

Ribosome profiling reveals the what, when, where and how of protein synthesis

Gloria A. Brar^{1,3,4} and Jonathan S. Weissman^{2–4}

Abstract | Ribosome profiling, which involves the deep sequencing of ribosome-protected mRNA fragments, is a powerful tool for globally monitoring translation *in vivo*. The method has facilitated discovery of the regulation of gene expression underlying diverse and complex biological processes, of important aspects of the mechanism of protein synthesis, and even of new proteins, by providing a systematic approach for experimental annotation of coding regions. Here, we introduce the methodology of ribosome profiling and discuss examples in which this approach has been a key factor in guiding biological discovery, including its prominent role in identifying thousands of novel translated short open reading frames and alternative translation products.

Ribosome footprints
mRNA fragments of ~30 nucleotides that result from nuclease treatment of translating ribosomes. These are mRNA regions that are protected by the ribosome as the mRNA is decoded to a protein sequence.

¹Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA.

²Howard Hughes Medical Institute, Department of Cellular and Molecular Pharmacology, University of California, San Francisco, California 94158, USA.

³Center for RNA Systems Biology, University of California, Berkeley, California 94720, USA.

⁴California Institute for Quantitative Biosciences (QB3), University of California, San Francisco, California 94158, USA.

e-mails: gabrar@berkeley.edu; jonathan.weissman@ucsf.edu
doi:10.1038/nrm4069
Published online
14 October 2015

Translation, which is the process by which a ribosome reads an mRNA template to guide protein synthesis, is a crucial step in gene expression. Translation is energetically costly and is therefore tightly regulated to conserve cellular resources, as well as to avoid mistakes that may result in the production of toxic proteins. Indeed, a wide range of disease states, including neurodegeneration, anaemia and specific developmental defects, result when the translational process is compromised (see selected REFS 1–6). Although much is known about the structure and function of the ribosome, our understanding of many aspects of the regulation of translation has been far more limited.

Efforts to globally monitor gene expression have historically focused on measuring mRNA levels (for example, using microarrays or RNA sequencing (RNA-seq)), although we know that translational control is an essential and regulated step in determining levels of protein expression. Until recently, precisely monitoring translation was far more challenging than was measuring mRNA levels. This has changed with the development of the ribosome profiling approach, which was first described in 2009 (REF. 7).

Ribosome profiling is a deep-sequencing-based tool that facilitates the detailed measurement of translation globally and *in vivo*⁷. At the core of this approach is the observation that a translating ribosome strongly protects about 30 nucleotides of an mRNA from nuclease activity^{8,9}. Sequencing of these ribosome-protected fragments, termed ribosome footprints, thus provides a precise record

of the position of the ribosome at the time at which translation was halted. Measuring the density of protected fragments on a given transcript provides a proxy for the rate of protein synthesis. In addition, determining the positions of the protected fragments makes it possible to empirically measure the identity of translation products (for example, where they begin and end and even the frame being read). This has led to the discovery of many novel or alternative protein products^{10–19}. The distribution of ribosome footprints can provide insights into the mechanism of translational control (for example, it can be used to identify regulatory translational pauses and translated upstream open reading frames (uORFs)). Finally, novel adaptations of the ribosome profiling approach make it possible to monitor translation mediated by subsets of ribosomes on the basis of their physical location in the cell or their interaction partners.

Here, we discuss the principles of the ribosome profiling approach, its strengths and limitations, and recent examples in which it has guided biological discovery. We focus on the value of ribosome profiling as a tool to interrogate what is being translated, how this translation is regulated and where in the cell the translation of specific sets of proteins occurs.

What is ribosome profiling and what can it reveal?

Ribosome profiling exploits the classical molecular method of ribosome footprinting^{8,9}, in which *in vitro* translated mRNAs are treated with nuclease to destroy the regions that are not protected by the ribosome^{7,9}.

Upstream open reading frames (uORFs). ORFs in the 5' leader region of a characterized mRNA transcript. Translation of uORFs may regulate translation of a downstream ORF. Ribosome profiling allows for the empirical identification of all translated uORFs *in vivo* under a condition of interest. Although uORFs are short, here we do not include them in the class of 'short ORFs', which are on an mRNA that was not previously thought to encode a protein.

