



Start codon-associated ribosomal frameshifting mediates nutrient stress adaptation

Received: 6 December 2022

Accepted: 7 September 2023

Published online: 13 November 2023

 Check for updates

Yuanhui Mao ^{1,2,3}, Longfei Jia^{1,3}, Leiming Dong¹, Xin Erica Shu¹ & Shu-Bing Qian ¹ 

A translating ribosome is typically thought to follow the reading frame defined by the selected start codon. Using super-resolution ribosome profiling, here we report pervasive out-of-frame translation immediately from the start codon. Start codon-associated ribosomal frameshifting (SCARF) stems from the slippage of ribosomes during the transition from initiation to elongation. Using a massively paralleled reporter assay, we uncovered sequence elements acting as SCARF enhancers or repressors, implying that start codon recognition is coupled with reading frame fidelity. This finding explains thousands of mass spectrometry spectra that are unannotated in the human proteome. Mechanistically, we find that the eukaryotic initiation factor 5B (eIF5B) maintains the reading frame fidelity by stabilizing initiating ribosomes. Intriguingly, amino acid starvation induces SCARF by proteasomal degradation of eIF5B. The stress-induced SCARF protects cells from starvation by enabling amino acid recycling and selective mRNA translation. Our findings illustrate a beneficial effect of translational ‘noise’ in nutrient stress adaptation.

Eukaryotic translation initiation involves more than a dozen initiation factors that orchestrate ribosome loading, scanning, and start codon selection^{1,2}. Although the main start codon almost exclusively establishes the primary open reading frame (ORF), frameshifting occurs during elongation, either spontaneously or in a sequence-programmed manner^{3,4}. Another mechanism contributing to out-of-frame translation is alternative initiation, which involves start codons in different reading frames⁵. Because non-AUG codons can serve as potential initiation sites⁶, it is not always apparent whether downstream out-of-frame translation is due to elongation-associated frameshifting or alternative initiation. The key mechanistic difference lies in the fidelity of start codon recognition, as start codon skipping leads to ‘leaky scanning’ and downstream initiation. Once a start codon is recognized by the scanning ribosome, the engagement of the initiator tRNA is followed by 60S joining, a process facilitated by the evolutionarily conserved GTPase eIF5B⁷. While our knowledge about start codon selection is

steadily increasing, the transition of the assembled 80S ribosome from an initiation complex to an elongation complex remains incompletely understood⁸. Important aspects that remain elusive are the timescale of the transition and the mechanisms responsible for quality control of the 80S assembly at the start codon before commitment to elongation.

Ribosome profiling (Ribo-seq) is a powerful technique that provides a snapshot of global translation by sequencing ribosome-protected mRNA fragments (RPFs)^{9,10}. Notably, Ribo-seq has revealed a substantial amount of out-of-frame footprints^{11,12}, but their origins remain largely ambiguous. We do not know exactly, beyond a few examples, whether off-track translation is a result of alternative initiation, frameshifting during elongation or simply experimental noise. Mass spectrometry (MS) is the main methodology that enables direct identification of translational products. Surprisingly, on average, 75% of spectra analyzed in a MS experiment cannot be identified¹³.

¹Division of Nutritional Sciences, Cornell University, Ithaca, NY, USA. ²Present address: Liangzhu Laboratory, Zhejiang University, Hangzhou, China. ³These authors contributed equally: Yuanhui Mao, Longfei Jia. ✉ e-mail: sq38@cornell.edu

Because search engines were built on theoretical spectra derived from user-defined protein sequences, it is possible that the current proteome database is far from complete. Even when all three reading frames are considered, the identity of millions of spectra remains elusive. Without knowing the scope of translational diversity, how translational ‘noise’ contributes to the proteome landscape remains a fundamental knowledge gap.

Here, we re-designed the Ribo-seq methodology by introducing easy RNA-adenylation sequencing (Ezra-seq), which not only simplifies library construction, but also improves the quality of Ribo-seq. With improved resolution, Ezra-seq revealed pervasive out-of-frame translation following the main start codon. We interpreted this translational noise as being a result of start codon-associated frameshifting (SCARF). Notably, we uncovered a regulatory role for eIF5B in controlling SCARF, suggesting that it is part of a quality-control mechanism in maintaining start codon-associated reading frame fidelity. Intriguingly, nutrient stress induces SCARF via eIF5B degradation, implying that translational noise has physiological significance.

Results

Ezra-seq offers high-resolution ribosome profiling

Typical Ribo-seq results are characterized by the 3-nucleotide (nt) periodicity of ribosome-protected mRNA fragments. The percentage of in-frame reads (in-frame ratio, IFR), or phasing, is often used to gauge the quality of Ribo-seq datasets. However, different methods result in varying accuracy of 5'-end reads. For instance, the standard Ribo-seq approach relies on circularization after reverse transcription, which is known to introduce untemplated nucleotide addition¹⁴. The RNA ligation method suffers from sequence-dependent biases¹⁵, resulting in altered footprint quantification. Additionally, varying read lengths often require offset adjustment so that the correct ribosome P-site position can be inferred¹⁴. Despite continuous optimization of Ribo-seq methodology, a substantial number of reads remain out of frame (Extended Data Fig. 1a). We re-designed the Ribo-seq methodology by introducing Ezra-seq (Extended Data Fig. 1b). Without 3'-end linker ligation and 5'-end circularization, Ezra-seq simplifies library construction and, notably, improves the 5'-end accuracy of RPFs with >90% of in-frame reads (Fig. 1a). Furthermore, the ribosome position inferred from the 5' end of footprints is independent of the read length (Fig. 1a, bottom), offering direct determination of P-site positions without in silico adjustment. As expected, Ezra-seq revealed prominent peaks around start and stop codons corresponding to initiating and terminating ribosomes, respectively (Extended Data Fig. 1c). The single-nucleotide resolution of Ezra-seq is highly reproducible in different cell types (Extended Data Fig. 1d). Additionally, Ezra-seq readily detected ribosome pausing sites on proline codon motifs (Extended Data Figs. 1e, f). The accuracy of P-site alignment was further substantiated by the highest variance (97.9%) achieved with a 12-nt offset (Extended Data Fig. 1g).

To affirm the improved resolution of Ezra-seq, we directly compared it with several Ribo-seq datasets that had been obtained using different methods (Extended Data Figs. 2a–h). Ezra-seq clearly offers the highest resolution of footprints, evidenced by increased 5'-end phasing and reduced 5'-end mismatches. Because a start codon is followed by increased in-frame reads downstream (Extended Data Fig. 2i), Ezra-seq's improved resolution allows us to evaluate the initiation potential of all 64 codons. This analysis uncovered 10 triplets, in addition to AUG, that are capable of initiation, as evidenced by a >5% increase of IFR for their downstream regions. The initiation potential of non-AUG codons is highly conserved across different cell types and species (Extended Data Fig. 2j). This result is not only consistent with previous profiling of initiating ribosomes¹⁶, but is also in line with the measurement of translation initiation in *Escherichia coli* using reporter-based assays¹⁷.

Characterizing out-of-frame footprints

Although out-of-frame reads are substantially reduced by Ezra-seq, the remaining noisy footprints are expected to be distributed evenly across individual codons. To our surprise, although ~30% of codons contain perfect in-frame reads, many are deprived of in-frame footprints (Fig. 1b). The wide range of IFRs across individual codons argues against the possibility of technical bias. Intriguingly, metagene analysis revealed an accumulation of out-of-frame reads in the beginning of the coding sequence (CDS) (Fig. 1c, right). Accordingly, the codon IFR averaged across the transcriptome shows a sharp drop (~20%) following the start codon. This feature is consistent regardless of cycloheximide treatment and is even discernable in regular Ribo-seq results (Extended Data Fig. 3a). We surveyed published Ribo-seq datasets and found that the poor codon IFR after the start codon is a common feature across different cell lines (Extended Data Fig. 3b). This feature is not due to the different read lengths of footprints near the start codon (Extended Data Fig. 3c). In fact, the same feature is maintained for reads of different lengths (Extended Data Fig. 3d). If out-of-frame reads represent footprints of actively translating ribosomes, the density of these reads is expected to drop at the out-of-frame stop codons. This is indeed the case (Extended Data Fig. 3e), suggesting that the observed out-of-frame reads are footprints of actively translating ribosomes.

Besides the asymmetric distribution of out-of-frame footprints, we observed large IFR variations across individual transcripts (Extended Data Fig. 4a). The 5' untranslated regions of transcripts with differing IFR values have similar properties, such as folding potential, GC content and length (Extended Data Fig. 4b). Upstream open reading frames (uORFs), when overlapping with the main CDS, result in out-of-frame footprints after the main start codon. However, a direct comparison between transcripts with or without uORF translation showed comparable IFR patterns in the beginning of the CDS (Extended Data Fig. 4c). Because most uORFs use non-AUG initiators, it is possible that overall uORF translation is not robust enough to contribute substantially to out-of-frame reads. Indeed, only overlapping uORFs with the AUG initiator have the lowest IFR (Extended Data Fig. 4d). To further exclude the possibility of hidden uORFs, we selected transcripts with an out-of-frame stop codon before the main start codon (for example, UGAUG). Even after excluding overlapping uORFs, these mRNAs still show an evident drop of IFR at the beginning of the CDS (Extended Data Fig. 4e).

Another possibility is leaky scanning, which often occurs when the start codon is suboptimal¹⁸. Indeed, mRNAs bearing weak start codons (without the Kozak sequence context) tend to have lower IFR values at the beginning of the CDS (Fig. 1d), although a strong start codon is still associated with a substantial number of out-of-frame footprints. Downstream initiation depends on the presence of downstream AUG (dAUG) triplets; however, in most cases, dAUG and its cognate sites (dNUGs) are >6 nt away from the main start codon (Extended Data Fig. 4f). Given the possibility of alternative initiation at non-AUG codons, we surveyed the initiation potential of entire out-of-frame codons immediately downstream of the annotated start codons (Fig. 1e). This analysis revealed that most second codons with poor IFR values possess negligible out-of-frame initiation potential. For a sequence (such as CAUGA) with the strongest out-of-frame initiation potential (that is, +1 AUG), the IFR value at the second codon CAU was barely changed. We conclude that the prevalence of out-of-frame footprints after the main start codon is due to mechanisms beyond leaky scanning.

Codon optimality contributes to out-of-frame footprints

In many organisms, including humans, poorly adapted codons tend to be enriched at the beginning of the CDS (Fig. 2a)¹⁹. It is possible that non-optimal codons promote ribosomal frameshifting and subsequent out-of-frame translation. We examined the reading frame fidelity of footprints aligned to individual codons. Intriguingly, the A-site codon

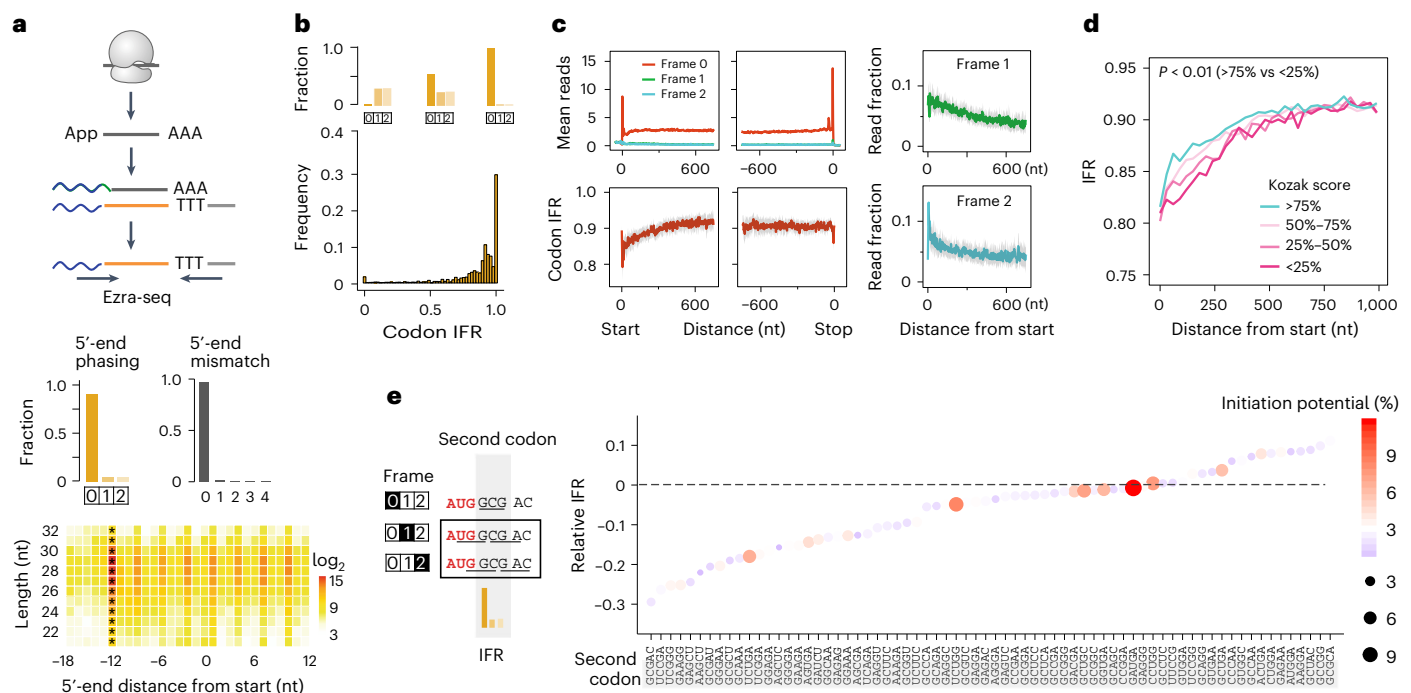


Fig. 1 Ezra-seq reveals prevailing out-of-frame footprints in the beginning of the CDS. **a**, Top, schematic of Ezra-seq procedures. Middle, fraction of reads in different reading frames (5'-end phasing) and the fraction of reads with varying numbers of 5'-end mismatches. Bottom, heatmap showing the distance of the 5' end to the start codon (x axis) for reads of different lengths (y axis). The color code represents the \log_2 (read count). **b**, Top, examples of codons with varied 5'-end phasing, corresponding to different codon in-frame ratios (IFR). Bottom, frequency of codons with different IFRs, with 0 indicating no in-frame reads and 1.0 indicating that all reads were in frame. **c**, Aggregation plots of ribosome density (top) and the IFR of ribosome footprints (bottom) across the transcriptome. Right, the fraction of out-of-frame reads across the

transcriptome. The gray shadow shows the variation of mean IFR estimated by the bootstrap method. Transcripts are aligned to start and stop codons, respectively. **d**, Comparison of IFR between mRNAs bearing start codons with varying strengths of the Kozak sequence context. IFR values are calculated by dividing the in-frame reads by total reads within a non-overlapping sliding window (30 nt). The *P* value was estimated by a permutation test. **e**, A scatter plot of the relative IFR values of the second codons subtracted from the IFR values of the same codons in the CDS, excluding the first 600 nt. The second codon and two downstream nucleotides are shown. The color and size of the dots represent the out-of-frame initiation potential around the second codon, obtained by averaging the initiation potential of four out-of-frame codons.

optimality is positively correlated with IFR ($P = 0.003$) (Fig. 2b, top, and Extended Data Fig. 5a). Two non-optimal serine codons, AGC and AGU, exhibit the lowest IFR values. No such correlation exists when the P-site codon identity is considered (Fig. 2b, right). The same feature holds true when uORF-containing mRNAs are not considered (Extended Data Fig. 5b). The relaxed reading frame fidelity when a non-optimal codon enters the A site suggests that delayed tRNA delivery induces frameshifting. The A-site codon-induced frameshifting is likely influenced by the P-site codon identity³. Indeed, pair-wise codon analysis revealed that certain triplets at the P site, such as the tryptophan codon UGG, coordinate with the non-optimal codons at the A site to lower IFR (Extended Data Fig. 5c).

By grouping transcripts on the basis of their codon optimality at the beginning of the CDS, we found that the presence of non-optimal codons indeed lowered IFR values in a dose-dependent manner (Fig. 2c). Of note, for many transcripts, the strongest codon bias occurs near the start codon. For instance, the non-optimal alanine codon GCG is mainly found at the second codon position, but is barely present in the remainder of the CDS (Extended Data Fig. 5d). To examine the positional effect of codon optimality, we devised a reporter assay by placing the firefly luciferase (Fluc) into different reading frames relative to an uORF (Extended Data Fig. 5e). We eliminated out-of-frame stop codons from the uORF and introduced optimal or non-optimal codons without altering the encoded amino acid sequence. To avoid transcription-associated variation, we synthesized mRNA reporters and monitored Fluc levels in transfected HEK293 cells. The presence of the non-optimal UUG, compared with the optimal CUG, at the second

codon position resulted in an increase in frameshifting of more than threefold (Fig. 2d). Importantly, the same codon replacement at the fourteenth codon near the end of the uORF led to minimal frameshifting, suggesting that initiating ribosomes are more susceptible to frameshifting than are elongating ribosomes (Fig. 2e).

Start codon-associated ribosomal frameshifting

Considering that the 80S ribosome assembled at the start codon has empty E and A sites¹⁶, we propose that the initiating ribosome is susceptible to SCARF. To test this possibility, we constructed SCARF reporters to monitor the translation of uORF in different reading frames (Fig. 3a). The uORF-encoded tracer peptide SIINFEKL is readily presented by the mouse major histocompatibility complex (MHC) class I molecule H-2K^b, and the amount of peptide-K^b complex can be quantified by a monoclonal antibody 25D1 (refs. 20,21). When the uORF is placed in different reading frames relative to the first AUG, any 25D1 signals must be a result of SCARF because neither leaky scanning nor elongation-associated frameshifting would produce the full-length peptides. Again, we used mRNA transfection to exclude transcription-associated variation. Compared with the green fluorescent protein (GFP)-expressing control reporter, which produced background levels of 25D1 signal, SCARF reporters with out-of-frame uORFs clearly produced elevated 25D1 signals in transfected HEK293 cells expressing H-2K^b (Fig. 3a).

