**, 50.
10. Zeng,C., Hou,X., Yan,J., Zhang,C., Li,W., Zhao,W., Du,S. and Dong,Y. (2020) Leveraging mRNA sequences and nanoparticles to deliver SARS-CoV-2 antigens In vivo. *Adv. Mater.*, **32**, 2004452.
11. Sultana,N., Hadas,Y., Sharkar,M.T.K., Kaur,K., Magadum,A., Kurian,A.A., Hossain,N., Alburquerque,B., Ahmed,S., Chepurko,E., *et al.* (2020) Optimization of 5′ untranslated region of modified mRNA for use in cardiac or hepatic ischemic injury. *Mol. Ther. Methods Clin. Dev.*, **17**, 622–633.
12. Jia,L., Mao,Y., Ji,Q., Dersh,D., Yewdell,J.W. and Qian,S.-B. (2020) Decoding mRNA translatability and stability from the 5′ UTR. *Nat. Struct. Mol. Biol.*, **27**, 814–821.
13. Sample,P.J., Wang,B., Reid,D.W., Presnyak,V., McFadyen,I.J., Morris,D.R. and Seelig,G. (2019) Human 5′ UTR design and variant effect prediction from a massively parallel translation assay. *Nat. Biotechnol.*, **37**, 803–809.
14. Cao,J., Novoa,E.M., Zhang,Z., Chen,W.C.W., Liu,D., Choi,G.C.G., Wong,A.S.L., Wehrspaun,C., Kellis,M. and Lu,T.K. (2021) High-throughput 5′ UTR engineering for enhanced protein production in non-viral gene therapies. *Nat. Commun.*, **12**, 4138.
15. To,K.K.W. and Cho,W.C.S. (2021) An overview of rational design of mRNA-based therapeutics and vaccines. *Expert Opin Drug Discov*, **16**, 1307–1317.
16. Zhang,H., Zhang,L., Lin,A., Xu,C., Li,Z., Liu,K., Liu,B., Ma,X., Zhao,F., Jiang,H., *et al.* (2023) Algorithm for optimized mRNA design improves stability and immunogenicity. *Nature*, **621**, 396–403.
17. Sumi,S., Hamada,M. and Saito,H. (2024) Deep generative design of RNA family sequences. *Nat. Methods*, **21**, 435–443.
18. Akiyama,M. and Sakakibara,Y. (2022) Informative RNA base embedding for RNA structural alignment and clustering by deep representation learning. *NAR Genomics Bioinformatics*, **4**, lqac012.
19. Fan,X., Chang,T., Chen,C., Hafner,M. and Wang,Z. (2023) Accurate prediction of RNA translation with a deep learning architecture. bioRxiv doi: https://doi.org/10.1101/2023.07.08.548206, 08 July 2023, preprint: not peer reviewed.
20. Castillo-Hair,S.M. and Seelig,G. (2022) Machine learning for designing next-generation mRNA therapeutics. *Acc. Chem. Res.*, **55**, 24–34.
21. Boyd,N., Anderson,B.M., Townshend,B., Chow,R., Stephens,C.J., Rangan,R., Kaplan,M., Corley,M., Tambe,A., Ido,Y., *et al.* (2023) ATOM-1: a foundation model for RNA structure and function built on chemical mapping data. bioRxiv doi: https://doi.org/10.1101/2023.12.13.571579, 14 December 2023, preprint: not peer reviewed.
22. Andreev,D.E., O'Connor,P.B.F., Loughran,G., Dmitriev,S.E., Baranov,P.V. and Shatsky,I.N. (2017) Insights into the mechanisms of eukaryotic translation gained with ribosome profiling. *Nucleic Acids Res.*, **45**, 513–526.
23. Zinshteyn,B., Wangen,J.R., Hua,B. and Green,R. (2020) Nuclease-mediated depletion biases in ribosome footprint profiling libraries. *RNA*, **26**, 1481–1488.
24. Hinnebusch,A.G. (1993) Gene-specific translational control of the yeast *GCN4* gene by phosphorylation of eukaryotic initiation factor 2. *Mol. Microbiol.*, **10**, 215–223.
25. Crawford,R.A. and Pavitt,G.D. (2019) Translational regulation in response to stress in Saccharomyces cerevisiae. *Yeast*, **36**, 5–21.
26. Hinnebusch,A.G. (2017) Structural insights into the mechanism of scanning and start codon recognition in eukaryotic translation initiation. *Trends Biochem. Sci*, **42**, 589–611.
27. Guzikowski,A.R., Harvey,A.T., Zhang,J., Zhu,S., Begovich,K., Cohn,M.H., Wilhelm,J.E. and Zid,B.M. (2022) Differential translation elongation directs protein synthesis in response to acute glucose deprivation in yeast. *RNA Biol*, **19**, 636–649.