Such treatment leaves 'footprints' of ~30 nucleotides, which can be mapped back to the original mRNA to define the exact location of the translating ribosome. Ribosome profiling extends this method by mapping and measuring the full complement of *in vivo* ribosome footprints to quantify new protein synthesis and to annotate coding regions globally^{7,10–12} (FIGS 1, 2). Extraordinary advances in sequencing technology²⁰ now make it possible to deeply sample all translating ribosomes. In mammalian cells, for example, which encode ~20,000 proteins with an average mRNA coding region of ~500 nucleotide triplets, nuclease digestion of all translating ribosome–mRNA complexes yields 10 million possible footprints. The billions of reads that are now possible with next-generation sequencing enable the reliable quantification of the set of footprints tiling across all but the rarest mRNAs, and a recently developed kit facilitates sample preparation^{21,22}. With such easily attainable and quantitative information,

ribosome profiling has a range of uses, from a broad proteomic tool to a specific probe of translation in an *in vivo* setting, and as a valuable complement to mRNA-seq.

Ribosome profiling requires collection of a physiological sample; inhibition of translation to freeze ribosomes in the act of translation; nuclease digestion to produce ribosome-protected fragments; and isolation of ribosomes and, subsequently, of ribosome footprints²¹. Ribosome footprints are converted to a strand-specific library and subjected to next-generation sequencing, and the fragments are then mapped to the appropriate reference genome. Ribosome profiling is typically carried out on a split sample, with parallel libraries constructed for measuring mRNA abundance by mRNA-seq. Comparison between the rates of protein synthesis and the abundance of mRNAs makes it possible to determine the translational efficiency for each mRNA⁷ (FIGS 1 a,b; 2b,c). The common biophysical properties of the ribosome and