Although the SCARF reporter enables us to rule out leaky scanning, it is still possible that SIINFEKL peptides are produced through alternative initiation upstream. We devised two reporters to exclude this possibility. First, we introduced a mutation that changed AUG to ACC,

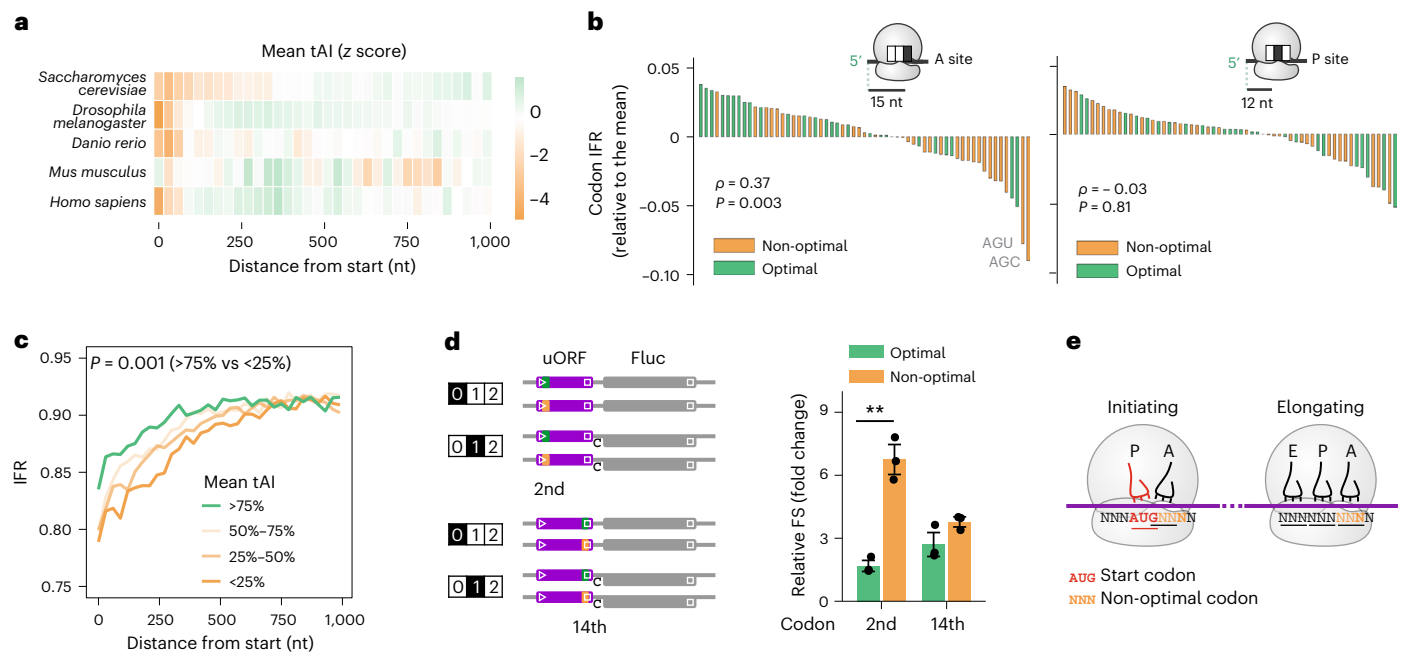


Fig. 2 | Non-optimal codons induce ribosomal frameshifting. **a**, Heat maps showing codon optimality based on the tRNA adaptation index (tAI) along the CDS in different species. tAI values were calculated as a geometric mean of the tAI values of the codons in a non-overlapping sliding window (30 nt) along the CDS. To compare codon optimality across species, the tAI values were normalized by z score. **b**, Bar plots showing the IFR of footprints aligned to individual codons at the ribosome A site (left) or P site (right). Optimal and non-optimal codons are defined on the basis of relative tAI values. Spearman's ρ between tAI and codon IFR values, as well as P values, are indicated. **c**, Comparison of IFR between mRNAs enriched with differential codon optimality in the first 100 codons. IFR values were calculated within a non-overlapping sliding window of 30 nt.

$P < 0.001$ between the top 25% and bottom 25% mRNAs (permutation test). **d**, Left, schematic of uORF reporters with either the second codon or 14th codon replaced by a synonymous optimal (green) or non-optimal (orange) codon. The highlighted numbers in frames indicate the reading frame of Fluc relative to uORF. Right, relative frameshifting rates, determined by the downstream Fluc activities (frame 1/frame 0). Error bars, mean \pm s.e.m.; two-tailed t -test, $n = 3$, $**P = 0.007$. The positions of start and stop codons are indicated by triangles and squares, respectively. **e**, Differential frameshifting susceptibility between initiating and elongating ribosomes in response to non-optimal codons at the A site.

eliminating the main start codon (Fig. 3b) without affecting potential upstream hidden start codons. When AUG was absent, 25D1 signals were abolished, suggesting that SCARF is coupled to the main start codon. Second, we introduced a stop codon in frame 1 between the start codon and the sequence encoding SIINFEKL (Fig. 3c). This design also eliminated the 25D1 signals from transfected cells, providing evidence of an immediate frameshifting event. Therefore, SCARF represents a previously unrecognized phenomenon by enabling out-of-frame translation from the same start codon.

SCARF regulation by sequence contexts

To explore SCARF-associated sequence contexts in an unbiased manner, we used our existing datasets based on a massively paralleled uORF reporter system²¹. By replacing the start codon of uORF with a random 10-nt sequence, this system allows us to identify SCARF enhancers and suppressors from more than one million sequence variants (Fig. 4a). To enrich mRNA variants with uORF translation, we separated mRNA reporters on the basis of the number of associated ribosomes using a sucrose gradient. Given the small uORF size (42 nt), mRNAs with preferential uORF translation are expected to reside in the monosome. When an out-of-frame start codon is present, the mRNA variants tend to be enriched in the polysome, owing to the longer CDS (Fig. 4a). This approach allows us to separate different mRNAs transfected into the same cell. As expected, AUG was predominantly recovered from mRNAs enriched in the monosome fraction (Extended Data Fig. 6a). Intriguingly, mRNAs with AUG placed in out-of-frame positions had comparable monosome/polysome (M/P) ratios (Fig. 4a), suggesting that many initiating ribosomes are not fixed to the reading frame set by the initiator AUG.

Notably, the retrieval of AUG within the 10-nt insert was uneven; it most frequently occurred at the second in-frame position. Unlike the second position, where the -3 nucleotide is fixed, the AUG at the fifth position is surrounded by random sequences. This feature allows us to analyze the sequence context in start codon selection. As expected, the recognition of AUG is primarily dependent on the -3 nucleotide, with the $+4$ nucleotide having minor influence (Extended Data Fig. 6b). The same feature holds true for out-of-frame AUG initiators (Fig. 4b, top, and Extended Data Fig. 6c,d), suggesting that SCARF is also influenced by the Kozak sequence context. Intriguingly, when the AUG is situated within the optimal sequence context (that is, RNNAUGG), the $+5$ nucleotide becomes the most influential in the recovery of out-of-frame AUG codons (Fig. 4b, bottom). Analysis of the $+5$ nucleotide identity revealed that U contributes the least to SCARF (Fig. 4c), indicating that AUGGU suppresses SCARF. Strikingly, the initial IFR analysis of Ezra-seq also uncovered U at the $+5$ nucleotide position for codons with excellent reading frame fidelity (Fig. 1e). As an independent validation, we constructed SCARF reporters with different $+5$ nucleotides and confirmed that AUGGU largely prevents SCARF induced by AUGGA (Fig. 4d). The unexpected contribution of the $+5$ nucleotide identity to SCARF expands the role of the sequence context from recognition of the start codon to reading-frame maintenance.

Detecting endogenous SCARF products

SCARF is expected to increase proteome diversity as a result of out-of-frame translation from the start codon. To detect endogenous SCARF products on a global scale, we surveyed the Proteomics IDentifications (PRIDE) database, the world's largest data repository

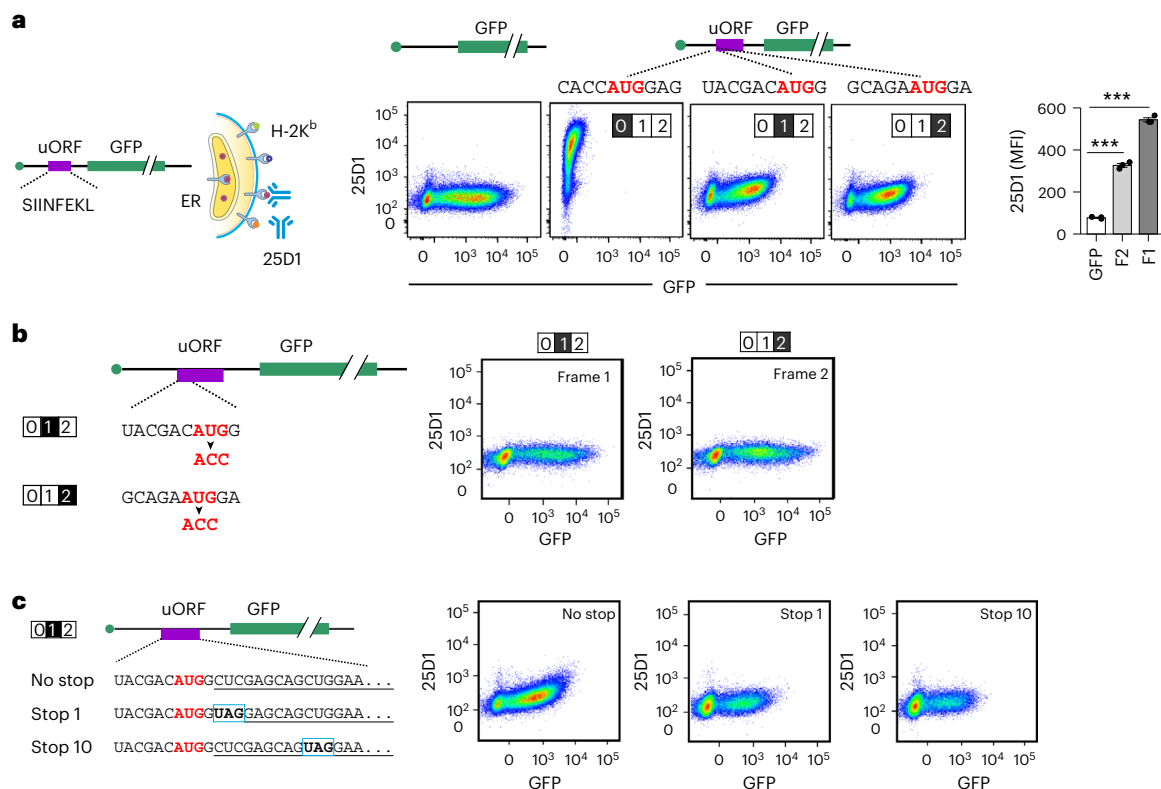


Fig. 3 | Characterizing start codon-associated ribosomal frameshifting. a, Top, a schematic of SCARF reporters that were used to measure start codon-associated frameshifting. The encoded tracer peptide SIINFEKL, presented by H-2K^b, is detected by monoclonal antibody 25D1 and can be quantified using flow cytometry. Bottom, representative flow cytometry scatter plots of HEK293-K^b cells transfected with SCARF reporters or a GFP control. The highlighted numbers in frames indicate the reading frame of the encoded SIINFEKL relative to the AUG codon. The bar graph shows the 25D1 mean fluorescence intensity (MFI) of SCARF reporters and the GFP control. Error bars, mean ± s.e.m.; two-tailed *t*-test, *n* = 3, ****P* = 0.0006 for frame 1 (F1) versus GFP, and *P* = 0.0002 for

frame 2 (F2) versus GFP. **b**, Left, sequence context of the SCARF reporters with an ACC start codon in place of AUG. Right, representative flow cytometry scatter plots of HEK293-K^b cells transfected with the SCARF reporters shown on the left. **c**, The left panel shows the sequence context of the SCARF reporters with an internal stop codon in-frame with SIINFEKL. The highlighted numbers in frames indicate the position of the stop codons. Of note, the region encoding SIINFEKL (underlined) is placed to reading frame 1 relative to the start codon (shown in red). Right, representative flow cytometry scatter plots of HEK293-K^b cells transfected with the SCARF reporters shown on the left.

of MS-based proteomic data²². We first built a human out-of-frame proteome database by including all in silico translation events initiated from frameshifted start codons. Among 18,058 newly identified out-of-frame peptides (Supplementary Table 1), surprisingly the majority accumulated following the main start codon (Fig. 5a). In support of their SCARF origin, transcripts containing out-of-frame peptides exhibit lower IFR values after the start codon (Fig. 5b).

One important feature of SCARF is that it commences with recognition of the main start codon by the initiator tRNA (Met-tRNA^{Met}). Therefore, the first amino acid of SCARF peptides should be methionine, even though their corresponding ORFs lack the AUG codon. By manually adding an extra M residue at the amino termini (position 0) of out-of-frame peptides, we uncovered 1,403 peptides starting with M (Supplementary Table 1). This number is likely an underestimate because some M residues are removed post-translationally²³, leaving a perfect match from position 1 (a total of 685 peptides). Notably, 62.7% of the extra N-terminal M residues are acetylated, an orthogonal signature of translation initiation²⁴, indicating that they are bona fide SCARF products. We next examined whether SCARF peptides could be detected in a single database. From a multi-plex tandem mass tagging (TMT)-based proteomics study²⁵, we uncovered 167 peptides resulting from out-of-frame translation (OFT); 65 were derived from the start codon (Fig. 5c and Supplementary Table 2). To further substantiate this finding, we examined a MS dataset of N-terminal peptides that had been enriched using terminal amine isotopic labeling of substrates²⁶.

Among 45 OFT peptides, 16 originated from SCARF products (Supplementary Table 3).

Because strong start codons like AUGG contribute to the majority of translational products, we examined the identity of amino acids at position 1 to shed light on the molecular details of SCARF (Fig. 5d). In the PRIDE database, about one-third of the SCARF peptides contain glycine at position 1, an indication of -1 frameshifting because G is encoded by GGN codons; 21% contain tryptophan, an indication of -2 frameshifting because W is encoded exclusively by UGG (Fig. 5d). These results not only exclude alternative initiation, but strongly suggest that SCARF is due to unstable interaction between the P-site initiator tRNA and the start codon, thereby permitting A-site tRNA mispairing (Fig. 5e).

SCARF regulation by eIF5B

We next wondered whether SCARF is subjected to regulation. To search for potential SCARF regulators, we focused on initiation factors involved in start codon recognition and 60S joining (Fig. 6a). Upon start codon recognition, eIF1 dissociation enables eIF5 to interact with eIF1A, forming a closed initiation complex to arrest scanning²⁷. Mutants of eIF1 have been shown to reduce the stringency of start codon selection²⁸. However, silencing of eIF1 had a negligible effect on IFR (Extended Data Fig. 7a), suggesting that SCARF occurs after start codon recognition. Similarly, eIF5 overexpression did not alter the IFR pattern (Extended Data Fig. 7b), although it did compromise the stringency of start codon recognition²⁹. Unlike eIF1 knockdown, eIF1A depletion lowered the IFR

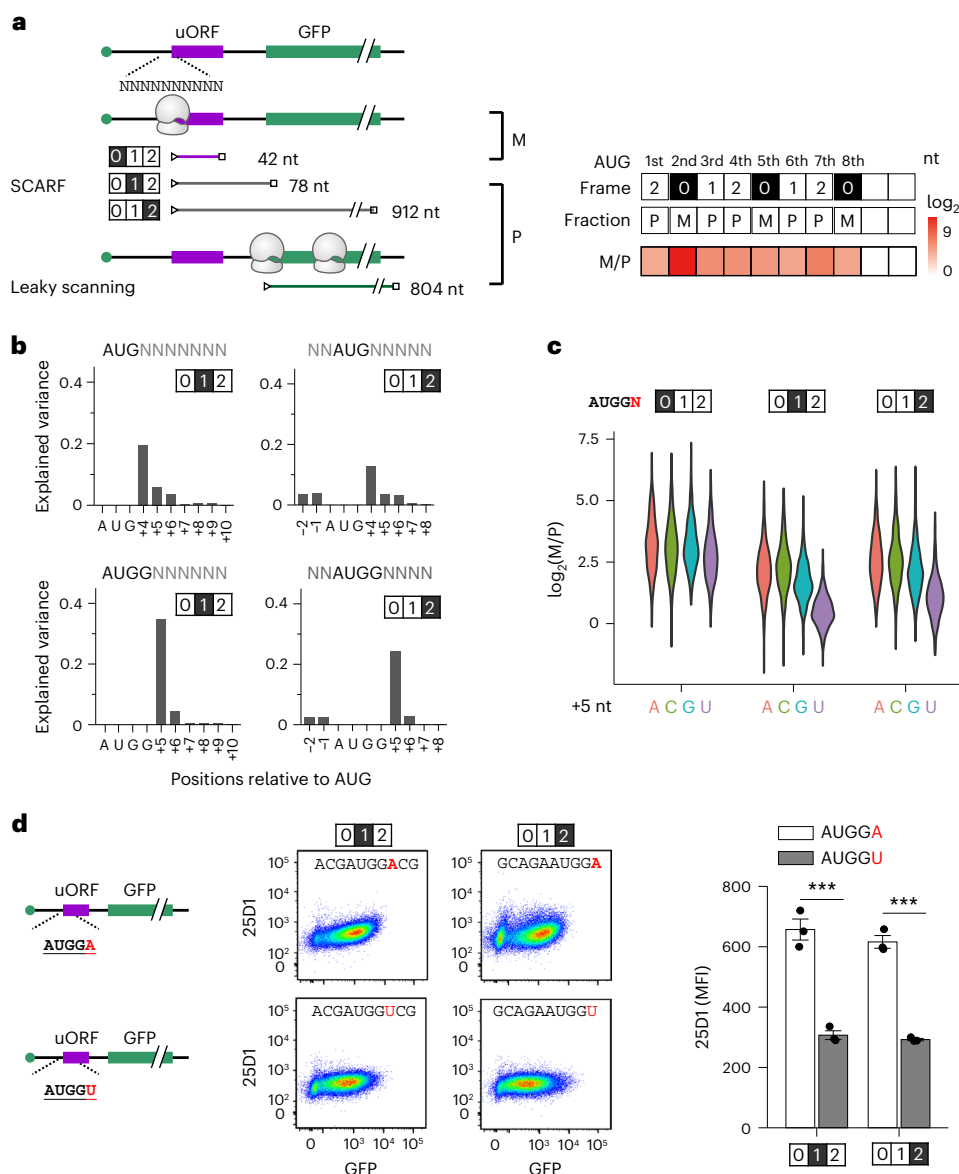


Fig. 4 | SCARF is dependent on sequence context. **a**, Left, schematic of a massively parallel SCARF reporter assay in which the start codon of the uORF is replaced with a random 10-nt sequence. uORF translation is monitored using the number of associated ribosomes separated by a sucrose gradient. Right, the reading frames relative to the uORF when AUG was placed at various locations, and the expected ribosome fractions (P, polysome; M, monosome). The heatmap shows the M/P ratio resulting from the placement of AUG at different positions. **b**, Relative contributions of the nucleotide identity at different positions to uORF translation, on the basis of M/P ratio. The highlighted numbers indicate the reading frame of the encoded SIINFEKL relative to the AUG codon. **c**, The influence of the identity of the +5 nucleotide on M/P ratios when a Kozak initiation context (RNNAUGG) is placed in different frames. The highlighted

numbers indicate the reading frame of the encoded SIINFEKL relative to the AUG codon. A two-sided Wilcoxon test was performed to test the null hypothesis that the +5 U yields the same M/P ratio as the other nucleotides. For out-of-frame reporters, all P values were $<2.2 \times 10^{-16}$. **d**, Left, schematic of SCARF reporter assay. Middle, representative flow cytometry scatter plots of HEK293-K⁰ cells transfected with SCARF reporters bearing a repressor (thymine at fifth position relative to start codon) or enhancer (adenine at fifth position relative to start codon). Right, relative 25D1 MFI of SCARF reporters. The highlighted numbers refer to the reading frame of the encoded SIINFEKL relative to the AUG codon. Error bars, mean \pm s.e.m.; two-tailed t -test, $n = 3$, *** $P = 0.004$ for the comparison of frame 1, and $P = 0.003$ for the comparison of frame 2.

immediately after the start codon (Extended Data Fig. 7c). It is worth mentioning that silencing eIF1A led to severe cellular toxicity, presumably owing to the pleiotropic functions of eIF1A in nearly all steps of initiation, including 60S joining. The 60S joining is catalyzed by eIF5B⁷, which also interacts with eIF1A³⁰. When eIF5B was depleted, the IFR values at the beginning of the CDS were reduced (Extended Data Fig. 7d). To exclude the possibility that the lowered IFR is an artifactual noise of reduced global translation, we separated mRNAs with differential translation status in cells lacking eIF5B. Remarkably, only

mRNAs with high ribosome occupancy experienced reduced IFR (Fig. 6b), indicating that SCARF is coupled to active translation. Supporting the crucial role for eIF5B in start codon-associated reading frame fidelity, SCARF reporter assays revealed increased signals of 25D1, but not GFP, in the absence of eIF5B (Fig. 6c and Extended Data Fig. 7e).