28. Brauer,M.J., Saldanha,A.J., Dolinski,K. and Botstein,D. (2005) Homeostatic adjustment and metabolic remodeling in glucose-limited yeast cultures. *MBoC*, **16**, 2503–2517.
29. Buchan,J.R., Muhlrad,D. and Parker,R. (2008) P bodies promote stress granule assembly in Saccharomyces cerevisiae. *J. Cell Biol.*, **183**, 441–455.
30. Shah,K.H., Zhang,B., Ramachandran,V. and Herman,P.K. (2013) Processing body and stress granule assembly occur by independent and differentially regulated pathways in Saccharomyces cerevisiae. *Genetics*, **193**, 109–123.
31. Rao,B.S. and Parker,R. (2017) Numerous interactions act redundantly to assemble a tunable size of P bodies in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, E9569–E9578.
32. Grousl,T., Vojtova,J., Hasek,J. and Vomastek,T. (2022) Yeast stress granules at a glance. *Yeast*, **39**, 247–261.
33. Galdieri,L., Mehrotra,S., Yu,S. and Vancura,A. (2010) Transcriptional regulation in yeast during diauxic shift and stationary phase. *OMICS*, **14**, 629–638.
34. Zaman,S., Lippman,S.I., Schneper,L., Slonim,N. and Broach,J.R. (2009) Glucose regulates transcription in yeast through a network of signaling pathways. *Mol. Syst. Biol.*, **5**, 245.
35. Castelli,L.M., Lui,J., Campbell,S.G., Rowe,W., Zeef,L.A.H., Holmes,L.E.A., Hoyle,N.P., Bone,J., Selley,J.N., Sims,P.F.G., *et al.* (2011) Glucose depletion inhibits translation initiation via eIF4A loss and subsequent 48S preinitiation complex accumulation, while the pentose phosphate pathway is coordinately up-regulated. *Mol. Biol. Cell*, **22**, 3379–3393.
36. Sen,N.D., Zhang,H. and Hinnebusch,A.G. (2021) Down-regulation of yeast helicase Ded1 by glucose starvation or heat-shock differentially impairs translation of Ded1-dependent mRNAs. *Microorganisms*, **9**, 2413.
37. Montero-Lomelí,M., Morais,B.L.B., Figueiredo,D.L., Neto,D.C.S., Martins,J.R.P. and Masuda,C.A. (2002) The initiation factor eIF4A is involved in the response to lithium stress in Saccharomyces cerevisiae. *J. Biol. Chem.*, **277**, 21542–21548.
38. Bresson,S., Shchepachev,V., Spanos,C., Turowski,T.W., Rappsilber,J. and Tollervey,D. (2020) Stress-induced translation inhibition through rapid displacement of scanning initiation factors. *Mol. Cell*, **80**, 470–484.
39. Meydan,S. and Guydosh,N.R. (2020) Disome and trisome profiling reveal genome-wide targets of ribosome quality control. *Mol. Cell*, **79**, 588–602.
40. Wu,C.C.-C., Peterson,A., Zinshteyn,B., Regot,S. and Green,R. (2020) Ribosome collisions trigger general stress responses to regulate cell fate. *Cell*, **182**, 404–416.
41. Yu,R., Vorontsov,E., Sihlbom,C. and Nielsen,J. (2021) Quantifying absolute gene expression profiles reveals distinct regulation of central carbon metabolism genes in yeast. *eLife*, **10**, e65722.
42. Luan,Y., Tang,N., Yang,J., Liu,S., Cheng,C., Wang,Y., Chen,C., Guo,Y., Wang,H., Zhao,W., *et al.* (2022) Deficiency of ribosomal proteins reshapes the transcriptional and translational landscape in human cells. *Nucleic Acids Res.*, **50**, 6601–6617.
43. Brar,G.A. and Weissman,J.S. (2015) Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nat. Rev. Mol. Cell Biol.*, **16**, 651–664.
44. Ingolia,N.T., Hussmann,J.A. and Weissman,J.S. (2019) Ribosome profiling: global views of translation. *Cold Spring Harb. Perspect. Biol.*, **11**, a032698.
45. Chassé,H., Boulben,S., Costache,V., Cormier,P. and Morales,J. (2017) Analysis of translation using polysome profiling. *Nucleic Acids Res.*, **45**, e15.
46. Coudert,L., Adjibade,P. and Mazroui,R. (2014) Analysis of translation initiation during stress conditions by polysome profiling. *J. Visual. Exp.*, **87**, e51164.
47. King,H.A. and Gerber,A.P. (2016) Translatome profiling: methods for genome-scale analysis of mRNA translation. *Brief. Funct. Genomics*, **15**, 22–31.