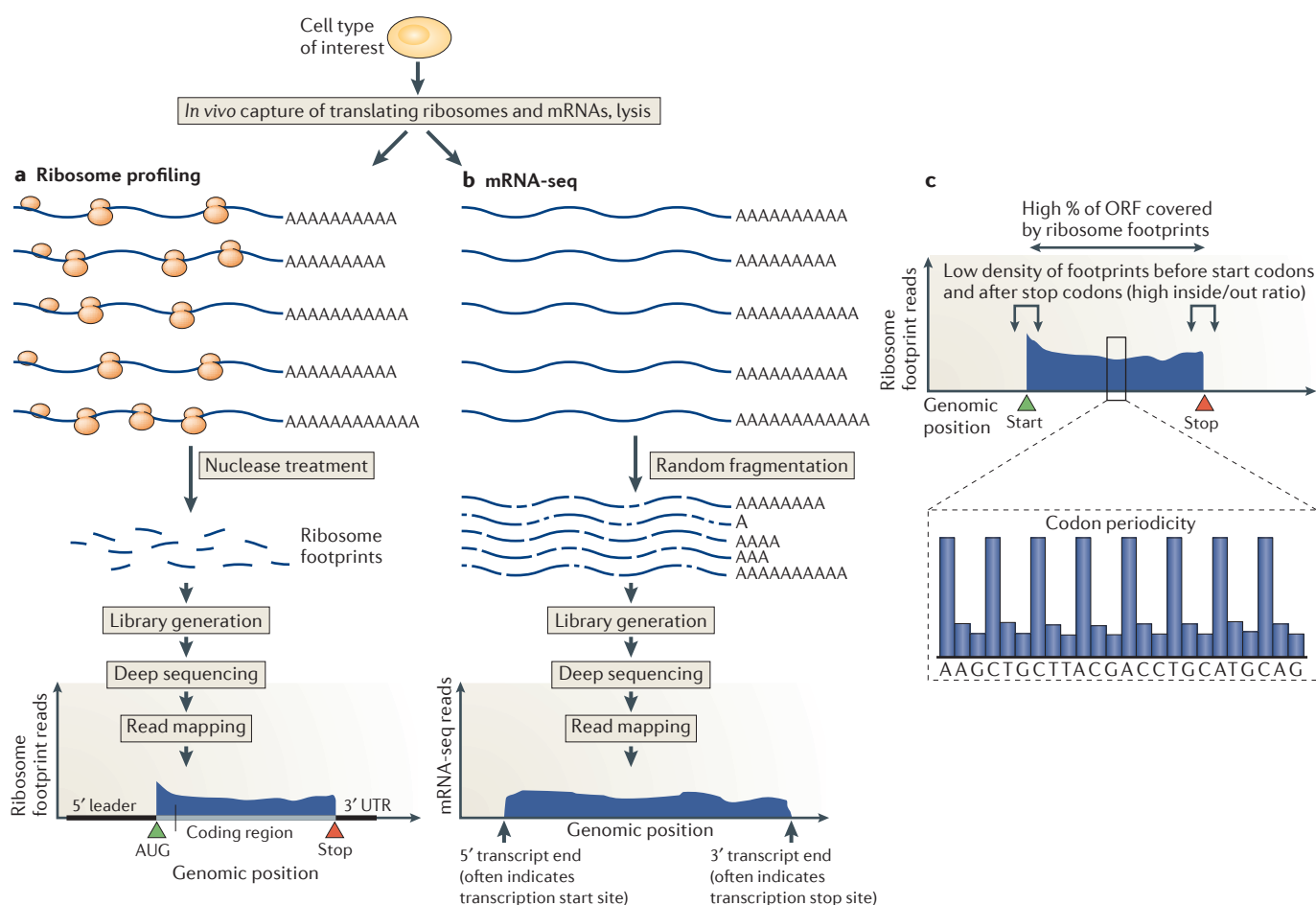


Figure 1 | An overview of ribosome profiling. **a** | Ribosome-bound mRNAs are isolated by size and treated with a nonspecific nuclease (typically RNase I or micrococcal nuclease), resulting in protected mRNA fragments termed 'footprints'. These ribosome footprints are isolated and converted to a library for deep sequencing. The ribosome footprints typically show precise positioning between the start and the stop codon of a gene, which facilitates global and experimental identification of genomic coding regions. **b** | By comparison, mRNA sequencing (mRNA-seq) captures random fragments covering the entire mRNA transcript. The positional information determined by standard mRNA-seq

allows approximate determination of transcript boundaries, but it is less precise than that collected by ribosome profiling, owing to the loss of 5' and 3' ends during the fragment generation method that is typically used. **c** | Translated open reading frames (ORFs) contain a stereotyped organization of ribosome footprints. Ribosome footprint density over ORFs begins sharply at the start codon, ends sharply at the stop codon and shows evidence of codon periodicity. True translated regions tend to show ribosome footprint coverage over the majority of the ORF and not typically in the regions before the putative start codon and after the putative stop codon. UTR, untranslated region.