As a universally conserved initiation factor, eIF5B gates the transition of 80S from initiation to elongation. Indeed, depletion of eIF5B in HEK293 cells markedly elevated ribosome density at the start codon (Fig. 6d). An earlier study has reported that eIF5B tends to have longer

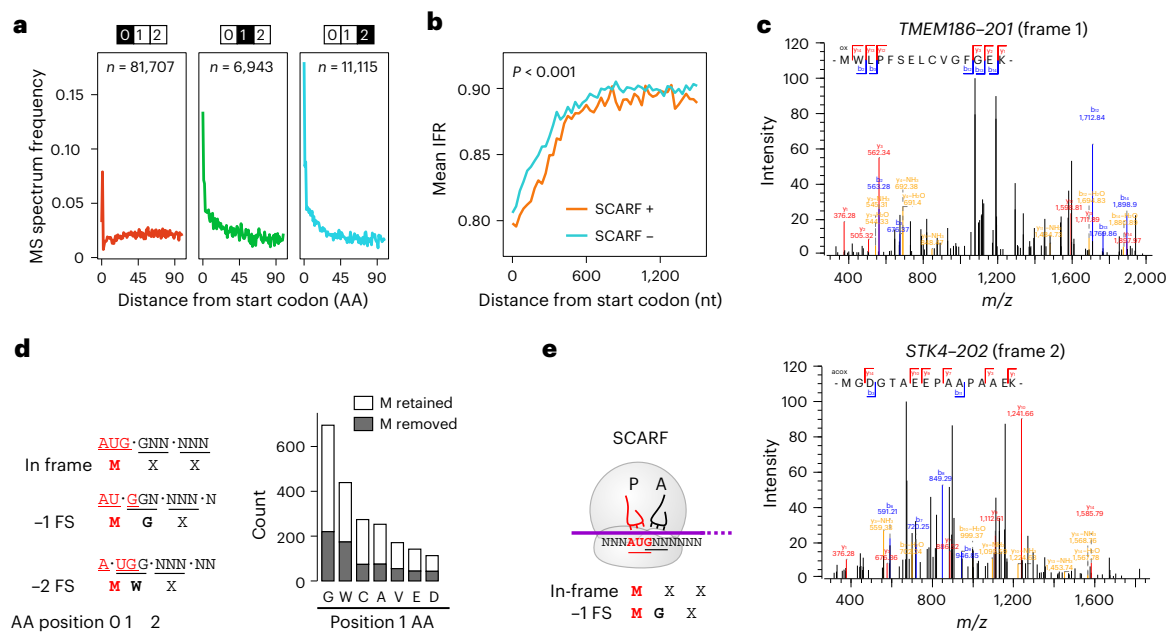


Fig. 5 | Detecting endogenous SCARF products. **a**, The frequency of mass spectra mapped to the in-frame and out-of-frame peptides was calculated using the PRIDE database. The out-of-frame spectra were obtained by searching unidentified spectra against the out-of-frame proteome database containing in silico translation of SCARF. The highlighted numbers indicate the reading frame. **b**, Comparison of IFR between mRNAs with (SCARF +) or without (SCARF -) out-of-frame peptides, generated from SCARF. IFR values were calculated within a non-overlapping sliding window of 30 nt. *P* values were calculated using

a permutation test. **c**, Representative examples of mass spectra²⁵ of N-terminal out-of-frame peptides from genes *TMEM186-201* and *STK4-202*. **d**, The count of different types of amino acids at the first N-terminal amino acid (AA) position after methionine in out-of-frame peptides. M retained, the first methionine is detected at the first N-terminal residue; M removed, the first methionine is not detected at the first N-terminal residue. **e**, Schematic of the process through which SCARF produces out-of-frame peptides.

residence time on initiating 80S ribosomes formed at non-Kozak start codons⁸. Intriguingly, transcripts bearing weak start codons are more susceptible to SCARF in cells lacking eIF5B (Extended Data Fig. 8a). A start codon without the Kozak sequence is often associated with leaky scanning¹⁸. To distinguish SCARF from leaky scanning, a new reporter was needed because the SIINFEKL-based reporter is not sensitive enough to measure SCARF from weak start codons. We replaced the tracer peptide SIINFEKL with a nano-luciferase-derived peptide HiBit that can be quantified by luminometry with exquisite sensitivity³¹ (Extended Data Fig. 8b). By placing HiBit at the reading frame 1, we confirmed that the HiBit signal could be used to assess SCARF, because mutating the AUG start codon to ACC eliminated HiBit signals (Extended Data Fig. 8c). To measure leaky scanning in parallel, we inserted the Fluc sequence further downstream. This design allows us to monitor both SCARF and leaky scanning in the same samples by measuring HiBit and Fluc levels, respectively. As expected, a weak start codon led to an increase in leaky scanning of >20-fold (Extended Data Fig. 8d, left). Upon normalization to the in-frame HiBit signals, the same reporters showed a fivefold increase of SCARF when the start codon was switched from strong to weak (from 4% to 21%) (Fig. 6e and Extended Data Fig. 8d, right). Therefore, a start codon without the Kozak sequence tends to be skipped, resulting in leaky scanning. But once selected, the assembled 80S ribosome is susceptible to SCARF.

In cells lacking eIF5B, we readily observed increased SCARF, as evidenced by the elevated HiBit signals relative to the in-frame translation (Fig. 6e). Notably, eIF5B depletion nearly doubled SCARF when the start codon lacks the Kozak sequence. The unexpected role of eIF5B in maintaining the start codon reading frame is consistent with the unique position of eIF5B within the initiating 80S complex^{30,32}. By stabilizing the initiator tRNA at the P site (Extended Data Fig. 8e), the presence of eIF5B likely prevents the slippage of the initiating ribosome when the start codon is suboptimal.

Nutrient stress induces SCARF via eIF5B degradation

Despite quality control by eIF5B in maintaining the start codon-associated reading frame fidelity, SCARF is seemingly an energy-wasting process by generating degradative polypeptides. Cells under nutrient starvation require intracellular amino acid recycling to support essential protein synthesis³³. To examine whether nutrient stress induces SCARF, we applied Ezra-seq to HEK293 cells with or without amino acid deprivation. Remarkably, the IFR in starved cells was substantially reduced at the beginning of the CDS (Fig. 7a). This was not due to repressed global protein synthesis because only mRNAs with relatively high ribosome occupancy experienced increased SCARF (Extended Data Fig. 9a). Amino acid starvation-induced SCARF was further confirmed by SIINFEKL-based SCARF reporters (Fig. 7b). HiBit-based SCARF reporters further revealed that mRNAs with weak start codons are more susceptible to SCARF in response to starvation (Fig. 7c and Extended Data Fig. 9b).

The similar effects that eIF5B knockdown and nutrient starvation have on SCARF prompted us to examine whether starvation modulates eIF5B. Intriguingly, amino acid deprivation decreased the steady-state levels of eIF5B, but not those of other initiation factors like eIF4E or eIF4A1 (Fig. 7d). We found that eIF5B underwent a faster turnover upon nutrient stress (Extended Data Fig. 9c, top) and was stabilized by the proteasome inhibitor MG132 (Extended Data Fig. 9c, bottom). Supporting the notion that nutrient-stress-induced SCARF is a result of eIF5B degradation, MG132 treatment rescued SCARF by 30% in starved cells (Extended Data Fig. 9d). Additionally, expression of exogenous eIF5B in starved cells suppressed SCARF (Extended Data Fig. 9e, f). This result prompted us to conduct rescue experiments using eIF5B mutants. Recent structural studies have revealed that there is close contact between domain IV of eIF5B and the acceptor stem of initiator tRNA, whereas the GTPase domain is near the mRNA entry site (Extended Data Fig. 10a)³². We created a G domain mutant (p.T665A) and

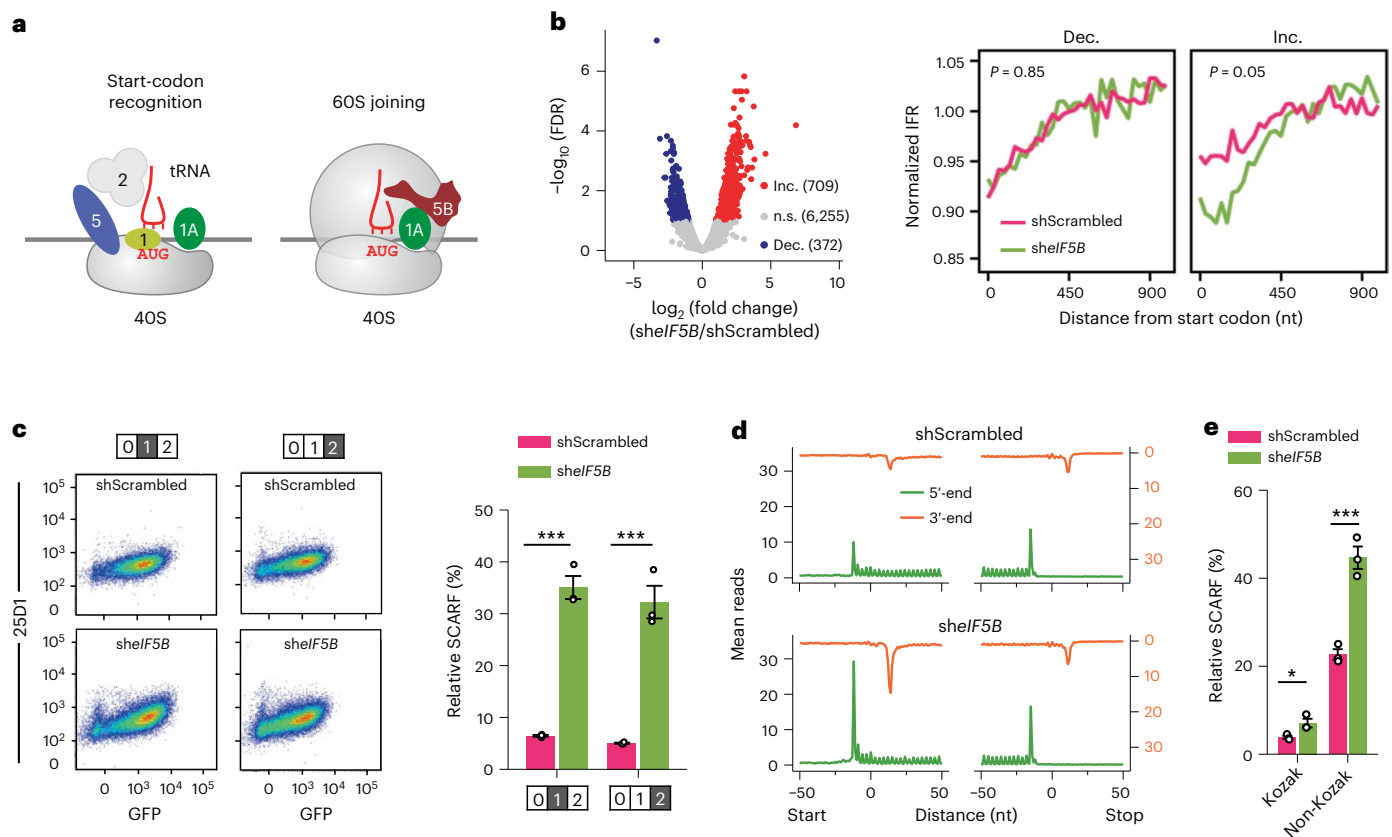


Fig. 6 | SCARF regulation by eIF5B. **a**, Schematic of translation-initiation factors involved in start codon recognition and 60S subunit joining. **b**, Left, fold change of ribosome density after eIF5B knockdown. The mRNAs with significantly increased ribosome density (Inc.; false discovery rate (FDR) < 0.05) are shown in red. The mRNAs with significantly decreased ribosome density (Dec.; FDR < 0.05) are shown in blue. n.s., not significant. Right, comparison of IFR along the CDS in the two mRNA groups in cells with or without eIF5B knockdown. IFR values were calculated within a non-overlapping sliding window of 45 nt, which was subsequently normalized by the CDS IFR. *P* values were calculated by a permutation test. **c**, Left, representative flow cytometry scatter plots of HEK293-K₆ cells transfected with SCARF reporters with or without eIF5B knockdown by

short hairpin RNA (shRNA). Right, the relative SCARF rate, calculated by the 25D1 mean fluorescence intensity (MFI) of SCARF reporters over the in-frame control. The highlighted numbers refer to the reading frame of the encoded SIINFEKL. Error bars, mean \pm s.e.m.; two-tailed *t*-test, *n* = 3, ****P* = 0.0002 for the comparison of frame 1, and *P* = 0.001 for the comparison of frame 2. **d**, Aggregation plots show the ribosome density across the transcriptome in cells with or without eIF5B knockdown. Both the 5'-end (green) and 3'-end (orange) of footprints were included. **e**, The relative SCARF rate in HEK293 cells with or without eIF5B knockdown. The relative SCARF rate was calculated using the HiBit activity (frame 1/ frame 0). Error bars, mean \pm s.e.m.; two-tailed *t*-test, *n* = 3, **P* = 0.04 (Kozak), ****P* = 0.001 (non-Kozak).

a domain IV mutant (p.R1174A) and transfected cells with these mutants and measured SCARF. Whereas the wild-type eIF5B readily repressed starvation-induced SCARF by 50%, neither the G domain mutant nor the domain IV mutant showed any effects (Fig. 7e and Extended Data Fig. 10b). Therefore, both GTP hydrolysis and interaction with the initiator tRNA are crucial for eIF5B in maintaining the reading frame fidelity during the transition from initiation to elongation.

Nutrient starvation-induced SCARF could supply degradative materials for intracellular amino acid recycling. In line with this notion, either amino acid deprivation or eIF5B silencing resulted in marked accumulation of polyubiquitinated species, which were rapidly degraded (Fig. 7f). In cells subjected to both starvation and eIF5B knockdown, we observed the largest accumulation of polyubiquitinated signals. Additionally, eIF5B overexpression reduced degradative materials in starved cells (Extended Data Fig. 10c). Internal amino acid supply is essential for selective translation of mRNAs, such as ATF4, which has a role in cell survival during starvation³⁴. Indeed, eIF5B overexpression in starved cells not only suppressed SCARF but also reduced ATF4 expression (Extended Data Fig. 10d,e), which was not seen after using eIF5B mutants. We predict that SCARF protects cells from starvation by enabling amino acid recycling and selective mRNA translation. This is indeed the case. Overexpression of wild-type eIF5B, but not the

G domain mutant (p.T665A) or domain IV mutant (p.R1174A), decreased the cell viability during prolonged starvation (Fig. 7g). We conclude that SCARF represents a cellular adaptation mechanism that is crucial for amino acid homeostasis during nutrient starvation.