48. Hedayioglu,F., Mead,E.J., O'Connor,P.B.F., Skiotys,M., Sansom,O.J., Mallucci,G.R., Willis,A.E., Baranov,P.V., Smales,C.M. and von der Haar,T. (2022) Evaluating data integrity in ribosome footprinting datasets through modelled polysome profiles. *Nucleic Acids Res.*, **50**, e112.
49. Bartholomäus,A., Campo,C.D. and Ignatova,Z. (2016) Mapping the non-standardized biases of ribosome profiling. *Biol. Chem.*, **397**, 23–35.
50. Mohammad,F., Green,R. and Buskirk,A.R. (2019) A systematically-revised ribosome profiling method for bacteria reveals pauses at single-codon resolution. *eLife*, **8**, e42591.
51. Hussmann,J.A., Patchett,S., Johnson,A., Sawyer,S. and Press,W.H. (2015) Understanding biases in ribosome profiling experiments reveals signatures of translation dynamics in yeast. *PLos Genet.*, **11**, e1005732.
52. Darnell,A.M., Subramaniam,A.R. and O'Shea,E.K. (2018) Translational control through differential ribosome pausing during amino acid limitation in mammalian cells. *Mol. Cell*, **71**, 229–243.
53. Schuller,A.P. and Green,R. (2018) Roadblocks and resolutions in eukaryotic translation. *Nat. Rev. Mol. Cell Biol.*, **19**, 526–541.
54. Zhang,S., Hu,H., Zhou,J., He,X., Jiang,T. and Zeng,J. (2017) Analysis of ribosome stalling and translation elongation dynamics by Deep learning. *Cell Syst.*, **5**, 212–220.
55. Yang,Q., Yu,C.-H., Zhao,F., Dang,Y., Wu,C., Xie,P., Sachs,M.S. and Liu,Y. (2019) eRF1 mediates codon usage effects on mRNA translation efficiency through premature termination at rare codons. *Nucleic Acids Res.*, **47**, 9243–9258.
56. Rubio,A., Ghosh,S., Mülleder,M., Ralser,M. and Mata,J. (2021) Ribosome profiling reveals ribosome stalling on tryptophan codons and ribosome queuing upon oxidative stress in fission yeast. *Nucleic Acids Res.*, **49**, 383–399.
57. Mitarai,N., Sneppen,K. and Pedersen,S. (2008) Ribosome collisions and translation efficiency: optimization by codon usage and mRNA destabilization. *J. Mol. Biol.*, **382**, 236–245.
58. Sabi,R. and Tuller,T. (2017) Computational analysis of nascent peptides that induce ribosome stalling and their proteomic distribution in Saccharomyces cerevisiae. *RNA*, **23**, 983–994.
59. Takamatsu,S., Ohashi,Y., Onoue,N., Tajima,Y., Imamichi,T., Yonezawa,S., Morimoto,K., Onouchi,H., Yamashita,Y. and Naito,S. (2020) Reverse genetics-based biochemical studies of the ribosomal exit tunnel constriction region in eukaryotic ribosome stalling: spatial allocation of the regulatory nascent peptide at the constriction. *Nucleic Acids Res.*, **48**, 1985–1999.
60. Arpat,A.B., Liechti,A., Matos,M.D., Dreos,R., Janich,P. and Gatfield,D. (2020) Transcriptome-wide sites of collided ribosomes reveal principles of translational pausing. *Genome Res.*, **30**, 985–999.
61. Corsi,A.K. and Schekman,R. (1996) Mechanism of polypeptide translocation into the endoplasmic reticulum. *J. Biol. Chem.*, **271**, 30299–30302.
62. Wolin,S.L. and Walter,P. (1988) Ribosome pausing and stacking during translation of a eukaryotic mRNA. *EMBO J.*, **7**, 3559–3569.
63. Zhao,T., Chen,Y.-M., Li,Y., Wang,J., Chen,S., Gao,N. and Qian,W. (2021) Disome-seq reveals widespread ribosome collisions that promote cotranslational protein folding. *Genome Biol.*, **22**, 16.
64. Liu,B., Han,Y. and Qian,S.-B. (2013) Cotranslational response to proteotoxic stress by elongation pausing of ribosomes. *Mol. Cell*, **49**, 453–463.
65. Han,Y., David,A., Liu,B., Magadán,J.G., Bennink,J.R., Yewdell,J.W. and Qian,S.-B. (2012) Monitoring cotranslational protein folding in mammalian cells at codon resolution. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 12467–12472.
66. Collart,M.A. and Weiss,B. (2020) Ribosome pausing, a dangerous necessity for co-translational events. *Nucleic Acids Res.*, **48**, 1043–1055.
67. Goldman,D.H., Livingston,N.M., Movsik,J., Wu,B. and Green,R. (2021) Live-cell imaging reveals kinetic determinants of quality control triggered by ribosome stalling. *Mol. Cell*, **81**, 1830–1840.