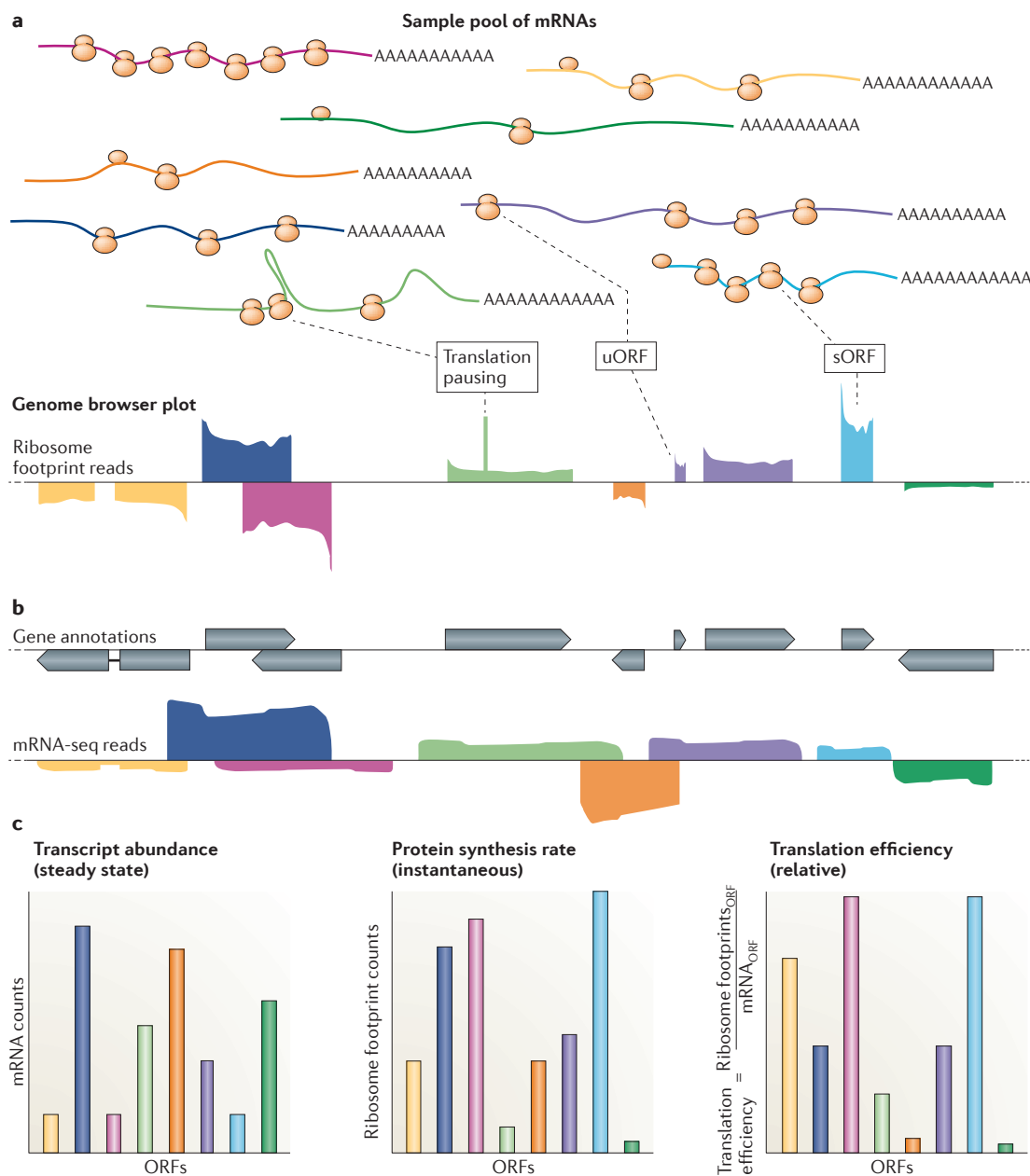


Figure 2 | Qualitative and quantitative data provided by ribosome profiling. **a** | A diverse sample pool of mRNAs, distinguished by colour, is shown, together with a corresponding representative genome browser plot of ribosome profiling data derived from this pool. Note that ribosome profiling facilitates experimental determination of translated regions, including short open reading frames (sORFs), which may be an important newly identified source of cellular peptides, and upstream ORFs (uORFs), which are thought to be largely regulatory. Pausing during translation elongation may result in peaks in ribosome footprint reads within ORFs. **b** | Overlaid gene annotations and mRNA sequencing (mRNA-seq) data for the examples in part **a** are shown. **c** | The graphs show examples of quantitative data derived from parts **a** and **b**. Note that transcript abundances may not correlate closely with the instantaneous protein synthesis rates. The collection of quantitative data for both transcript abundances and protein synthesis rates enables the relative translation efficiencies to be inferred. These can vary over several orders of magnitude within a given organism in a given state. The translation efficiency can also change over time for a given mRNA, reflecting dynamic regulation at the level of translation.

the lack of genetic manipulation that is required for this approach make ribosome profiling highly adaptable to cells or tissues from essentially any organism, with modest modifications. Organisms that have been investigated thus far by ribosome profiling include a variety of bacteria, yeast, parasitic protozoa, zebrafish, flies, nematodes,

mice, rats, plants, viruses and human cells^{7,10–12,19,23–30}. Even mitochondrial translation within human cells has been effectively assayed by this method³¹, and a similar approach has been applied to chloroplasts in plant cells³². Many of these data sets have been compiled and made readily accessible for data mining and comparison³³.

What are the strengths of ribosome profiling?

Despite its recent development, ribosome profiling has rapidly become a widely used tool for understanding diverse and complex biological issues. Three key features, which are outlined below, have facilitated the broad utility of this method.

Sensitivity and precision of quantification. Ribosome profiling provides a large dynamic range for the detection and quantification of translation in unperturbed cells. The sensitivity of the method, which results from the depth of sampling that is possible in sequencing studies, facilitates the measurement of even relatively rare translation events, with the range of detection generally limited only by the counting variability that is seen with very low numbers of sequencing reads. Complementary methods, including pulsed label-based mass spectrometry, analyses of transcript distributions on polysome gradients and ³⁵S Met-based metabolic labelling, enable sensitive measurement of new protein synthesis; however, the highly parallel sequencing readout of all ribosome positions that is provided by ribosome profiling typically yields more quantitative and detailed information than is currently accessible by alternative methods.