Discussion

Our study demonstrates that start codon recognition does not always guarantee the subsequent reading frame used in translation. The unique feature of initiating ribosomes necessitates additional quality-control mechanisms to ensure reading frame fidelity, especially when the start codon is suboptimal. Although it has been well-established that the Kozak sequence facilitates start codon selection, >50% of human transcripts do not contain optimal start codons. A weak start codon not only leads to leaky scanning, but is also susceptible to SCARF, a previously unappreciated phenomenon. SCARF is distinct from alternative initiation because it is coupled to the main start codon. SCARF also differs from elongation-associated frameshifting because it starts with the initiator tRNA_{i^{Met}}. As a result, SCARF products are often overlooked in proteomic studies because of out-of-frame translation with a mismatched first amino acid. Although Ezra-seq estimates that ~10% of translation initiation events are subjected to SCARF, most frameshifted peptides are short-lived. However, certain SCARF products could have

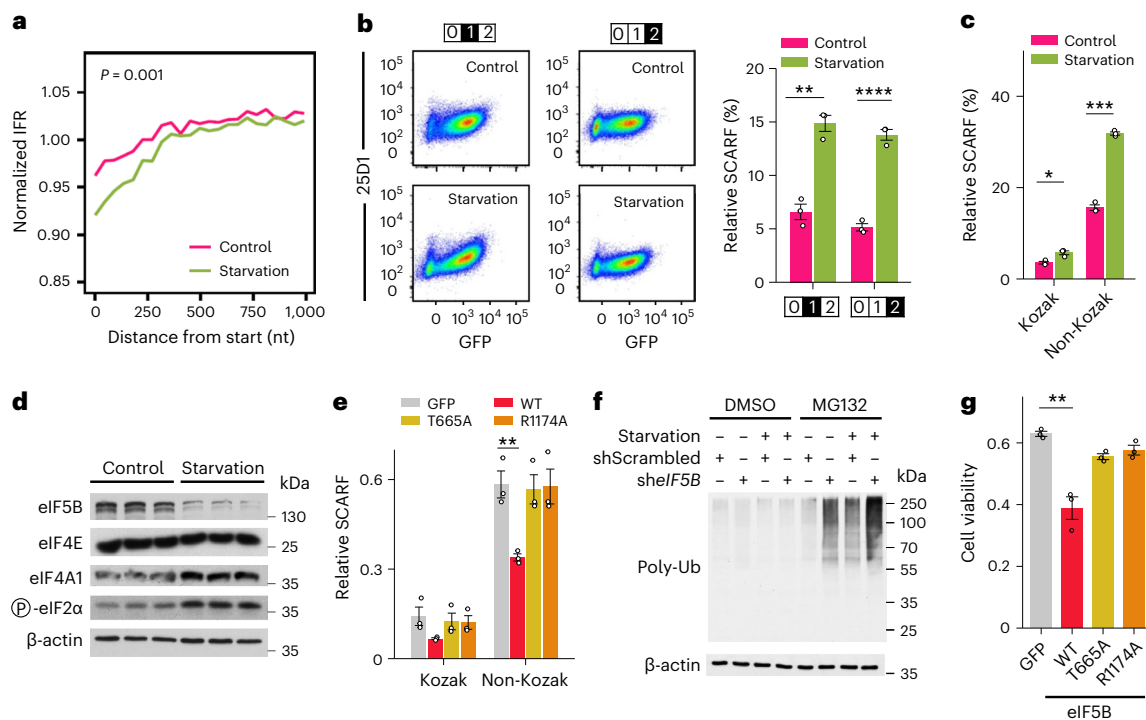


Fig. 7 | SCARF can be induced by amino acid starvation. **a**, Comparison of normalized IFR in cells with or without amino acid starvation. IFR values were calculated within a non-overlapping sliding window of 45 nt, which was subsequently normalized to the CDS IFR. P values are calculated using a permutation test. **b**, Left, representative flow cytometry scatter plots of HEK293-K³ cells transfected with SCARF reporters before and after amino acid starvation. Right, the relative SCARF rate, calculated using the 25D1 MFI of SCARF reporters over the in-frame control. The highlighted numbers refer to the reading frame of the encoded SIINFEKL relative to the AUG codon. Error bars, mean \pm s.e.m.; two-tailed t -test, $n = 3$, $**P = 0.001$ for the comparison of frame 1, $***P = 0.0001$ for the comparison of frame 2. **c**, The relative SCARF rate in HEK293 cells after 16 h of amino acid starvation. The SCARF rate was determined using HiBit activity (frame 1/ frame 0). Error bars, mean \pm s.e.m.; two-tailed

t -test, $n = 3$, $*P = 0.01$ for the comparison of Kozak context, $***P = 2.2 \times 10^{-5}$ for the comparison of non-Kozak context. **d**, Representative western blots of translation initiation factors in HEK293 cells before and after 16 h of amino acid starvation. The experiment was independently repeated three times, with similar results. P-eIF2 α , phosphorylated eIF2 α . **e**, The relative SCARF rate in starved HEK293 cells transfected with GFP, wild-type eIF5B, or mutated eIF5B (T665A or R1174A) plasmids. The SCARF rate was determined using HiBit activity (frame 1/ frame 0). Error bars, mean \pm s.e.m.; two-tailed t -test, $n = 3$, $**P = 0.006$. **f**, Representative western blots of polyubiquitinated species in HEK293 cells with or without eIF5B knockdown, in the absence or presence of MG132, before and after amino acid starvation. The experiment was independently repeated three times, with similar results. **g**, Viability of transfected HEK293 cells after 16 h of amino acid starvation. Error bars, mean \pm s.e.m.; two-tailed t -test, $n = 3$, $**P = 0.03$.

biological functions and contribute to proteome diversity. The SCARF reporters established here can be readily adapted to distinguish start codon-associated frameshifting from other non-canonical translation events.

The promiscuous translation from the start codon suggests that the transition from initiation to elongation is more complex than previously thought. The GTPase eIF5B is involved in the correct positioning of the initiator tRNA_i^{Met} on the 80S ribosome, thereby controlling the elongation commitment^{8,32}. We propose that eIF5B plays a crucial role in maintaining the reading frame fidelity during the transition from initiation to elongation. Intriguingly, eIF5B levels fluctuate during stress and along different developmental stages^{35,36}. It is likely that eIF5B senses environmental cues, thereby contributing to stress adaptation by increasing translational diversity from existing mRNAs. Nutrient stress rapidly alters the proteome landscape via translational reprogramming³⁷. Upon amino acid deprivation, general protein synthesis is rapidly suppressed, but a subset of mRNAs undergoes selective translation. To support selective protein synthesis, degradative systems are activated to recycle intracellular amino acids. However, it remains debated which protein sources are preferentially allocated for degradation. Previous studies have proposed that a ribosome autophagy (ribophagy) pathway supplies internal amino acids during acute nutrient stress³⁸, but systematic quantification of ribosome inventory showed minimal ribosome degradation²⁵. Considering that eIF5B degradation can be induced

by stress, we hypothesize that the subsequently increased SCARF products provide a degradative source for intracellular amino acid recycling during starvation.

A central tenet of biology is the accurate flow of genetic information from nucleic acids to proteins. Biological noise, however, is often overlooked. Translational noise derived from SCARF is analogous to divergent transcription at the promoter region by RNA polymerase II³⁹. Exploring translational noise under different experimental conditions could indicate whether translational infidelity arises from error or represents a potential feature conferring an advantage. For instance, tRNA misacylation and ribosome recoding have been shown to protect cells from oxidative stress⁴⁰. On the basis of the the nutrient stress-inducible nature of SCARF, its physiological significance is twofold: first, it offers an additional means to turn off translation of mRNAs that are already engaged with ribosomes. Second, it serves as an immediate source for intracellular amino acid recycling, to enable selective protein synthesis during prolonged starvation. It is equally possible that some SCARF products could be functional by acting as signaling factors. Broadly, the discovery of SCARF makes sense of translational noise, illustrating the beneficial effect of translational diversity in nutrient stress adaptation.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions

and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41594-023-01119-z>.

References

- Pelletier, J. & Sonenberg, N. The organizing principles of eukaryotic ribosome recruitment. *Annu. Rev. Biochem.* **88**, 307–335 (2019).
- Hinnebusch, A. G. Structural insights into the mechanism of scanning and start codon recognition in eukaryotic translation initiation. *Trends Biochem. Sci.* **42**, 589–611 (2017).
- Atkins, J. F. & Bjork, G. R. A gripping tale of ribosomal frameshifting: extragenic suppressors of frameshift mutations spotlight P-site realignment. *Microbiol. Mol. Biol. Rev.* **73**, 178–210 (2009).
- Dinman, J. D. Mechanisms and implications of programmed translational frameshifting. *Wiley Interdiscip. Rev. RNA* **3**, 661–673 (2012).
- Orr, M. W., Mao, Y., Storz, G. & Qian, S. B. Alternative ORFs and small ORFs: shedding light on the dark proteome. *Nucleic Acids Res.* **48**, 1029–1042 (2020).
- Kearse, M. G. & Wilusz, J. E. Non-AUG translation: a new start for protein synthesis in eukaryotes. *Genes Dev.* **31**, 1717–1731 (2017).
- Pestova, T. V. et al. The joining of ribosomal subunits in eukaryotes requires eIF5B. *Nature* **403**, 332–335 (2000).
- Wang, J. et al. eIF5B gates the transition from translation initiation to elongation. *Nature* **573**, 605–608 (2019).
- Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. & Weissman, J. S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223 (2009).
- Brar, G. A. & Weissman, J. S. Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nat. Rev. Mol. Cell Biol.* **16**, 651–664 (2015).
- Ingolia, N. T., Lareau, L. F. & Weissman, J. S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789–802 (2011).
- Gao, X. et al. Quantitative profiling of initiating ribosomes in vivo. *Nat. Methods* **12**, 147–153 (2015).
- Griss, J. et al. Recognizing millions of consistently unidentified spectra across hundreds of shotgun proteomics datasets. *Nat. Methods* **13**, 651–656 (2016).
- Ingolia, N. T., Brar, G. A., Rouskin, S., McGeachy, A. M. & Weissman, J. S. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat. Protoc.* **7**, 1534–1550 (2012).
- Hafner, M. et al. RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. *RNA* **17**, 1697–1712 (2011).
- Lee, S., Liu, B., Huang, S. X., Shen, B. & Qian, S. B. Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc. Natl Acad. Sci. USA* **109**, E2424–E2432 (2012).
- Hecht, A. et al. Measurements of translation initiation from all 64 codons in *E. coli*. *Nucleic Acids Res.* **45**, 3615–3626 (2017).
- Kozak, M. Pushing the limits of the scanning mechanism for initiation of translation. *Gene* **299**, 1–34 (2002).
- Tuller, T. et al. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* **141**, 344–354 (2010).
- Dersh, D., Yewdell, J. W. & Wei, J. A SIINFEKL-based system to measure MHC class I antigen presentation efficiency and kinetics. *Methods Mol. Biol.* **1988**, 109–122 (2019).
- Jia, L. et al. Decoding mRNA translatability and stability from the 5' UTR. *Nat. Struct. Mol. Biol.* **27**, 814–821 (2020).
- Perez-Riverol, Y. et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2019).
- Xiao, Q., Zhang, F., Nacev, B. A., Liu, J. O. & Pei, D. Protein N-terminal processing: substrate specificity of *Escherichia coli* and human methionine aminopeptidases. *Biochemistry* **49**, 5588–5599 (2010).
- Aksnes, H., Hole, K. & Arnesen, T. Molecular, cellular, and physiological significance of N-terminal acetylation. *Int. Rev. Cell Mol. Biol.* **316**, 267–305 (2015).
- An, H., Ordureau, A., Korner, M., Paulo, J. A. & Harper, J. W. Systematic quantitative analysis of ribosome inventory during nutrient stress. *Nature* **583**, 303–309 (2020).
- Na, C. H. et al. Discovery of noncanonical translation initiation sites through mass spectrometric analysis of protein N termini. *Genome Res* **28**, 25–36 (2018).
- Hinnebusch, A. G. The scanning mechanism of eukaryotic translation initiation. *Annu. Rev. Biochem.* **83**, 779–812 (2014).
- Thakur, A. & Hinnebusch, A. G. eIF1 Loop 2 interactions with Met-tRNAi control the accuracy of start codon selection by the scanning preinitiation complex. *Proc. Natl Acad. Sci. USA* **115**, E4159–E4168 (2018).
- Tang, L. et al. Competition between translation initiation factor eIF5 and its mimic protein 5MP determines non-AUG initiation rate genome-wide. *Nucleic Acids Res.* **45**, 11941–11953 (2017).
- Nag, N. et al. eIF1A/eIF5B interaction network and its functions in translation initiation complex assembly and remodeling. *Nucleic Acids Res.* **44**, 7441–7456 (2016).
- Dixon, A. S. et al. NanoLuc complementation reporter optimized for accurate measurement of protein interactions in cells. *ACS Chem. Biol.* **11**, 400–408 (2016).
- Wang, J. et al. Structural basis for the transition from translation initiation to elongation by an 80S–eIF5B complex. *Nat. Commun.* **11**, 5003 (2020).
- Suraweera, A., Munch, C., Hanssum, A. & Bertolotti, A. Failure of amino acid homeostasis causes cell death following proteasome inhibition. *Mol. Cell* **48**, 242–253 (2012).
- Kilberg, M. S., Shan, J. & Su, N. ATF4-dependent transcription mediates signaling of amino acid limitation. *Trends Endocrinol. Metab.* **20**, 436–443 (2009).
- Lee, S. et al. Upregulation of eIF5B controls cell-cycle arrest and specific developmental stages. *Proc. Natl Acad. Sci. USA* **111**, E4315–E4322 (2014).
- Chukka, P. A. R., Wetmore, S. D. & Thakor, N. Established and emerging regulatory roles of eukaryotic translation initiation factor 5B (eIF5B). *Front Genet* **12**, 737433 (2021).
- Shu, X. E., Swanda, R. V. & Qian, S. B. Nutrient control of mRNA translation. *Annu. Rev. Nutr.* **40**, 51–75 (2020).
- Wyant, G. A. et al. NUFIP1 is a ribosome receptor for starvation-induced ribophagy. *Science* **360**, 751–758 (2018).
- Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* **13**, 720–731 (2012).
- Netzer, N. et al. Innate immune and chemically triggered oxidative stress modifies translational fidelity. *Nature* **462**, 522–526 (2009).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2023

Methods

Cell lines and reagents

HEK293-K^b cells and Lenti-X 293T cells were maintained in DMEM (Corning 10-013-CV) with 10% FBS (Sigma 12306C). Antibodies used in immunoblotting experiments were: anti-eIF5B (Proteintech, 13527-1-AP, 1:500), anti-eIF4E (Cell Signaling, 9742 S, 1:1,000), anti-eIF4A1 (Abcam, ab31217, 1:1,000), anti-eIF4E (Cell Signaling, 9742S, 1:1,000), anti-Phospho-eIF2 α (Ser51) (Cell Signaling, 3398S, 1:1,000), anti-eIF1A (Abcam, ab177939, 1:1,000), anti-eIF1 (Proteintech, 15276-1-AP, 1:1,000), anti-eIF5 (Santa Cruz Biotechnology, sc-282, 1:200), anti-Ubiquitin (Santa Cruz Biotechnology P4D1, sc-8017, 1:200), anti-ATF4 (Santa Cruz Biotechnology, sc-390063, 1:200) and anti- β -actin (Sigma, A5441, 1:5,000).

Construction of plasmids and reporters

The full-length coding sequence of Fluc gene was cloned into the pcDNA3.1 vector (Invitrogen) to generate Fluc-pcDNA3.1. PCR was performed to generate products of the Fluc-based standard frameshifting reporter and the HiBit-based SCARF reporters using Fluc-pcDNA3.1 as a template. PCR products of SIINFEKL-based reporters were generated using a two-step PCR amplification approach. First, full-length SIINFEKL followed by enhanced GFP (eGFP) was amplified from pcDNA3-EGFP to generate SIINFEKL-eGFP. The resulting PCR product was used as a template to produce the full-length reporter using a second forward primer. For exogenous eIF5B, a truncated human eIF5B (580–1220 aa) was cloned into pcDNA3.1(myc-His B) using BamH I and Not I restriction sites. To create the eIF5B mutant, site-directed mutagenesis was performed using Q5 Site-Directed Mutagenesis Kit (New England Biolabs) according to the manufacturer's manual. Mutation was confirmed by Sanger DNA sequencing. DNA sequences of all primers used in this study are listed in Supplementary Table 4.

In vitro transcription

To prepare mRNA reporters, 1–2 μ g of the PCR products described above was utilized as templates to generate mRNAs suitable for transfection. In vitro transcription was performed with mMESAGE mMACHINE T7 transcription kit (Ambion), followed by poly(A) tailing kit (Ambion). mRNAs were purified following the manufacturer's instructions and were eluted in nuclease-free water.

Transfection

For mRNA reporter transfection, cells were transfected with 3 μ g mRNA at a mass ratio of 1:1 using Lipofectamine MessengerMAX (Invitrogen), according to the manufacturer's guide. Immediately after the mixture was added to cells, a real-time Fluc assay was performed. Alternatively, cells were incubated at 37 °C for 5 h, followed by flow cytometry or a HiBit assay. For eIF5B and eIF5 overexpression, 2 μ g plasmids were mixed with 4 μ L Lipofectamine 2000 (Invitrogen), and the mixture was added to cells, followed by incubation at 37 °C for at least 24 h. For flow cytometry, cells were transfected with SIINFEKL-based reporter mRNAs, and this was followed by amino acid starvation. For the HiBit assay, cells were transfected with HiBit dual-luciferase mRNA reporter, which was followed by amino acid starvation.

Firefly luciferase and HiBit assay

Cells grown in 35-mm dishes were transfected with luciferase reporter mRNAs (3 μ g) synthesized in vitro. Then, Fluc substrate D-luciferin (Regis Tech) was added to the culture medium (1 mM), which was gently mixed immediately after cell transfection. Luciferase activity was monitored and recorded using a Kronos Dio Luminometer (Atto). HiBit and Fluc activity were measured simultaneously using a Nano-Glo HiBit Dual-Luciferase Reporter System kit (Promega, N1630) with purified LgBiT Protein (Promega). Briefly, Fluc activity was measured first following addition of the ONE-Glo EX Reagent, which contains detergent to lyse cells and the firefly luciferin substrate. NanoDLR Stop

& Glo Reagent and LgBiT protein were then added to quench the Fluc signal and to provide the substrate needed to measure the activity of the HiBit–LgBiT complexes. HiBit and Fluc activities were recorded using the luminometer.

Cell treatment

Amino acid starvation was carried out by incubating cells in HBSS buffer (Lonza) with 10% dialyzed FBS (Sigma-Aldrich). Samples were collected at the indicated time points. For the protein turnover and ubiquitination assay, cells were treated with cycloheximide (CHX) at 100 μ g mL⁻¹ or with 5 μ M of MG132, followed by collection at the indicated time points.

Flow cytometry

Transfected HEK293-K^b cells were washed with PBS and collected using trypsin. Cells were pelleted at 300g for 2 min at 4 °C, followed by resuspension in the blocking buffer (1% BSA in PBS). Cells were aliquoted into a 96-well plate, followed by centrifugation at 300 \times g for 2 min. Cells were washed one more time, which was followed by staining with 25D1 antibody to Alexa Fluor 647 (1:1,000). After incubation in the dark at 4 °C for 30 min with gentle rocking, cells were washed three times with blocking buffer to remove unbound antibodies. Resuspended cells were subjected to single-cell filtering (Falcon) followed by analysis on a BD FACSAria Fusion flow cytometer (BD Biosciences). Analysis of the flow cytometry data was done using FlowJo.