68. Matsuo,Y., Ikeuchi,K., Saeki,Y., Iwasaki,S., Schmidt,C., Udagawa,T., Sato,F., Tsuchiya,H., Becker,T., Tanaka,K., *et al.* (2017) Ubiquitination of stalled ribosome triggers ribosome-associated quality control. *Nat. Commun.*, **8**, 159.
69. Wu,C.C.-C., Zinshteyn,B., Wehner,K.A. and Green,R. (2019) High-resolution ribosome profiling defines discrete ribosome elongation states and translational regulation during cellular stress. *Mol. Cell*, **73**, 959–970.
70. Veltri,A.J., D'Orazio,K.N., Lessen,L.N., Loll-Krippleber,R., Brown,G.W. and Green,R. (2022) Distinct elongation stalls during translation are linked with distinct pathways for mRNA degradation. *eLife*, **11**, e76038.
71. Weinberg,D.E., Shah,P., Eichhorn,S.W., Hussmann,J.A., Plotkin,J.B. and Bartel,D.P. (2016) Improved ribosome-footprint and mRNA measurements provide insights into dynamics and regulation of yeast translation. *Cell Rep.*, **14**, 1787–1799.
72. Wickramasinghe,V.O. and Laskey,R.A. (2015) Control of mammalian gene expression by selective mRNA export. *Nat. Rev. Mol. Cell Biol.*, **16**, 431–442.
73. Protter,D.S.W. and Parker,R. (2016) Principles and properties of stress granules. *Trends Cell Biol.*, **26**, 668–679.
74. Tauber,D., Tauber,G., Khong,A., Van Treeck,B., Pelletier,J. and Parker,R. (2020) Modulation of RNA condensation by the DEAD-box protein eIF4A. *Cell*, **180**, 411–426.
75. Khong,A., Matheny,T., Jain,S., Mitchell,S.F., Wheeler,J.R. and Parker,R. (2017) The stress granule transcriptome reveals principles of mRNA accumulation in stress granules. *Mol. Cell*, **68**, 808–820.
76. Archer,S.K., Shirokikh,N.E., Beilharz,T.H. and Preiss,T. (2016) Dynamics of ribosome scanning and recycling revealed by translation complex profiling. *Nature*, **535**, 570–574.
77. Shirokikh,N.E., Archer,S.K., Beilharz,T.H., Powell,D. and Preiss,T. (2017) Translation complex profile sequencing to study the in vivo dynamics of mRNA–ribosome interactions during translation initiation, elongation and termination. *Nat. Protoc.*, **12**, 697–731.
78. Wagner,S., Herrmannová,A., Hronová,V., Gunišová,S., Sen,N.D., Hannan,R.D., Hinnebusch,A.G., Shirokikh,N.E., Preiss,T. and Valášek,L.S. (2020) Selective translation complex profiling reveals staged initiation and Co-translational assembly of initiation factor complexes. *Mol. Cell*, **79**, 546–560.
79. Shirokikh,N.E., Archer,S.K., Beilharz,T.H., Powell,D. and Preiss,T. (2017) Translation complex profile sequencing to study the in vivo dynamics of mRNA-ribosome interactions during translation initiation, elongation and termination. *Nat. Protoc.*, **12**, 697–731.
80. Luthe,D.S. (1983) A simple technique for the preparation and storage of sucrose gradients. *Anal. Biochem.*, **135**, 230–232.
81. Archer,S.K., Shirokikh,N.E. and Preiss,T. (2015) Probe-directed degradation (PDD) for flexible removal of unwanted cDNA sequences from RNA-seq libraries. *Curr. Protoc. Hum. Genet.*, **85**, 11.15.1–11.15.36.
82. Archer,S.K., Shirokikh,N.E. and Preiss,T. (2014) Selective and flexible depletion of problematic sequences from RNA-seq libraries at the cDNA stage. *BMC Genomics*, **15**, 401.
83. Sneddon,A., Mateos,P.A., Shirokikh,N. and Eyras,E. (2022) Language-Informed Basecalling Architecture for Nanopore Direct RNA Sequencing. In: *Proceedings of the 17th Machine Learning in Computational Biology meeting, in Proceedings of Machine Learning Research*. PMLR, pp. 150–165.
84. Visser,I. and Speekenbrink,M. (2010) depmixS4: an R package for hidden markov models. *J. Stat. Softw.*, **36**, 1–21.
85. Riba,A., Nanni,N.D., Mittal,N., Arhné,E., Schmidt,A. and Zavolan,M. (2019) Protein synthesis rates and ribosome occupancies reveal determinants of translation elongation rates. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 15023–15032.
86. Hothorn,T. and Zeileis,A. (2015) partykit: A Modular Toolkit for Recursive Partytioning in R. *Journal of Machine Learning Research*, **16**, 3905–3909.