Precision of positional information. In addition to its broad dynamic range of detection, ribosome profiling provides uniquely rich and precise positional information. The almost universal biophysical properties of ribosomes across species yield a characteristic footprint size that allows prediction of the codon in the ribosome P site (that is, the position of peptide bond formation) and the detection of codon periodicity⁷ (FIG. 1c). Analyses of ribosome footprint positions can be used to mechanistically probe aspects of translation, thus far identifying many novel examples of ribosomal frameshifting, stop codon readthrough, ribosome pausing, translation initiation at non-AUG codons and uORF translation^{7,10,12,34–37} (FIG. 2a). Furthermore, years after many genomes were originally annotated, the precise positional information obtained from ribosome profiling experiments has provided the first opportunity to experimentally define translated ORFs^{11–13,38,39} (FIGS. 1, 2), resulting in the identification of new classes of coding regions in diverse organisms.

Instantaneous measurements. A final valuable property of ribosome profiling is the instantaneous nature of the information that is collected, which reflects a snapshot of the dynamic process of translation. Although mRNA-seq and standard genome-scale mass spectrometry experiments are valuable in following gene expression globally, these widely used measurements report on steady-state levels of mRNA and protein, respectively. This information is important, but it may not reflect the rapid cellular decision making that accompanies developmental transitions and environmental responses. Ribosome profiling enables sensitive detection of changes in cellular protein expression as they occur^{7,12,40}. The common, quantitative output from ribosome profiling and mRNA-seq further allows for direct comparison of instantaneous protein

synthesis and steady-state transcript levels, providing an opportunity to quantify *in vivo* translation efficiencies in detail (FIG. 2c).

What are the limitations of the method?

We discuss below notable weaknesses and caveats of ribosome profiling that should be considered when using the method or interpreting data derived from its use.

Experimentally introduced distortions. The key technical challenge of ribosome profiling is the need to rapidly inhibit translation to capture a snapshot of ribosomes in a particular physiological state. The reliability of this step is particularly important for any analyses of translation pausing, as the fast rate of translation elongation may result in signal blurring or the artificial accumulation of ribosomes at specific positions if inhibition is slow. The use of a translation elongation inhibitor (such as cycloheximide) can be valuable; however, it is clear that such inhibitors can alter the local distributions of ribosomes on an mRNA, especially near translation start sites^{7,18,21,41}. Although this does not seem to interfere with the global measurements of the density of ribosomes on an mRNA that are used to determine rates of protein synthesis, it can cause spurious peaks of ribosome binding at particular sites. Thus far, flash freezing has been the most robust approach in a wide range of diverse organisms and has enabled the physiological capture of local and global ribosome distributions²¹. In general, each experimental step — from cell harvesting to nuclease digestion to library generation — has the potential to cause distortions in the data output. These distortions must be accounted for carefully, as the degree to which any given distortion might be problematic will depend strongly on the questions being addressed and the system being probed.

The need to infer protein synthesis rates. A caveat to consider when interpreting ribosome profiling data is that rates of synthesis are typically inferred from the average ribosome density along the mRNA in question. The accuracy of this measure depends on the premise that all ribosomes finish translation and that, on average, the translation elongation rate is similar among different mRNAs in a cell. These assumptions can be tested and are appropriate for a wide range of conditions, but this will not always be the case. Known exceptions^{42–44} — including the build-up of ribosomes at and immediately proximal to the start codon in a partially cycloheximide-dependent manner⁷ or regulated translation pausing and abortion under starvation conditions⁴⁵ — can be corrected for to increase measurement accuracy, but there may be cases in which these and other, currently unknown, exceptions pose challenges for proper data analysis.

Contaminating footprint-sized fragments. Another important issue for ribosome profiling experiments is that footprints are inferred on the basis of their size and their association with assembled (80S) ribosomes. Contaminating RNA fragments, including those from

Polysome gradients

A method for fractionating ribosomes that are bound to mRNAs by velocity centrifugation of cell extract on sucrose gradients, allowing for the separation of mRNAs that are associated with one ribosome (monosome) from those being translated by multiple ribosomes (polysome). Sucrose gradient fractionation facilitates qualitative analysis of the translation status of cells.

Ribosome P site

The site within an actively translating ribosome that is usually associated with the tRNA attached to the growing peptide chain.