Immunoblotting

Cells were washed twice in ice-cold PBS and lysed on ice in SDS–PAGE sample buffer (50 mM Tris pH 6.8, 100 mM dithiothreitol, 2% SDS, 0.1% bromophenol blue, 10% glycerol), followed by heating for 10 min at 95 °C. Proteins were separated on SDS–PAGE and transferred to PVDF membranes (Fisher). Subsequently, membranes were blocked in TBS containing 5% non-fat milk and 0.1% Tween-20 for 1 h, followed by incubation with primary antibodies overnight at 4 °C. After incubation with anti-mouse IgG horseradish peroxidase (HRP)-conjugated secondary antibody (Sigma-Aldrich, A0168, 1:10,000) or anti-rabbit IgG secondary antibody conjugated to peroxidase (Sigma-Aldrich, A9169, 1:10,000) at room temperature for 1 h, immunoblots were visualized using enhanced chemiluminescence (GE Healthcare).

shRNA knockdown

A shRNA targeting eIF5B was designed from BROAD RNAi consortium database (<https://portals.broadinstitute.org/gpp/public/>) and is listed in Supplementary Table 4. Oligonucleotides were annealed and then cloned into DECIPHER pRS19-U6-(sh)-UbiC-TagRFP-2A-Puro (Cellecta), according to the manufacturer's instructions. Lentiviral particles were packaged using Lenti-X 293T cells (Clontech), following the manufacturer's instructions. Virus-containing supernatants were collected 48 h after transfection and filtered through a 0.45- μ m filter (Millipore) using a syringe to eliminate cell contaminants. HEK293-K^b cells were infected by shRNA lentivirus for 48 h before selection by 2 μ g mL⁻¹ puromycin. Knockdown efficiency was verified by western blotting. A scrambled shRNA was used as control.

Cell viability assay

HEK293-K^b cells were preincubated in a 96-well plate for 24 h at a density of 2,000 cells per well. Cell viability was evaluated using the Cell Counting Kit-8 (CCK-8) (APExBIO), according to the manufacturer's instructions.

Polysome profiling

Four 10-cm plates containing HEK293-K^b cells, grown to 80% confluency, were refreshed with fresh medium 2 h before collection to remove the dead cells. Cells were washed with cold PBS and lysed in the polysome lysis buffer (10 mM HEPES, pH 7.4, 100 mM KCl, 5 mM MgCl₂,

100 $\mu\text{g mL}^{-1}$ cycloheximide with 2% Triton X-100). The lysates were cleared by centrifugation at 20,000g for 10 min at 4 °C. Then, 15–45 % (wt/vol) sucrose density gradients were freshly prepared in a SW41 ultracentrifuge tube (Beckman) using a Gradient Master (BioComp Instruments); 500 μL of cytosolic extracts were loaded onto sucrose gradients, followed by ultracentrifugation at 180,000g for 2.5 h at 4 °C in a SW41 rotor. Polysome profiles were recorded at A254 using the Brandel Gradient Fractionation System and an ISCO UA-6 UV/Vis detector. An aliquot of ribosome fractions representing monosome or polysome were collected, followed by digestion with *Escherichia coli* RNase I (Ambion, 750 U per 100 A260 units) by incubation at 4 °C for 1 h. Total RNA was extracted using TRIzol LS reagent (Invitrogen). Purified RNAs were used for cDNA library construction.

Ezra-seq and deep sequencing

The ribosome-protected mRNA fragments were separated on a 15% polyacrylamide TBE-urea gel (Invitrogen) and visualized using SYBR Gold (Invitrogen). Selected regions in the gel corresponding to 25–35 nt were excised. RNA fragments were dissolved by soaking overnight in 400 μL RNA elution buffer (300 mM sodium acetate, pH 5.2, 1 mM EDTA, 0.1 U μL^{-1} SUPERase_In). The gel debris was removed using a Spin-X column (Corning), followed by ethanol precipitation. Purified RNA fragments were resuspended in nuclease-free water and quantified using Qubit 2.0 Fluorometer (Invitrogen). A fraction of RNAs (10–200 ng) were used for cDNA library construction. In brief, 4 μL RNAs was mixed thoroughly with 10 U T4 PNK (NEB), 20 U SUPERase_In (Invitrogen), 5 U *E. coli* Poly(A) Polymerase (New England Biolabs), 1 μL of 1 mM ATP, and 1 μL purified home-made Ezra enzyme in 1 μL of 10 \times Ezra buffer. The mixture was incubated at 37 °C for 30 min, followed by 70 °C for 10 min. After placing the mixture on ice for 1 min, 1 μL of 1 μM 5' end adapter (Supplementary Table 4) was added, followed by ligation for 60 min at 25 °C in a 9 μL reaction mixture (1 \times T4 Rnl2 reaction buffer, 20 U SUPERase_In, 25% PEG8000 and 200 U T4 RNA ligase 2 truncated KQ (New England Biolabs)). The ligated RNA sample was mixed with pre-washed streptavidin beads (5 μL beads, washed and resuspended in 20 μL of 2 \times SSC) and incubated at room temperature for 10 min. After being subjected to magnetic bead separation for 2 min, the supernatant was removed and the beads were washed with 100 μL 2 \times SSC. The beads were re-suspended in 12 μL nuclease-free water and mixed with 8 μL cDNA synthesis mix (1 μL of 10 μM RT primer, 4 μL of 5 \times first strand buffer, 1 μL of 0.1 M DTT, 1 μL of 10 mM dNTP and 1 μL of 100 $\mu\text{g mL}^{-1}$ m-MLV-mut5), followed by incubation at 50 °C for 20 min. After the tube was placed on a magnetic stand for 2 min, the supernatant was removed and the beads were washed with 100 μL 2 \times SSC. The beads were re-suspended in 20 μL nuclease-free water and incubated at 95 °C for 2 min, and were then immediately placed on ice and kept there for at least 1 min. After the tube was placed on a magnetic stand for 2 min, the supernatant was transferred to a new tube. The synthesized cDNAs were mixed with 20 μL rRNA depletion oligonucleotide mix tagged with biotin (1 μL of 20 μM oligonucleotide mix, 4 μL 20 \times SSC buffer, and 15 μL nuclease-free water) and heated at 95 °C for 30 s, and was then slowly cooled (3 °C min^{-1}) to 25 °C. Twenty microliters of pre-washed streptavidin magnetic beads were added to each of the samples, which were kept at room temperature for 10 min. After being subjected to magnetic bead separation for 2 min, the supernatant was transferred to a new tube and precipitated with ethanol. rRNA-depleted cDNAs were amplified by PCR using barcoded sequencing primers (Supplementary Table 4). The PCR contains 4 μL of 5 \times HF buffer, 0.5 μL of 10 mM dNTP, 0.5 μL of 10 μM oligonucleotide primers and 0.25 μL of 5 U Phusion polymerase. PCR was carried out under the following conditions: one cycle at 98 °C, 30 s; 12 cycles of 98 °C (5 s), 67 °C (15 s) and 72 °C (20 s); and one cycle of 72 °C, 2 min. PCR products were separated on a 8% polyacrylamide TBE gel (Invitrogen). Expected PCR product at 180 bp was excised and recovered from DNA gel elution. After quantification using the Agilent

BioAnalyzer DNA 1000 assay, equal amounts of barcoded samples were pooled into sequencing lanes, which were sequenced using NextSeq 500 (Illumina).

Ribo-seq analysis

Sequencing reads alignment. To compile the human and mouse transcriptomes, we downloaded annotation files from the ENSEMBL database (GRCh38.81 for human, and GRCm38.83 for mouse). Protein-coding transcripts were extracted on the basis of the annotation files, using in-house scripts. For each gene, the transcript with the longest CDS was selected. In the case of equal CDS length, the longest transcript was used. rRNA sequences were downloaded from the NCBI's nucleotide database and RNACentral⁴¹. To align sequencing reads, the 3' adapters of the reads were trimmed by Cutadapt⁴², using parameters: -a AAAAAA-max-n = 0.1 -m 15. The trimmed reads with length shorter than 15 nucleotides (nt) were excluded from analysis. To keep an accurate reading frame, low-quality bases at both ends of the reads were not clipped. The clean reads were first aligned to rRNAs using Bowtie⁴³, with parameters: -v0-norc. The unaligned reads were then mapped to human or mouse transcriptome using STAR⁴⁴, with default parameters. To avoid ambiguity, reads mapped to multiple positions or with >2 mismatches were disregarded from further analyses. Reads with unaligned bases at the 5' end (that is, those flagged as soft-clip by STAR aligner) were further excluded from analyses. The ribosome P site was defined as the 12th, 13th and 14th positions from the 5' end of the read (position 0). The A site was defined as the 15th, 16th and 17th positions.

Mean footprint density along the CDS. For each mRNA, footprints at individual sites were normalized by mean footprints of the CDS. mRNAs with <16 total reads in the CDS and those in which <10% of CDS sites were covered by footprints were excluded. The normalized values of the sites with the same distance relative to the start codon or stop codon were averaged across the transcriptome.

In-frame ratio along the CDS. We first calculated the codon in-frame ratio. For each codon position within a CDS, the in-frame ratio was calculated by dividing in-frame footprint by the total footprints at the codon. The codons with less than ten footprints were excluded. Then, in-frame ratios of the codons at the same distance relative to start or stop codons were averaged across the transcriptome. A bootstrap method was used to estimate the variation of mean in-frame ratio. To this end, we repeated the process to calculate the mean in-frame ratios along CDS, using an artificial transcriptome that was generated by randomly selecting mRNAs from real transcriptome. The artificial transcriptome has the same number of mRNAs as the real transcriptome. The standard variation was calculated on the basis of the mean in-frame ratios from 100 artificial transcriptomes. To compare mean in-frame ratios between mRNA groups or samples, we instead calculated the in-frame ratio using a non-overlapping sliding window that was 30 nt in length. First, in the same way that the codon in-frame ratio was calculated, the in-frame ratio of the window was calculated. Then, mean in-frame ratios were calculated by taking the average of the in-frame ratios of the windows located at the same distance relative to start or stop codons. Because both starvation and eIF5B knockdown inhibit global translation, the sliding window was increased to 45 nt to reduce the noise of in-frame ratios.

In-frame ratio at ribosome A site or P site. The counts of total footprints were determined for the 61 codons (excluding the three stop codons), on the basis of the reads with A-site or P-site locations in the CDS. To focus our analysis on elongation, the first and last 30 codons of the CDS were excluded from analysis. For each codon, read ratios in different reading frames were calculated as the proportion of the reads with the P-site or A-site located in-frame or in frame 1 or 2, relative to

the total reads at the codon. To study the relationship between codon optimality and reading fidelity, the 61 codons were divided into optimal and non-optimal codons, on the basis of tAI⁴⁵. Codons with tAI < 0.3 were defined as non-optimal codons, and the others were considered optimal codons.

Codon occupancy at ribosome A-site or P-site. For each mRNA, footprints at the same codons within the CDS were counted, and were then normalized by the average footprint of the CDS. mRNAs with <32 footprints in the CDS were not included. Footprints with the P-site or A-site within the first or last 30 codons were also excluded. The normalized ribosome occupancies at the same codons were averaged over the transcriptome.

Differentially expressed genes upon eIF5B knockdown. Ribosome footprints mapped to the CDS of individual mRNAs were counted using an in-house script. mRNAs with less than ten total reads in all samples were excluded. The count table was then analyzed using DESeq2 (ref. 46). mRNAs with FDR < 0.05 and fold change > 2 were defined as significantly upregulated. mRNAs with FDR < 0.05 and fold change < 0.5 were considered significantly downregulated. The significantly changed mRNAs were then analyzed using PANTHER⁴⁷ to search enriched biological processes and pathways.

Calculation of initiation potential. We used ribosome footprints in the 5' UTR to calculate the initiation potential of a triplet when a scanning subunit reached the triplet. If a triplet can serve as a start codon, there should be more reads in the downstream coding frame relative to the triplet than in the same frame before the triplet. Under this assumption, for all triplets in the 5' UTR, we counted IFR values in a small window (30 nt, or to a stop codon if the ORF was shorter than 30 nt) after the triplet, which was then subtracted from the IFR value in the window (30 nt) before the triplet. Triplets with fewer than five total reads in any of the two windows were excluded. Considering that AUG is usually a strong initiation codon, we also excluded triplets containing AUG in any of the two windows. Although multiple alternative initiation sites in the 5' UTR may overlap, thereby affecting the IFR difference between the two windows, averaging the IFR difference values for the same triplets in the 5' UTR can eliminate the effect of overlapping hidden ORFs and other random noise, thus providing an overview of the initiation potential of a triplet. For a triplet, we defined the average IFR difference as the initiation potential of the triplet.

uORF prediction. We predicted uORFs with robust translation on the basis of Ezra-seq data. For each mRNA, we first extracted all possible uORFs, that is, mRNA regions starting with ATG or one of the other ten non-canonical initiators (as indicated by a higher initiation potential) and ending with TAG, TGA or TAA. We used the method from ref. 48 to find the potential uORFs. In brief, we applied a Wilcoxon test to test whether there are significantly more in-frame reads than in the other two frames. The two *P* values were then combined to a single *P* value using a Stouffer's method. The *P* values were then corrected using the FDR. The uORFs with FDR < 0.1 were defined as having robust translation. Overlapping uORFs located in the same coding frame were merged into a single one by selecting the stop codon located at the most 5' end of transcript, and the start codon that maximizes the fraction of in-frame reads.

mRNA sequence analysis

Definition of mRNAs enriched with non-optimal codons. To compare in-frame ratio between mRNAs groups with different fractions of non-optimal codons in the beginning of the CDS, we calculated a geometric mean of the tAI value for the first 100 codons. mRNAs were categorized into four groups on the basis of the quartiles (25%, 50% and 75%) of the mean tAI values.

Calculation of Kozak score. We calculated a Kozak score for an mRNA on the basis of a position weight matrix. In brief, we downloaded the initiation strength from previous work⁴⁹, which evaluated the initiation efficiency by placing random sequences around the AUG codon (that is, NNNNNNAUGNN). The 100 random sequences with highest initiation efficiencies were used to construct a position probability matrix. The position weight matrix (PWM) was calculated using following equation:

$$\text{PWM} = \log_2 \frac{p_{ij}}{b_i}$$

Where, $p_{i,j}$ refers to the possibility of *i*th nucleotide at position *j*. *i* is either A, T, G or C. b_i is the background frequency of the *i*th nucleotide. For simplicity, we assumed equal frequencies, that is each nucleotide had a frequency of one-fourth.

For each mRNA, we extracted six nucleotides before the start codon, and two nucleotides after the start codon. A Kozak score was defined as the sum of the values of corresponding nucleotides at each position in the PWM. mRNAs were categorized into four groups on the basis of the quartiles (25%, 50% and 75%) of the Kozak score.

RNA secondary structure analysis. First, a 50-nt sequence with the start codon in the center was extracted. The minimum fold free energy (MFE) was calculated using ViennaRNA⁵⁰ with default parameters. The mRNAs with lowest MFEs (bottom 25%) were considered start codon-structured mRNAs, and those with highest MFEs (top 20%) were defined as non-structured mRNA.

Permutation test. We performed a permutation test to calculate significance of the IFR between two native groups, under the null hypothesis that the difference between the two native groups is equal to the background difference. To this end, mRNAs were randomly selected from the transcriptome, and the IFR of the first 70 codons was calculated. The number of randomly selected mRNAs matched the number of mRNAs in the respective native groups. The IFR difference was calculated between the two groups of randomly selected mRNAs. This process was repeated 1,000 times. A *P* value was calculated by dividing the total number of permutation tests, 1,001 (adding a pseudo-number 1), by the number of permutation tests, adding a pseudo-number 1, with an IFR difference higher than that between the two native groups.

Analysis of massively parallel reporter assay

Count of random sequences. For each sequencing reads, from the raw sequencing file, the first 21 nt at the 5' end, the 3' adapter and low-quality bases were trimmed using Cutadapt. The trimmed reads with a length that was unequal to 10 nt were excluded from analysis. The remaining trimmed reads were counted, and an RPM value (reads per million) was obtained by dividing the resultant read count by the total count.

Triplet frequency along random sequences. We separated mRNA reporters into different groups on the basis of the number of associated ribosomes. For mRNA reporters with more than five sequencing reads, a M/P ratio was calculated as the ratio of the RPM value in the monosome fraction over the RPM value in the polysome fraction. The mRNA reporters with the highest M/P ratio (top 10% of the total mRNA reporters) were defined as monosome-enriched mRNA reporters. For mRNA reporters sorted by 25D1 and GFP signals, the 25D1 or GFP signal was calculated using the method described in Jia et al.²¹. The mRNA reporters with the highest ratio of 25D1 signal over GFP intensity (top 10% of the total mRNA reporters) were defined as 25D1-enriched mRNA reporters. We counted triplet frequencies along random sequences inserted in monosome or 25D1-enriched mRNA reporters, respectively.

Analysis of variance. To understand the relative contribution of position-specific nucleotide to frameshifting rate, we conducted a multiway analysis of variance (ANOVA)^{51,52} using the following equation:

$$E(Y_{ij}) = \mu + \sum P_{ij}$$

where Y_{ij} refers to the log-transformed ratio of RPM in the monosome fraction to RPM in the polysome fraction ($\log_2(M/P)$). μ is the overall average effect for all levels. P_{ij} represents the additional effect of the i th nucleotide at position j . i is either A, T, G or C. The combination effects among positions were not considered. We performed the ANOVA described above using the 'aov' function in R. The sum-of-squares for each position was directly retrieved from the ANOVA result, accessed using the 'summary' function in R. The relative contribution of each position to the frameshifting rate was defined as the fraction of sum-of-squares for the position divided by the total sum-of-squares.

Proteomic analysis

We downloaded high-quality human unidentified spectra from PRIDE cluster database²². The database includes all proteins encoded by human mRNAs, and the predicted peptide sequences generated by out-of-frame translation initiated from the start codon. For each out-of-frame peptide, an artificial methionine was added at the N terminus. We also added contaminants obtained from MaxQuant⁵³. The searches were performed by X!Tandem⁵⁴ with default parameters. The trypsin was set as the enzyme to digest peptides. The output of X!Tandem was further analyzed by PeptideProphet following the Trans-Proteomic Pipeline tutorial⁵⁵. TMT-MS spectra were downloaded from An et al.²⁵. The spectra were analyzed by MaxQuant. For group-specific parameters, the spectra type was set as 11plexTMT for MS3. Oxidation of methionine and N-terminal acetylation were set as variable modifications. Cysteine carbamidomethylation was set as a fixed modification. The digestion enzyme was set as Trypsin/P and LysC. The MS dataset encompassing N-terminal peptides enriched using terminal amine isotopic labeling of substrates was downloaded from Na et al.²⁶. Acetylation and demethylation of peptide N termini and oxidation of methionine were set as variable modifications. Cysteine carbamidomethyl modification and dimethyl modification of lysine were set as fixed modifications. The digestion enzyme was set as ArgC. For PRIDE unidentified spectra, all identified peptides were used for metagene analysis. For the single MS dataset, the peptides with reported FDR < 0.01 and posterior error probability < 0.05 were used. The peptides from all three datasets were exported to Microsoft Excel for further analysis.