87. Kolde,R. (2012) Pheatmap: pretty heatmaps. *R Package Version*. **1**, 726.
88. Yu,G., Wang,L.-G., Han,Y. and He,Q.-Y. (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics*, **16**, 284–287.
89. Wolfram Research, Inc (2024) Mathematica. Wolfram Research, Inc., Version 14.0.
90. Sharma,A.K., Sormanni,P., Ahmed,N., Ciryam,P., Friedrich,U.A., Kramer,G. and O'Brien,E.P. (2019) A chemical kinetic basis for measuring translation initiation and elongation rates from ribosome profiling data. *PLoS Comput. Biol.*, **15**, e1007070.
91. Boersma,S., Khuperkar,D., Verhagen,B.M.P., Sonneveld,S., Grimm,J.B., Lavis,L.D. and Tanenbaum,M.E. (2019) Multi-color single-molecule imaging uncovers extensive heterogeneity in mRNA decoding. *Cell*, **178**, 458–472.
92. Han,P., Shichino,Y., Schneider-Poetsch,T., Mito,M., Hashimoto,S., Udagawa,T., Kohno,K., Yoshida,M., Mishima,Y., Inada,T., *et al.* (2020) Genome-wide survey of ribosome collision. *Cell Rep.*, **31**, 107610.
93. Juszkiewicz,S., Chandrasekaran,V., Lin,Z., Kraatz,S., Ramakrishnan,V. and Hegde,R.S. (2018) ZNF598 Is a quality control sensor of collided ribosomes. *Mol. Cell*, **72**, 469–481.
94. Chandrasekaran,V., Juszkiewicz,S., Choi,J., Puglisi,J.D., Brown,A., Shao,S., Ramakrishnan,V. and Hegde,R.S. (2019) Mechanism of ribosome stalling during translation of a poly(A) tail. *Nat. Struct. Mol. Biol.*, **26**, 1132–1140.
95. Richter,J.D. and Coller,J. (2015) Pausing on polyribosomes: make way for elongation in translational control. *Cell*, **163**, 292–300.
96. Buskirk,A.R. and Green,R. (2017) Ribosome pausing, arrest and rescue in bacteria and eukaryotes. *Philos. Trans. Roy. Soc. B: Biol. Sci.*, **372**, 20160183.
97. Gelsinger,D.R., Dallon,E., Reddy,R., Mohammad,F., Buskirk,A.R. and DiRuggiero,J. (2020) Ribosome profiling in archaea reveals leaderless translation, novel translational initiation sites, and ribosome pausing at single codon resolution. *Nucleic Acids Res.*, **48**, 5201–5216.
98. Schmidt,T., Schütz,G.J., Baumgartner,W., Gruber,H.J. and Schindler,H. (1996) Imaging of single molecule diffusion. *Proc. Natl Acad. Sci. U.S.A.*, **93**, 2926–2929.
99. Serag,M.F., Abadi,M. and Habuchi,S. (2014) Single-molecule diffusion and conformational dynamics by spatial integration of temporal fluctuations. *Nat. Commun.*, **5**, 5123.
100. Hendriks,F.C., Meirer,F., Kubarev,A.V., Ristanović,Z., Roeffaers,M.B.J., Vogt,E.T.C., Bruijnincx,P.C.A. and Weckhuysen,B.M. (2017) Single-molecule fluorescence microscopy reveals local diffusion coefficients in the pore network of an individual catalyst particle. *J. Am. Chem. Soc.*, **139**, 13632–13635.
101. Szavits-Nossan,J. and Ciandrini,L. (2020) Inferring efficiency of translation initiation and elongation from ribosome profiling. *Nucleic Acids Res.*, **48**, 9478–9490.
102. Chevalier,C., Dorignac,J., Ibrahim,Y., Choquet,A., David,A., Ripoll,J., Rivals,E., Geniet,F., Walliser,N.-O., Palmeri,J., *et al.* (2023) Physical modeling of ribosomes along messenger RNA: estimating kinetic parameters from ribosome profiling experiments using a ballistic model. *PLoS Comput. Biol.*, **19**, e1011522.
103. Ingolia,N.T., Ghaemmaghami,S., Newman,J.R.S. and Weissman,J.S. (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*, **324**, 218–223.
104. Steitz,J.A. (1969) Polypeptide chain initiation: nucleotide sequences of the three ribosomal binding sites in bacteriophage R17 RNA. *Nature*, **224**, 957–964.
105. Lareau,L.F., Hite,D.H., Hogan,G.J. and Brown,P.O. (2014) Distinct stages of the translation elongation cycle revealed by sequencing ribosome-protected mRNA fragments. *eLife*, **3**, e01257.
106. Staehelin,T., Wettstein,F.O., Oura,H. and Noll,H. (1964) Determination of the coding ratio based on molecular weight of messenger ribonucleic acid associated with ergosomes of different aggregate size. *Nature*, **201**, 264–270.
107. Bergmann,J.E. and Lodish,H.F. (1979) A kinetic model of protein synthesis. Application to hemoglobin synthesis and translational control. *J. Biol. Chem.*, **254**, 11927–11937.
108. Lauria,F., Tebaldi,T., Lunelli,L., Struffi,P., Gatto,P., Pugliese,A., Brigotti,M., Montanaro,L., Ciribilli,Y., Inga,A., *et al.* (2015) RiboAbacus: a model trained on polyribosome images predicts ribosome density and translational efficiency from mammalian transcriptomes. *Nucleic Acids Res.*, **43**, e153.
109. Haight,F.A. (1963) Chapter 3: Fundamental characteristics of road traffic In: *Mathematics in Science and Engineering*. Elsevier, Vol. **7**, pp. 67–95.
110. Zia,R.K.P., Dong,J.J. and Schmittmann,B. (2011) Modeling translation in protein synthesis with TASEP: a tutorial and recent developments. *J. Stat. Phys.*, **144**, 405.
111. Bifeld,E., Lorenzen,S., Bartsch,K., Vasquez,J.-J., Siegel,T.N. and Clos,J. (2018) Ribosome profiling reveals HSP90 inhibitor effects on stage-specific protein synthesis in Leishmania donovani. *mSystems*, **3**, e00214-18.
112. Ciandrini,L., Crisostomo,R.L. and Szavits-Nossan,J. (2023) TASEPy: A Python-based package to iteratively solve the inhomogeneous exclusion process. *SciPost Phys. Codebases*, https://doi.org/10.21468/SciPostPhysCodeb.22.
113. Argüello,R.J., Reverendo,M., Mendes,A., Camosseto,V., Torres,A.G., Ribas de Pouplana,L., van de Pavert,S.A., Gatti,E. and Pierre,P. (2018) SunRiSE - measuring translation elongation at single-cell resolution by means of flow cytometry. *J. Cell Sci.*, **131**, jcs214346.
114. Petrov,A., Chen,J., O'Leary,S., Tsai,A. and Puglisi,J.D. (2012) Single-Molecule Analysis of Translational Dynamics. *Cold Spring Harb. Perspect. Biol.*, **4**, a011551.
115. Möckl,L., Roy,A.R., Petrov,P.N. and Moerner,W.E. (2020) Accurate and rapid background estimation in single-molecule localization microscopy using the deep neural network BGnet. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 60–67.
116. Johnson,A.G., Grosely,R., Petrov,A.N. and Puglisi,J.D. (2017) Dynamics of IRES-mediated translation. *Philos. Trans. Roy. Soc. B: Biol. Sci.*, **372**, 20160177.
117. Wang,J., Johnson,A.G., Lapointe,C.P., Choi,J., Prabhakar,A., Chen,D.-H., Petrov,A.N. and Puglisi,J.D. (2019) eIF5B gates the transition from translation initiation to elongation. *Nature*, **573**, 605–608.
118. Johansson,M., Chen,J., Tsai,A., Kornberg,G. and Puglisi,J.D. (2014) Sequence-dependent elongation dynamics on macrolide-bound ribosomes. *Cell Rep.*, **7**, 1534–1546.
119. Ye,S. and Lehmann,J. (2022) Genetic code degeneracy is established by the decoding center of the ribosome. *Nucleic Acids Res.*, **50**, 4113–4126.
120. Meng,X., Reed,A., Lai,S., Szavits-Nossan,J. and McCarthy,J.E.G. (2023) Stochastic scanning events on the GCN4 mRNA 5' untranslated region generate cell-to-cell heterogeneity in the yeast nutritional stress response. *Nucleic Acids Res.*, **51**, 6609–6621.
121. Khuperkar,D., Hoek,T.A., Sonneveld,S., Verhagen,B.M.P., Boersma,S. and Tanenbaum,M.E. (2020) Quantification of mRNA translation in live cells using single-molecule imaging. *Nat. Protoc.*, **15**, 1371–1398.
122. Westhof,E., Thornlow,B., Chan,P.P. and Lowe,T.M. (2022) Eukaryotic tRNA sequences present conserved and amino acid-specific structural signatures. *Nucleic Acids Res.*, **50**, 4100–4112.
123. Kim,Y., Eggers,C., Shvetsova,E., Kleemann,L., Sin,O. and Leidel,S.A. (2021) Analysis of codon-specific translation by ribosome profiling. *Methods Enzymol.*, **658**, 191–223.
124. Nedialkova,D.D. and Leidel,S.A. (2015) Optimization of codon translation rates via tRNA modifications maintains proteome integrity. *Cell*, **161**, 1606–1618.