Statistics and reproducibility

For Ezra-seq libraries, two biologically independent replicates were used in most experiments. Immunoblots and flow cytometry plots are representative of at least three independent experiments. Data are presented as the mean \pm s.e.m., with two-tailed student's t -tests or a Wilcoxon signed-rank test used to assess statistical significance of differences between groups. All P values are given in the figure legends. All statistical analyses and data graphing were done using R (4.03) software.

Figure preparation

Figures were prepared using GraphPad Prism, R, ggplot2 and CorelDraw.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All sequencing data have been deposited in the National Center for Biotechnology Information Gene Expression Omnibus (GEO) and

are accessible through the GEO Series accession number [GSE184825](https://doi.org/10.1038/s41594-023-01119-z). Source data are provided with this paper.

Code availability

All Perl Scripts used in this study have been deposited in GitHub: https://github.com/QianLab-Cornell/Count_Ribo_Reads.

References

- Consortium, T. R. RNAcentral: a hub of information for non-coding RNA sequences. *Nucleic Acids Res.* **47**, D221–D229 (2019).
- Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12 (2011).
- Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, 1–10 (2009).
- Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
- Reis, M. D., Savva, R. & Wernisch, L. Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res.* **32**, 5036–5044 (2004).
- Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).
- Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P. D. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* **47**, D419–D426 (2019).
- Xiao, Z. et al. De novo annotation and characterization of the translome with ribosome profiling data. *Nucleic Acids Res.* **46**, e61 (2018).
- Noderer, W. L. et al. Quantitative analysis of mammalian translation initiation sites by FACS-seq. *Mol. Syst. Biol.* **10**, 748 (2014).
- Lorenz, R. et al. ViennaRNA package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011).
- Cambay, G., Guimaraes, J. C. & Arkin, A. P. Evaluation of 244,000 synthetic sequences reveals design principles to optimize translation in *Escherichia coli*. *Nat. Biotechnol.* **36**, 1005–1015 (2018).
- Mutalik, V. K. et al. Quantitative estimation of activity and quality for collections of functional genetic elements. *Nat. Methods* **10**, 347 (2013).
- Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **11**, 2301 (2016).
- Vaudel, M., Barsnes, H., Berven, F. S., Sickmann, A. & Martens, L. SearchGUI: an open-source graphical user interface for simultaneous OMSSA and X! Tandem searches. *Proteomics* **11**, 996–999 (2011).
- Deutsch, E. W. et al. A guided tour of the Trans-Proteomic Pipeline. *Proteomics* **10**, 1150–1159 (2010).

Acknowledgements

We thank J. W. Yewdell (NIAID, NIH) for providing 25D1 reagents and HEK293-K^b cells. We thank the Cornell University Life Sciences Core Laboratory Center for sequencing and FACS support. This work was supported by US National Institutes of Health (R01GM1222814 and DP1GM142101) and HHMI Faculty Scholar (55108556) to S.-B.Q.

Author contributions

Y.M. and S.-B.Q. conceived the study and designed the experiments. Y.M. conducted the majority of data analysis, and L.J. performed the majority of experiments. L.D. contributed to the Ezra-seq development. X.E.S. helped with starvation Ribo-seq. S.-B.Q. and Y.M. wrote the manuscript. All authors discussed the results and edited the manuscript.

Competing interests

L.D., X.E.S. and S.-B.Q. are inventors of Ezra-seq technology (9217-02 PCT). X.E.S. and S.-B.Q. are co-founders of EzraBio Inc. The remaining authors declare no competing interests.

Additional information

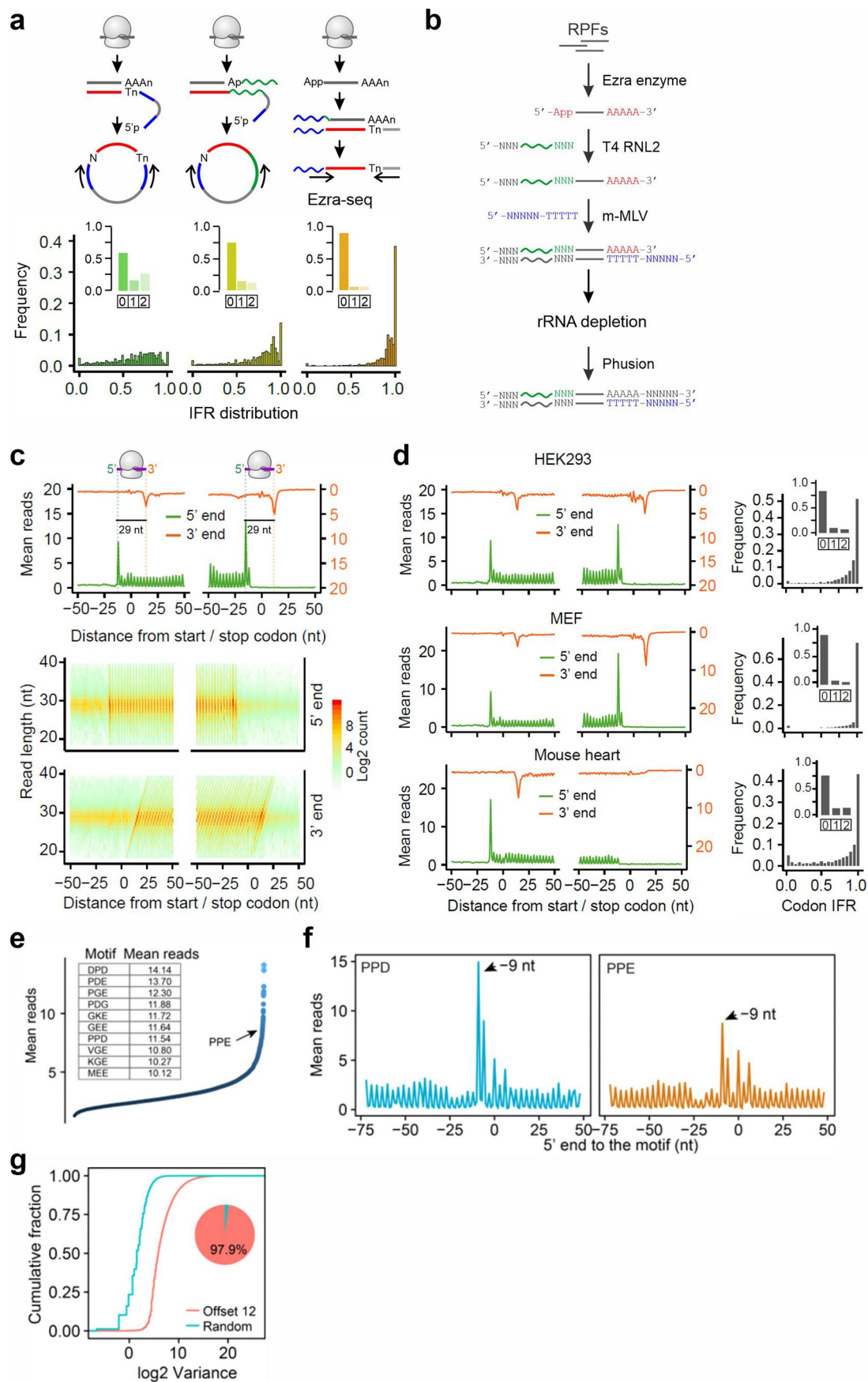
Extended data is available for this paper at <https://doi.org/10.1038/s41594-023-01119-z>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41594-023-01119-z>.

Correspondence and requests for materials should be addressed to Shu-Bing Qian.

Peer review information *Nature Structural & Molecular Biology* thanks Jonathan Schatz and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Sara Osman was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.



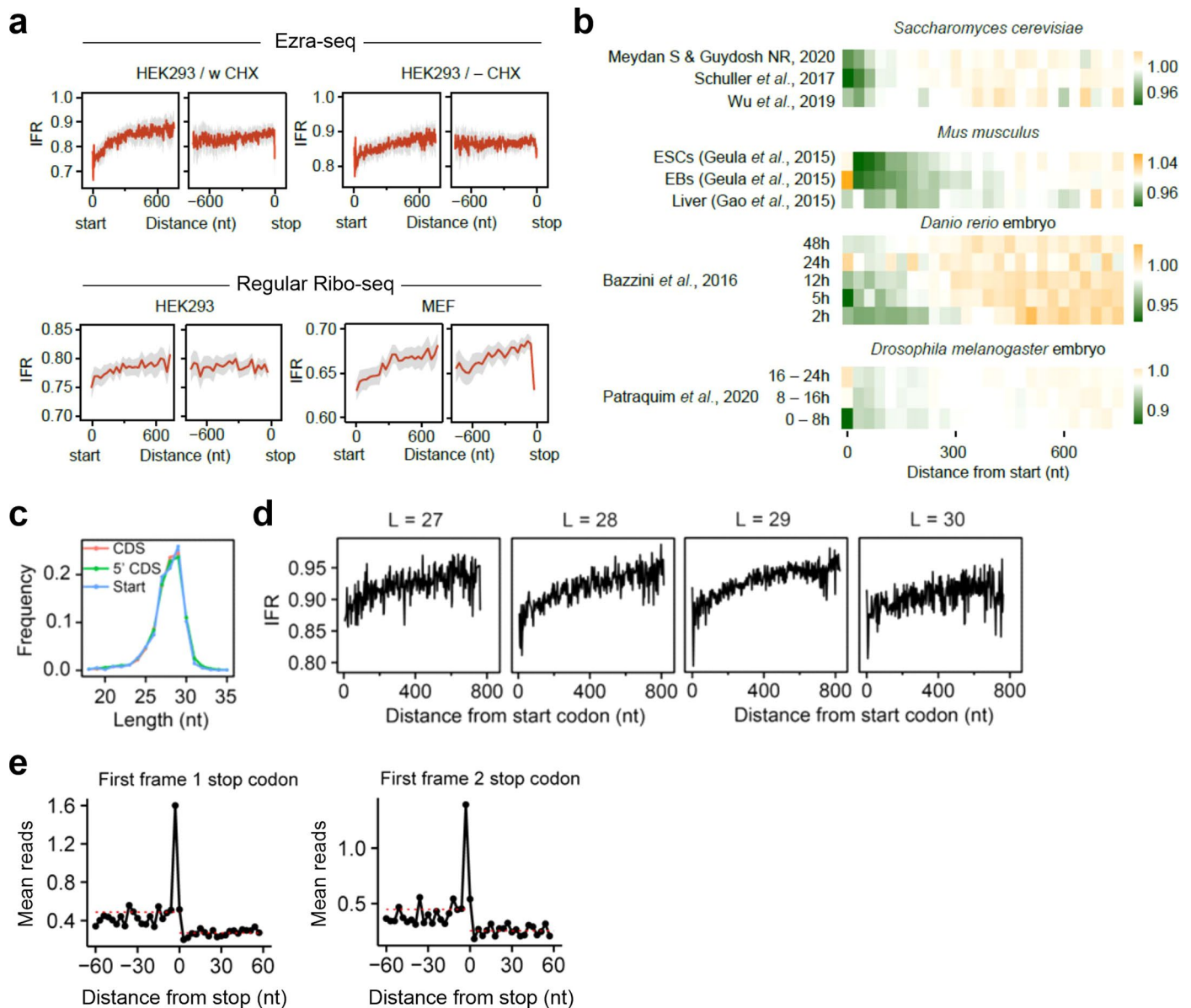
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Development of Ezra-seq. (a) The top panels show the comparison of Ezra-seq and two ligation-based Ribo-seq procedures. The bottom panels show in-frame ratio (IFR) of reads at codons with 1.0 indicating complete in-frame. The inserted bar plots show the fraction of reads in different reading frames. (b) An outline of Ezra-seq procedure (see Materials and Methods for the detail). (c) Aggregation plots show the ribosome density across the transcriptome. Transcripts are aligned to start and stop codons, respectively. Both 5' end (green) and 3' end (orange) of footprints are used for plotting. The heatmap shows the density of ribosome footprints with different length. (d) Aggregation plots show the ribosome density across the transcriptome in different cell lines and tissues. Transcripts are aligned to start and stop codons,

respectively. Both 5' end (green) and 3' end (orange) of footprints are used for plotting. The right panel shows the range of codon IFR with 1.0 indicating complete in-frame reads. The inserted bar plot shows the fraction of reads in different reading frames. (e) Mean ribosome reads on individual triple-motifs. The top 10 motifs with the highest occupancy were listed. (f) Aggregation plots show ribosome densities around the motif PPD (left) and PPE (right). The 5' end of the reads was used. (g) Variance analysis of ribosome density at individual P-sites when different offset values were considered. The pie chart shows the fraction of P-sites that have variation of ribosome density (12-nt offset) higher (red) or lower (blue) than the variation of ribosome density with random offset.

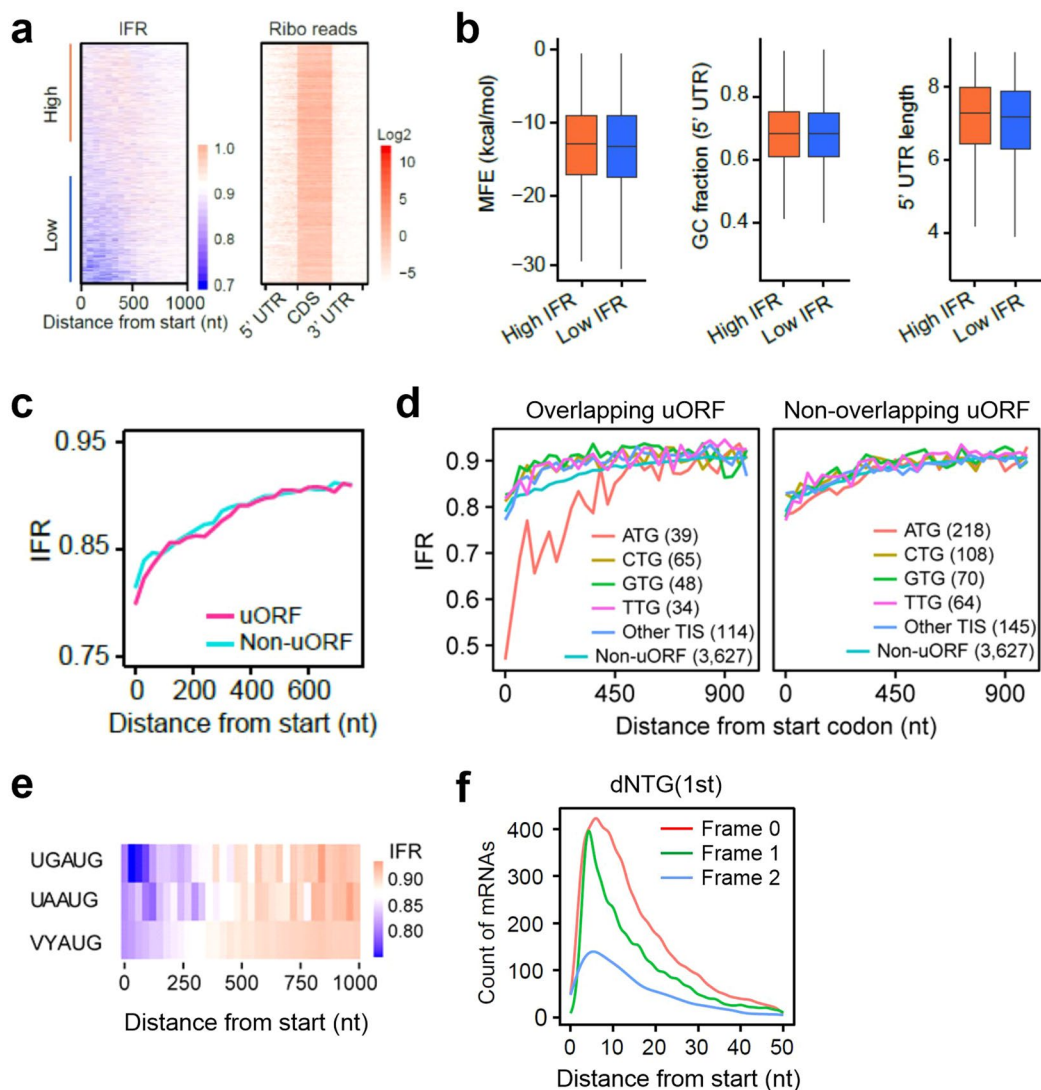
Extended Data Fig. 2 | Comparison of 5' end accuracy across representative Ribo-seq data sets. (a)-(h) For each panel, the left bar plots show the number of mismatches at the 5' end of footprints. The heatmap shows the distance of 5' end to the start codon. All footprints were classified into different length (y-axis), the color represents the \log_2 count of footprint. For each length group, the submit peak was indicated by a star, which indicates the distance of P-site to the start codon (the number at the right side of heatmap). The number in parentheses is

in-frame rate when the left P-site offset was used. **(i)** Schematic of IFR changes after a triplet with high or low initiation potential (left panel). A scatter plot shows initiation potential of 64 triplets based on IFR changes (right panel). **(j)** Correlation of initiation potential of 64 triplets between biological replicates (left), between HEK293 and MEF cell lines (middle), or between MEF cell lines and mouse heart (right). The AUG and 10 non-AUG triplets with highest initiation potential in HEK293 were highlighted in red.