125. Karlin,S. and McGregor,J. (1958) Many server queueing processes with Poisson input and exponential service times. *Pac. J. Math.*, **8**, 87–118.

126. Mito,M., Mishima,Y. and Iwasaki,S. (2020) Protocol for disome profiling to survey ribosome collision in humans and zebrafish. *STAR Protoc*, **1**, 100168.
127. Ikeuchi,K., Tesina,P., Matsuo,Y., Sugiyama,T., Cheng,J., Saeki,Y., Tanaka,K., Becker,T., Beckmann,R. and Inada,T. (2019) Collided ribosomes form a unique structural interface to induce Hel2-driven quality control pathways. *EMBO J.*, **38**, e100276.
128. Juszkiewicz,S., Slodkowicz,G., Lin,Z., Freire-Pritchett,P., Peak-Chew,S.-Y. and Hegde,R.S. (2020) Ribosome collisions trigger cis-acting feedback inhibition of translation initiation. *eLife*, **9**, e60038.
129. Saito,K., Kratzat,H., Campbell,A., Buschauer,R., Burroughs,A.M., Berninghausen,O., Aravind,L., Green,R., Beckmann,R. and Buskirk,A.R. (2022) Ribosome collisions induce mRNA cleavage and ribosome rescue in bacteria. *Nature*, **603**, 503–508.
130. Flügel,T., Schacherl,M., Unbehaun,A., Schroeer,B., Dabrowski,M., Bürger,J., Mielke,T., Sprink,T., Diebolder,C.A., Guillén Schlippe,Y.V., *et al.* (2024) Transient disome complex formation in native polysomes during ongoing protein synthesis captured by cryo-EM. *Nat. Commun.*, **15**, 1756.
131. Tesina,P., Lessen,L.N., Buschauer,R., Cheng,J., Wu,C.C.-C., Berninghausen,O., Buskirk,A.R., Becker,T., Beckmann,R. and Green,R. (2020) Molecular mechanism of translational stalling by inhibitory codon combinations and poly(A) tracts. *EMBO J.*, **39**, e103365.
132. Yordanova,M.M., Loughran,G., Zhdanov,A.V., Mariotti,M., Kiniry,S.J., O'Connor,P.B.F., Andreev,D.E., Tzani,I., Saffert,P., Michel,A.M., *et al.* (2018) AMD1 mRNA employs ribosome stalling as a mechanism for molecular memory formation. *Nature*, **553**, 356–360.
133. Karousis,E.D., Gurzeler,L.-A., Annibaldis,G., Dreos,R. and Mühlemann,O. (2020) Human NMD ensues independently of stable ribosome stalling. *Nat. Commun.*, **11**, 4134.
134. Karamyshev,A.L. and Karamysheva,Z.N. (2018) Lost in translation: ribosome-associated mRNA and protein quality controls. *Front. Genet.*, **9**, 431.
135. Bicknell,A.A., Reid,D.W., Licata,M.C., Jones,A.K., Cheng,Y.M., Li,M., Hsiao,C.J., Pepin,C.S., Metkar,M., Levdansky,Y., *et al.* (2024) Attenuating ribosome load improves protein output from mRNA by limiting translation-dependent mRNA decay. *Cell Rep.*, **43**, 114098.
136. Svitkin,Y.V., Gingras,A.-C. and Sonenberg,N. (2022) Membrane-dependent relief of translation elongation arrest on pseudouridine- and N1-methyl-pseudouridine-modified mRNAs. *Nucleic Acids Res.*, **50**, 7202–7215.
137. Jobava,R., Mao,Y., Guan,B.-J., Hu,D., Krokowski,D., Chen,C.-W., Shu,X.E., Chukwurah,E., Wu,J., Gao,Z., *et al.* (2021) Adaptive translational pausing is a hallmark of the cellular response to severe environmental stress. *Mol. Cell*, **81**, 4191–4208.
138. Panasenko,O.O., Somasekharan,S.P., Villanyi,Z., Zagatti,M., Bezrukov,F., Rashpa,R., Cornut,J., Iqbal,J., Longis,M., Carl,S.H., *et al.* (2019) Co-translational assembly of proteasome subunits in NOT1-containing assemblysomes. *Nat. Struct. Mol. Biol.*, **26**, 110–120.
139. Schwanhäusser,B., Busse,D., Li,N., Dittmar,G., Schuchhardt,J., Wolf,J., Chen,W. and Selbach,M. (2011) Global quantification of mammalian gene expression control. *Nature*, **473**, 337–342.
140. Uebbing,S., Konzer,A., Xu,L., Backström,N., Brunström,B., Bergquist,J. and Ellegren,H. (2015) Quantitative mass spectrometry reveals partial translational regulation for dosage compensation in chicken. *Mol. Biol. Evol.*, **32**, 2716–2725.