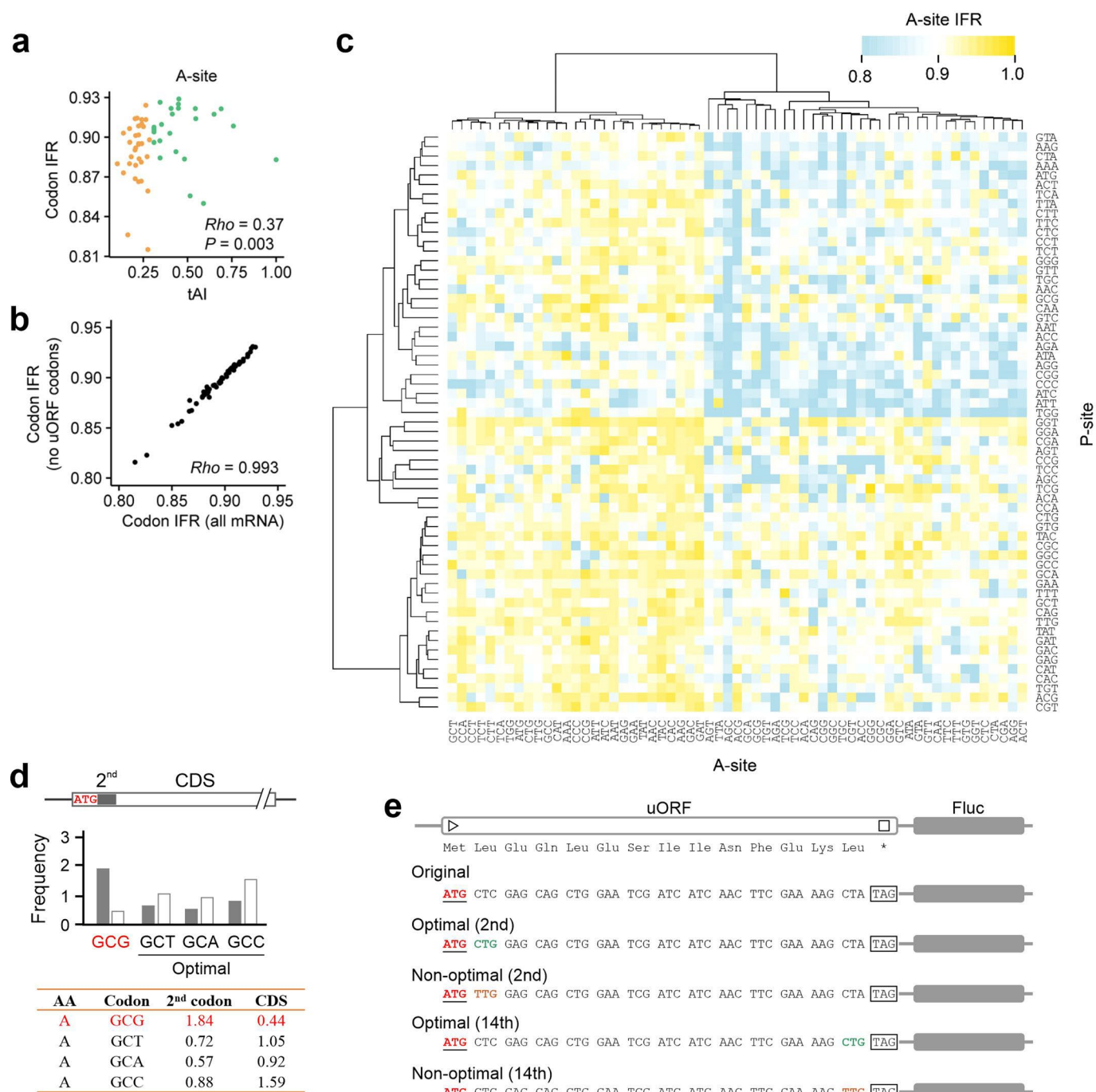
Extended Data Fig. 3 | Characterizing in-frame and out-of-frame RPFs. (a) Top panel shows in-frame ratio of ribosome footprints across the transcriptome in cells with (left panel) or without (right panel) cycloheximide (CHX) treatment (100 $\mu\text{g}/\text{mL}$) for 30 min. Bottom panel shows in-frame ratio of ribosome footprints across the transcriptome in HEK293 and MEF cells. Ribo-seq data were obtained by ligation-based Ribo-seq methods. Due to the relatively low resolution, IFR values were calculated within a non-overlapping sliding window (30 nt). Grey shadow shows the variation of mean IFR estimated by bootstrap method. Transcripts are aligned to start and stop codons, respectively. **(b)** Heat maps show normalized IFR of CDS across different species and cell lines

using published Ribo-seq data sets. IFR values were calculated within a non-overlapping sliding window (30 nt), which was subsequently normalized by CDS IFR. **(c)** The read length distribution for footprints at the start codon (Start), in the 5' end of CDS (the first 60 nt, 5' CDS), or in the CDS region. **(d)** In-frame ratio was calculated based on footprints with different lengths (from 27 nt to 30 nt), which represent the major groups of footprints. **(e)** An aggregation plot shows out-of-frame reads around the 1st frame 1 or frame 2 stop codons. The mean out-of-frame reads before and after the out-of-frame stop codons are indicated by dashed lines.



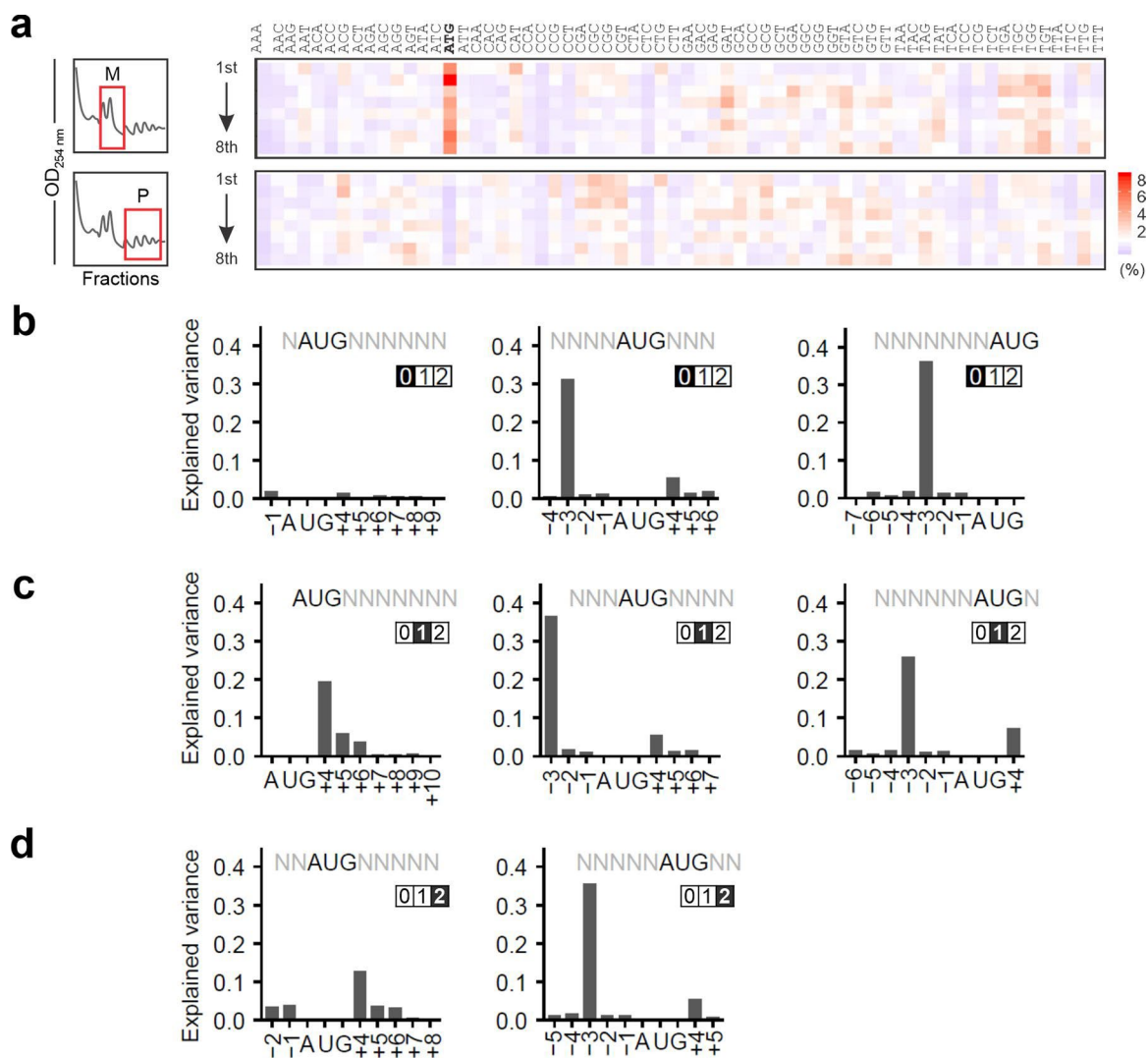
Extended Data Fig. 4 | uORF and leaky scanning minimally contribute to reduced IFR. (a) A heat map shows the IFR values at the first 333 codons of individual mRNAs. The right heat map shows ribosome densities in different regions of individual mRNAs. (b) Boxplots show RNA fold free energy (MFE) around start codon, GC fraction in 5' UTR and 5' UTR length between mRNAs with low or high IFR in the beginning of CDS. High and low IFR groups refer to mRNAs with top and bottom 15% of IFR values ($n = 889$ for each group) respectively. The median of MFE, GC and 5' UTR length in each group is indicated by a center line, the box shows the upper and lower quantiles, the whisker shows the 1.5 \times interquartile range. The outliers are not shown. (c) Comparison of IFR between mRNAs with or without uORF. IFR values are calculated within a non-overlapping

sliding window (30 nt). (d) Effects on uORF translation on in-frame ratio in the beginning of CDS. uORFs were identified by Ezra-seq data in this study. All uORFs were separated into different groups based on the initiators. Overlapping uORFs are defined if the stop codon of uORF is beyond the start codon of main CDS. The numbers indicate the number of mRNAs used for analysis. Of note, when uORFs strongly inhibit main CDS translation, those mRNAs were not included in analysis due to the lack of sufficient reads on the main CDS. (e) A heatmap shows IFR values between mRNAs with or without a stop codon UGA before start codon. V: not uracil, Y: pyrimidine. (f) Distance of the first downstream NTG (dNTG) in different reading frames relative to the annotated start codon.

**Extended Data Fig. 5 | Non-optimal codons induce ribosome frameshifting.**

(a) A scatter plot shows the correlation between codon optimality and IFR when the A-site codon is considered. Spearman's Rho correlation between tAI values and codon IFR, as well as the P value, was indicated. (b) A scatter plot shows the correlation of IFR values for (b) transcripts with or without uORFs. (c) A heatmap shows the effect of P-site and A-site combinations on reading frame fidelity

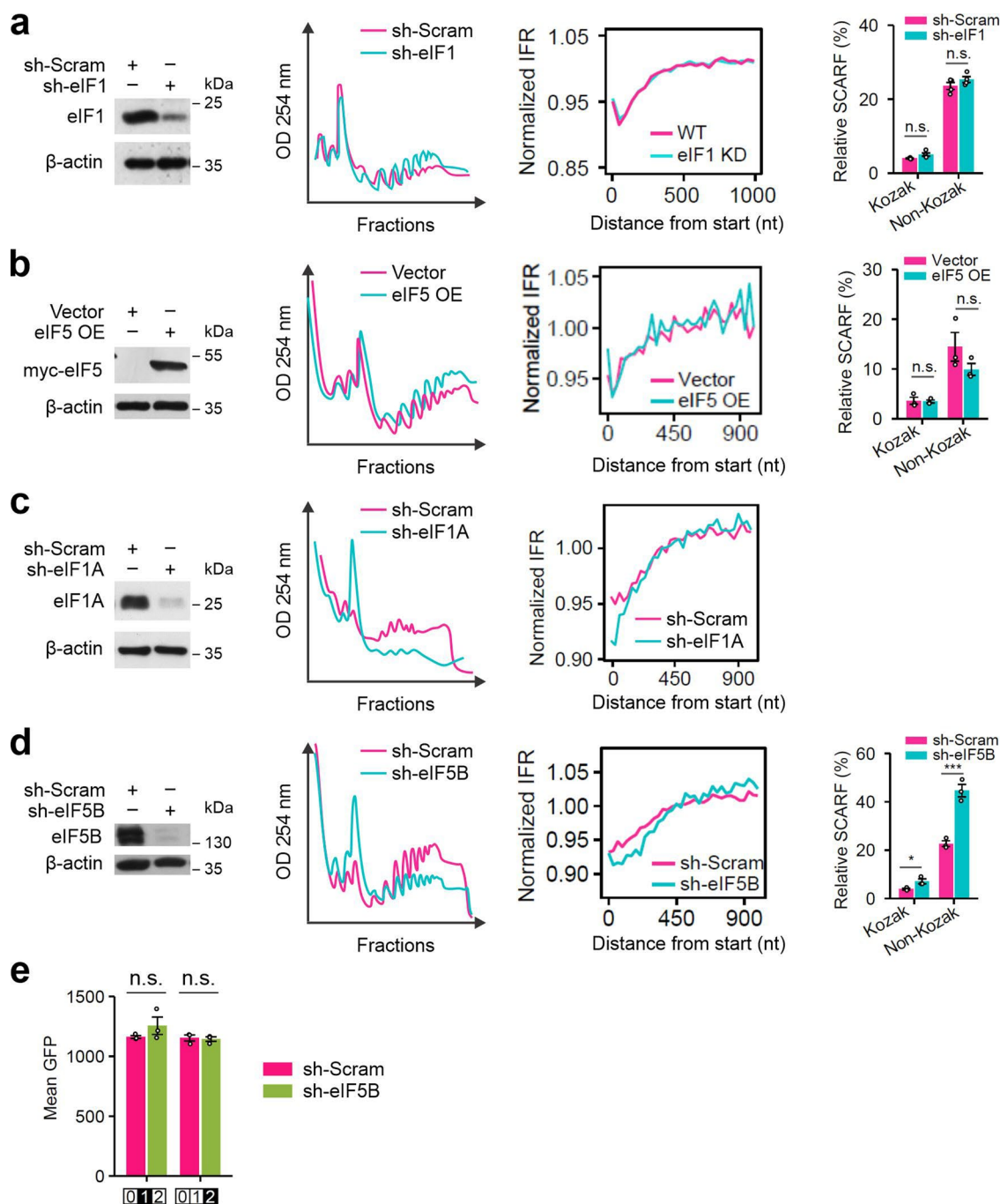
at ribosome A-sites. (d) Analysis of codon usage bias of the first codon after the start codon. Bar plot and the table show the relative synonymous codon usage (RSCU) of the most prevalence amino acid Alanine at the second codon of CDS. (e) Sequences of uORF reporter with either the 2nd codon or 14th codon replaced by a synonymous optimal (green) or non-optimal (orange).



Extended Data Fig. 6 | SCARF regulation by the start codon sequence context.

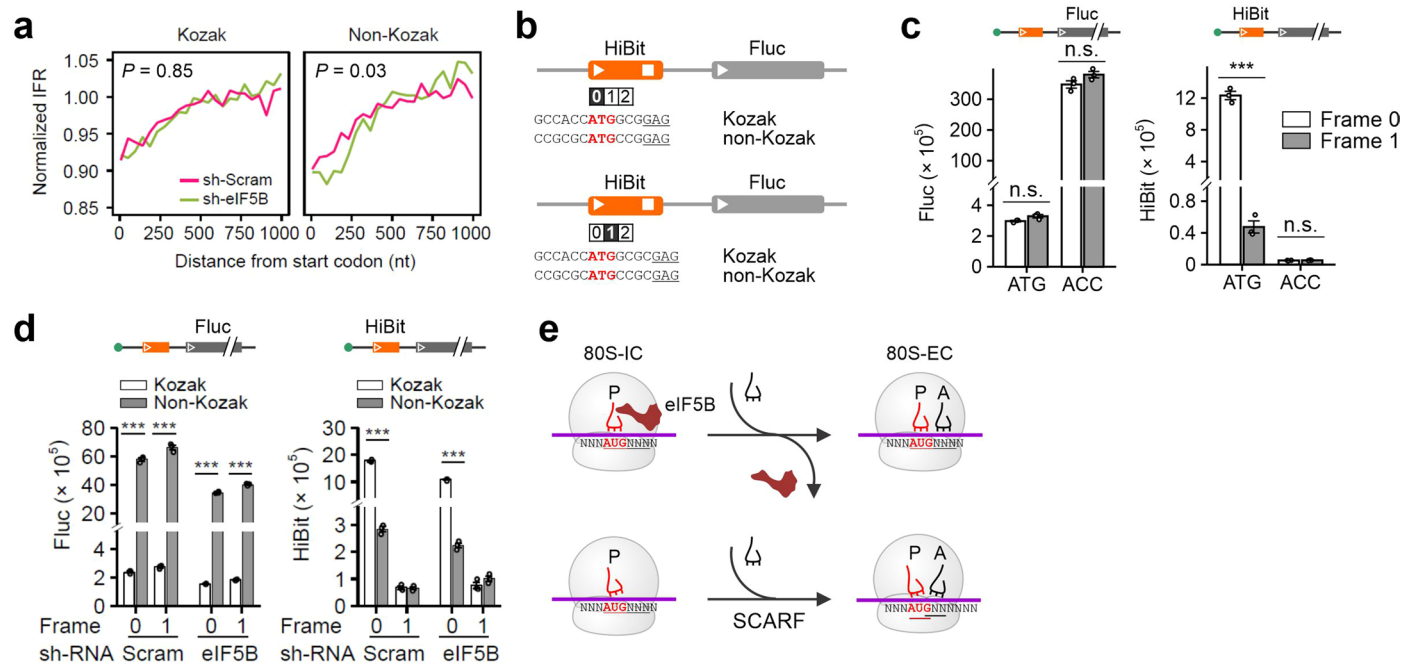
(a) HEK293-K⁹ cells were transfected with massively parallel mRNA reporters followed by sucrose gradient separation into monosome (M) and polysome (P) fractions. The original frequency of triplets in different populations is shown as

heat maps. (b–d) Relative contributions of the nucleotide identity in different positions to the uORF translation based on the M/P ratio. The highlighted numbers refer to the reading frame of the encoded SIINFELK relative to the AUG codon.



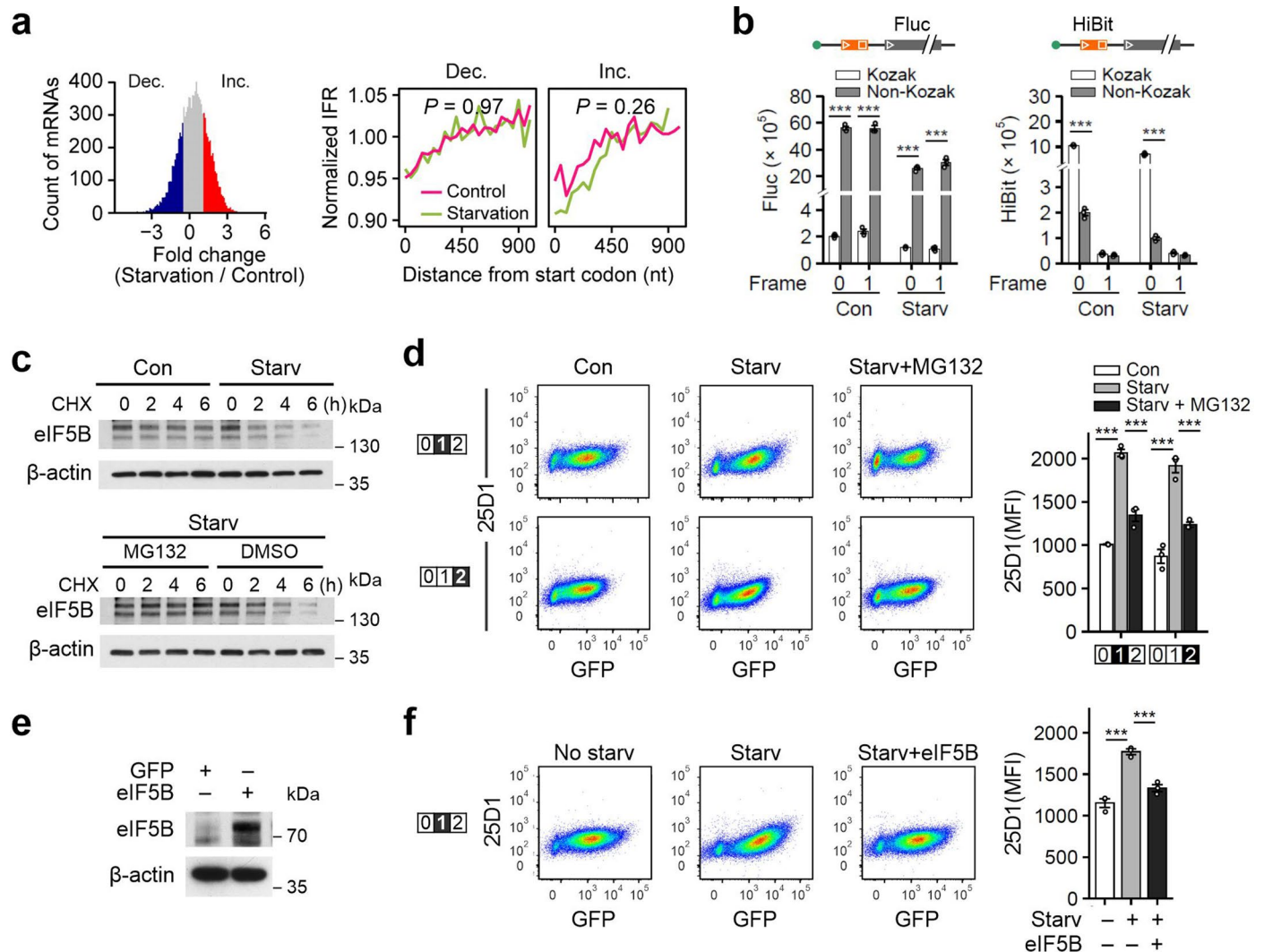
Extended Data Fig. 7 | Characterizing the regulatory role of translation initiation factors in SCARF. (a) The left panel shows the western blots of HEK293 cells with or without eIF1 knockdown. The middle panel shows the comparison of polysome profiles of cells with or without eIF1 knockdown. The right panel shows the comparison of normalized IFR in cells with or without eIF1 knockdown. IFR values are calculated within a non-overlapping sliding window (45 nt), which was subsequently normalized by CDS IFR. The right panel shows the HiBit-based SCARF reporter assay. Error bars, mean \pm s.e.m.; Two-tailed *t*-test, $n = 3$, n.s.

no significant change. (b) Same as (a) using cells with eIF5 overexpression. (c) Same as (a) using cells with eIF1A knockdown. (d) Same as (a) using cells with eIF5B knockdown. Error bars, mean \pm s.e.m.; two-tailed *t*-test, $n = 3$, * $P = 0.04$ for the comparison of Kozak context (left bars), *** $P = 0.001$ for the comparison of Non-Kozak context (right bars). (e) Bar plots show the relative GFP mean fluorescence intensity (MFI) of SCARF reporters over the in-frame control. Error bars, mean \pm s.e.m.; two-tailed *t*-test, $n = 3$, n.s. no significant change.



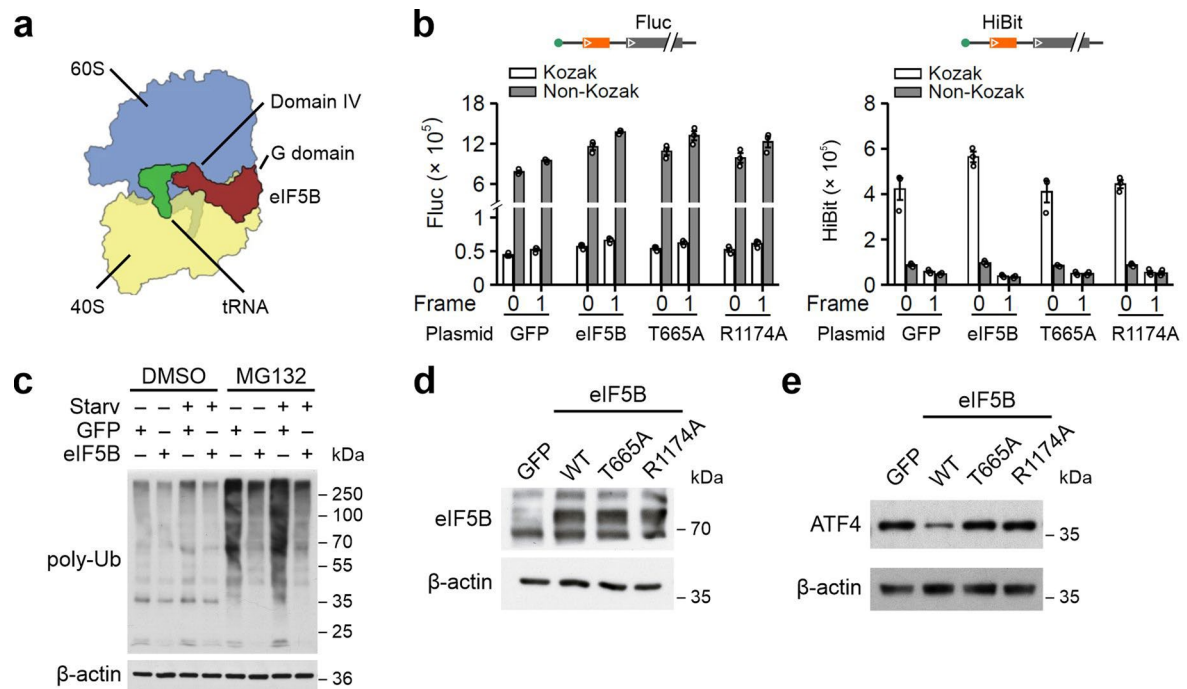
Extended Data Fig. 8 | Characterizing the regulatory role of translation initiation factors in SCARF. (a) Comparison of normalized IFR for mRNAs with or without the Kozak sequence context on start codons in cells with or without eIF5B knockdown. IFR values are calculated within a non-overlapping sliding window (45 nt), which was subsequently normalized by CDS IFR. P value was calculated by a permutation test. (b) Sequence information of HiBit-based SCARF reporter. (c) Bar graphs show the HiBit-based SCARF reporter assays in HEK293 cells. Both Fluc (left panel) and HiBit (right panel) signals were measured from cells transfected with SCARF reporters bearing a uORF start codon or a uORF with start codon mutated to ACC. Error bars, mean \pm s.e.m.; two-tailed t -test, $n = 3$, *** $P = 2.3 \times 10^{-5}$. n.s. no significant change. (d) Bar graphs show the HiBit-based

SCARF reporter assays in cells with or without eIF5B knockdown. Both the Fluc (left panel) and HiBit (right panel) signals were measured from cells transfected with SCARF reporters bearing a uORF start codon with or without the Kozak sequence context. Error bars, mean \pm s.e.m.; two-tailed t -test, $n = 3$, *** $P = 1.2 \times 10^{-6}$ and $P = 3.1 \times 10^{-6}$ for comparisons of Fluc activities in frame 0 and frame 1 of scramble cells (right panels); $P = 8.6 \times 10^{-8}$ and $P = 1.1 \times 10^{-7}$ for comparisons of Fluc activities in frame 0 and frame 1 of eIF5B knockdown cells (right panels). $P = 9.5 \times 10^{-7}$ and $P = 3.0 \times 10^{-6}$ for comparisons of HiBit activities in frame 0 of scramble and eIF5B knockdown cells (right panels). (e) A model depicting the role of eIF5B in stabilizing initiator tRNA at the P-site, thereby maintaining the reading frame during the transition from initiation to elongation.

**Extended Data Fig. 9 | Nutrient starvation induces SCARF via eIF5B**

degradation. (a) Histogram showing the changes of ribosome density upon amino acid starvation. The mRNAs with top 25% increase (Inc) and decrease (Dec) are colored coded in red and blue, respectively. The right panels show the CDS IFR of both mRNA groups. P value was calculated by a permutation test. (b) Bar graphs show SCARF reporter assays in cells before and after amino acid starvation. Error bars, mean \pm s.e.m.; two-tailed t -test, $n = 3$, $***P = 2.9 \times 10^{-6}$ and $P = 1.0 \times 10^{-5}$ for comparison of Fluc levels in frame 0 or frame 1 of control cells; $P = 2.3 \times 10^{-5}$ and $P = 1.5 \times 10^{-4}$ for Fluc levels in cells with starvation. $P = 5.6 \times 10^{-6}$ and $P = 6.5 \times 10^{-3}$ for comparison of HiBit activities in frame 0 of cells with and without starvation. (c) The top panel shows the turnover of eIF5B in HEK293 cells before and after amino acid starvation. The bottom panel shows the turnover of eIF5B in starved HEK293 cells in the presence of 5 μ M MG132. (d) Representative

flow cytometry of HEK293-K^b cells transfected with SCARF reporters before and after amino acid starvation, in the absence or presence of 5 μ M MG132. Bar plots show the relative 25D1 mean fluorescence intensity (MFI) of SCARF reporters. Error bars, mean \pm s.e.m.; two-tailed t -test, $n = 3$, $***P = 1.2 \times 10^{-5}$ and $P = 8.6 \times 10^{-4}$ for comparisons of 25D intensities in Frame 1; $P = 7.5 \times 10^{-4}$ and $P = 1.3 \times 10^{-3}$ for comparisons of 25D intensities in Frame 2. (e) Western blots of exogenous eIF5B in transfected HEK293 cells. (f) Representative flow cytometry of HEK293-K^b cells transfected with SCARF reporters before and after amino acid starvation, in the absence or presence of exogenous eIF5B. Bar plots show the relative 25D1 mean fluorescence intensity (MFI) of SCARF reporters. Error bars, mean \pm s.e.m.; two-tailed t -test, $n = 3$, $***P = 5.8 \times 10^{-4}$ for comparison of control vs starvation; $P = 1.2 \times 10^{-3}$ for comparison of exogenous eIF5B vs vector under starvation.



Extended Data Fig. 10 | The rescuing effect of eIF5B in nutrient starvation induced SCARF. (a) A schematic image of 80S-eIF5B complex adopted from Wang *et al.*, Nat Commun 2020. (b) Bar graphs show the HiBit-based SCARF reporter assays in cells transfected with wild type eIF5B or mutants. Both the Fluc (left panel) and HiBit (middle panel) signals were measured from cells transfected with SCARF reporters bearing a uORF start codon with or without the Kozak sequence context. Error bars, mean \pm s.e.m.; $n = 3$. (c) Representative western blots of polyubiquitinated species in HEK293 cells with or without

eIF5B overexpression before and after amino acid starvation. The experiment was independently repeated three times with similar results. (d) Representative western blots of exogenous eIF5B in transfected HEK293 cells. The experiment was independently repeated three times with similar results. (e) Representative western blots of ATF4 in starved HEK293 cells transfected with eIF5B wild type or mutants. The experiment was independently repeated three times with similar results.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection High through-put sequencing data was collected by illumina HiSeq Control Software v2.2.58 for HiSeq2500 System.

Data analysis Cutadapt v1.18 was used to filter low quality reads. Bowtie v1.1.2 and STAR v2.7.0 was used to align reads to human and mouse transcriptome. MaxQuant v2.0.3.0 and The Trans-Proteomic Pipeline V5.2 were used to analyze proteomics data. R v3.5.1 was used to perform all statistical analysis. ggplot2 for R was used to make all statistical figures. All Perl Scripts used in this study have been deposited in GitHub: https://github.com/QianLab-Cornell/Count_Ribo_Reads.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All sequencing data have been deposited in the National Center for Biotechnology Information Gene Expression Omnibus (GEO) and are accessible through the GEO Series accession number GSE184825. Source data are provided with this paper.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	N/A
Data exclusions	No data exclusion
Replication	Experiments in this study were reproduced at least three times. Replication were described in figure legends.
Randomization	N/A
Blinding	Investigators were blinded to group allocation during data collection and analysis.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

Methods

- | n/a | Involved in the study |
|-------------------------------------|---|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Antibodies |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Human research participants |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern |

- | n/a | Involved in the study |
|-------------------------------------|--|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |

Antibodies

Antibodies used

25D1 Alexa 647 antibody (gift from laboratory of Jonathan W. Yewdell), Mouse monoclonal anti- β -actin (Sigma-Aldrich, A5441, clone AC-15), Rabbit polyclonal anti-eIF5B (Proteintech, 13527-1-AP), Rabbit polyclonal anti-eIF4A1 (Abcam, ab31217), Rabbit polyclonal anti-eIF4E (Cell Signaling, 9742S), anti-Phospho-eIF2 α (Ser51) (Cell Signaling, 3398S, clone D9G8), Rabbit monoclonal anti-eIF1A (Abcam, ab177939, clone EPR12466(B)), Rabbit polyclonal anti-eIF1 (Proteintech, 15276-1-AP), Rabbit polyclonal anti-eIF5 (Santa Cruz Biotechnology, sc-282), Mouse monoclonal anti-ATF4 (Santa Cruz Biotechnology, sc-390063, clone B-3), Mouse monoclonal anti-Ubiquitin (Santa Cruz Biotechnology, sc-8017, clone P4D1). Anti-mouse IgG (Fc specific) horseradish peroxidase (HRP)-conjugated secondary antibody (Sigma-Aldrich, A0168). Anti-rabbit IgG (whole molecule) secondary antibody conjugated to peroxidase (Sigma-Aldrich, A9169).

Validation

25D1 Alexa 647 antibody was verified by the authors in the previous work (Dersh D et al. Methods Mol Biol. 2019;1988:109-122.PMID: 31147936).

Other antibodies are from commercially available sources and have been validated by the manufacturers with supporting data and publications found on the manufacturers' websites. See below for a summary:

anti- β -actin (Sigma-Aldrich, A5441)

Species reactivity: sheep, carp, feline, chicken, rat, mouse, *Hirudo medicinalis*, rabbit, canine, pig, human, bovine, guinea pig
Application: WB, IHC, IF, ELISA

anti-eIF5B (Proteintech, 13527-1-AP)

Species reactivity: Human, Mouse, Rat
Application: WB, IHC, IF, ELISA

anti-eIF4A1 (Abcam, ab31217)

Species reactivity: Mouse, Rat, Cow, Human, African green monkey

Application: WB, IP, ICC/IF, IHC-P

anti-eIF4E (Cell Signaling, 9742S)

Species reactivity: Human, Mouse, Rat, Monkey, Mink, Zebrafish
Application: WB, IHC

anti-Phospho-eIF2 α (Ser51) (Cell Signaling, 3398S)

Species reactivity: Human, Mouse, Rat, Monkey, D. melanogaster
Application: WB, IP, IHC

anti-eIF1A (Abcam, ab177939)

Species reactivity: Mouse, Rat, Human
Application: Flow Cyt (Intra), WB, ICC/IF, IHC-P, IP

anti-eIF1 (Proteintech, 15276-1-AP)

Species reactivity: Human
Application: WB, IP, IHC, ELISA

anti-eIF5 (Santa Cruz Biotechnology, sc-282)

Species reactivity: mouse, rat and human
Application: WB, IP and IF

anti-Ubiquitin (Santa Cruz Biotechnology, sc-8017)

Species reactivity: mouse, rat, human and Drosophila
Application: WB, IP, IF, IHC(P), FCM and ELISA

anti-ATF4 (Santa Cruz Biotechnology, sc-390063, clone B-3)

Species reactivity: mouse, rat and human

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)

HEK293-Kb cells (human, provided by J. W. Yewdell lab);
Lenti-X 293T Cell Line (Human, purchased from Clontech, 632180)

Authentication

Cell lines were not authenticated by ourselves.

Mycoplasma contamination

Cell lines were tested negative for mycoplasma contamination prior to our studies.

Commonly misidentified lines
(See [ICLAC](#) register)

No commonly misidentified cell lines were used.

Flow Cytometry

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Transfected cells are washed with PBS and harvested by trypsin. Cells are then re-suspended in blocking buffer (1% bovine serum albumin in PBS). Cells are aliquoted into a 96-well plate followed by 300g spinning for 2 min. After removal of blocking buffer, cells are washed one more time followed by staining with antibody. After incubation in the dark with gentle rocking at 4 °C for 30 minutes, cells are washed three times with 200 μ L of the blocking buffer to remove unbound antibodies. Cells are re-suspended in 300 μ L of blocking buffer followed by single cell filtering (Falcon). Cells are analyzed on a BD FACSAria Fusion flow cytometer (BD Biosciences). Cytometry data analysis is conducted using FlowJo.

Instrument

BD FACSAria Fusion flow cytometer.

Software

FlowJo software (Version 10).

Cell population abundance

At least 10,000 counts were recorded using a 0.5 mL s⁻¹ flow rate.

Gating strategy

Cells were first assessed in the FSC-A/SSC-A dot plot to exclude cell debris (P1) and doublet discrimination was carried out by additional plots to remove doublets and cell clumps (P2 and P3). Finally, fluorescence was detected in the FITC,

APC or PE channel. For a negative control, cells are transfected with Lipofectamine MessengerMAX only. Tools for fluorescence compensation were applied whenever needed. A detailed description of the experimental design can be found in the main manuscript.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.