141. Mergner,J., Frejno,M., List,M., Papacek,M., Chen,X., Chaudhary,A., Samaras,P., Richter,S., Shikata,H., Messerer,M., *et al.* (2020) Mass-spectrometry-based draft of the Arabidopsis proteome. *Nature*, **579**, 409–414.
142. Zhou,F., Lu,Y., Ficarro,S.B., Adelmant,G., Jiang,W., Luckey,C.J. and Marto,J.A. (2013) Genome-scale proteome quantification by DEEP SEQ mass spectrometry. *Nat. Commun.*, **4**, 2171.
143. Blevins,W.R., Tavella,T., Moro,S.G., Blasco-Moreno,B., Closa-Mosquera,A., Díez,J., Carey,L.B. and Albà,M.M. (2019) Extensive post-transcriptional buffering of gene expression in the response to severe oxidative stress in baker's yeast. *Sci. Rep.*, **9**, 11005.
144. McManus,C.J., May,G.E., Spealman,P. and Shteyman,A. (2014) Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. *Genome Res.*, **24**, 422–430.
145. Wang,S.H., Hsiao,C.J., Khan,Z. and Pritchard,J.K. (2018) Post-translational buffering leads to convergent protein expression levels between primates. *Genome Biol.*, **19**, 83.
146. Kusnadi,E.P., Timpone,C., Topisirovic,I., Larsson,O. and Furic,L. (2022) Regulation of gene expression via translational buffering. *Biochim. Biophys. Acta. Mol. Cell Res.*, **1869**, 119140.
147. Oertlin,C., Lorent,J., Murie,C., Furic,L., Topisirovic,I. and Larsson,O. (2019) Generally applicable transcriptome-wide analysis of translation using anota2seq. *Nucleic Acids Res.*, **47**, e70.
148. Shirokikh,N.E. and Spirin,A.S. (2008) Poly(A) leader of eukaryotic mRNA bypasses the dependence of translation on initiation factors. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 10738–10743.
149. Meade,N., DiGiuseppe,S. and Walsh,D. (2019) Translational control during poxvirus infection. *WIREs RNA*, **10**, e1515.
150. Stern-Ginossar,N., Thompson,S.R., Mathews,M.B. and Mohr,I. (2019) Translational control in virus-infected cells. *Cold Spring Harb. Perspect. Biol.*, **11**, a033001.
151. Walsh,D. (2017) Poxviruses: slipping and sliding through transcription and translation. *PLoS Pathog.*, **13**, e1006634.
152. Gibson,B.R., Lawrence,S.J., Boulton,C.A., Box,W.G., Graham,N.S., Linforth,R.S.T. and Smart,K.A. (2008) The oxidative stress response of a lager brewing yeast strain during industrial propagation and fermentation. *FEMS Yeast Res.*, **8**, 574–585.
153. Petti,A.A., Crutchfield,C.A., Rabinowitz,J.D. and Botstein,D. (2011) Survival of starving yeast is correlated with oxidative stress response and nonrespiratory mitochondrial function. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, E1089–E1098.
154. Wierman,M.B., Maqani,N., Strickler,E., Li,M. and Smith,J.S. (2017) Caloric restriction extends yeast chronological life span by optimizing the Snf1 (AMPK) signaling pathway. *Mol. Cell. Biol.*, **37**, e00562-16.
155. Petkova,M.I., Pujol-Carrion,N. and de la Torre-Ruiz,M.A. (2012) Mtl1 O-mannosylation mediated by both Pmt1 and Pmt2 is important for cell survival under oxidative conditions and TOR blockade. *Fungal Genet. Biol.*, **49**, 903–914.
156. Björkeroth,J., Campbell,K., Malina,C., Yu,R., Bartolomeo,F.D. and Nielsen,J. (2020) Proteome reallocation from amino acid biosynthesis to ribosomes enables yeast to grow faster in rich media. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 21804–21812.
157. Iyer,K.V., Müller,M., Tittel,L.S. and Winz,M.-L. (2023) Molecular highway patrol for ribosome collisions. *Chem. Biol. Chem.*, **24**, e202300264.
158. Jiang,W., Wagner,J., Du,W., Plitzko,J., Baumeister,W., Beck,F. and Guo,Q. (2022) A transformation clustering algorithm and its application in polyribosomes structural profiling. *Nucleic Acids Res.*, **50**, 9001–9011.
159. Woodward,K. and Shirokikh,N.E. (2021) Translational control in cell ageing: an update. *Biochem. Soc. Trans.*, **49**, 2853–2869.
160. Metkar,M., Pepin,C.S. and Moore,M.J. (2024) Tailor made: the art of therapeutic mRNA design. *Nat. Rev. Drug Discov.*, **23**, 67–83.

Received: January 9, 2024. **Revised:** March 21, 2024. **Editorial Decision:** April 11, 2024. **Accepted:** April 25, 2024