Codon periodicity

The three-nucleotide pattern of ribosome occupancy, reflecting mRNA translocation in the ribosome by codon as translation occurs.

structured non-coding RNAs or large ribonucleoprotein complexes that co-migrate in a sucrose gradient with the ribosome, may be processed with a ribosome profiling library and provide false readouts of translation (see [Supplementary information S1](#) (figure)). A recent approach, termed fragment length organization similarity score (FLOSS) analysis, aims to identify such fragments and remove them post-experimentally (that is, *in silico*)³⁹. FLOSS analysis is based on the observation that bona fide ribosome footprints have stereotypical distributions of footprint sizes (see [Supplementary information S1](#) (figure, parts a and b)). The distribution of typical 80S footprint sizes used in FLOSS analysis is empirically measured for each experiment, by examining the sizes of footprints in that same experiment from known protein-coding regions, and can be used to computationally identify contaminating fragments for removal. Nonetheless, there are examples in which genuine 80S mRNA footprints do not conform to the typical size pattern. Two recent cases that highlight interesting biology that was determined by analysis of alternatively sized ribosome footprints indicate effects that are due to both alternative ribosome conformations⁴⁶ and alternative mRNA properties⁴¹ (see below). Nuclease protection assays can be a useful adjunct control for identifying the full range of ribosome footprint sizes in a new organism or condition, thus informing the design of a ribosome profiling experiment to best capture all translating ribosomes in a given system.

Ribosomal RNA (rRNA) fragments commonly result from the nuclease-treatment step of ribosome profiling and may substantially decrease the ribosome footprint sequencing space in a ribosome profiling experiment⁷, particularly under conditions in which global translation levels are low. Whereas mRNA-seq often uses poly(A) selection as an effective method for the isolation of desired sequences, this approach is not possible with ribosome profiling. Selective subtraction of ribosomal fragments, however, is highly effective and is recommended, particularly for samples in which a small number of specific footprint-sized rRNA fragments are seen as contaminants²¹.

Mapping ambiguous reads. A general challenge in the analysis of sequencing data is determining the correct alignment position for reads from repetitive or highly similar regions, such as gene families, or from alternative transcript variants. In the case of genome sequencing or mRNA-seq, longer reads or paired-end⁴⁷ approaches can help to resolve such ambiguities, but the inherently short size of ribosome footprints precludes these experimental approaches. However, computational methods that have been developed for mRNA-seq data to assign multiply mapping reads in a probabilistic manner on the basis of overall read distributions⁴⁸ can be applied to ribosome profiling data to mitigate this limitation.

Material quantities. Currently, the main limitation of ribosome profiling compared with mRNA-seq approaches is the requirement for relatively large samples. In contrast to mRNA-seq⁴⁹, ribosome profiling

cannot yet be applied to single cells. This limitation results from the extra processing step that is required to isolate ribosomes²¹, as well as the small proportion of any given mRNA molecule that is being translated at any given instant and thus recoverable as footprints (FIG. 1a). It is likely that the types of technical advances that have greatly enhanced the sensitivity of RNA-seq approaches to small cell numbers will also be applicable in the future to ribosome profiling, although no such major effort has yet been undertaken.

Insights provided by ribosome profiling

With these advantages and disadvantages in mind, the application of ribosome profiling to specific biological questions has confirmed much of what we know about translation mechanism from decades of elegant structural, biochemical and genetic studies⁵⁰. Ribosome profiling has also made it possible to monitor translation with unprecedented depth and precision, providing important — and at times surprising — insights. The application of this method to numerous organisms and cellular states has illuminated fundamental aspects of cell biology that were previously challenging to probe experimentally, providing measurements for how much of each protein is synthesized, how translation is regulated, where synthesis starts and stops and what is being synthesized.

How much? A quantitative view of protein synthesis. The simplest and broadest application of ribosome profiling is as a quantitative proteomics tool to monitor which proteins are being synthesized, and at what levels, thus providing rich molecular insight into a given cell state. Ribosome footprint density reflects the number of ribosomes at a given position. Assuming that the average translation elongation rate is similar for different genes, ribosome profiling provides direct, global and quantitative measurements of rates of protein synthesis, thereby capturing information that has been largely invisible to gene expression measurements of mRNA levels alone. Mass spectrometry can, in principle, be used to measure rates of protein synthesis; however, this is technically difficult, as it typically requires metabolic labelling and multiple measurements per sample. Analysis of the positions of mRNAs in polysome gradients provides valuable complementary information to that obtained with ribosome profiling, but again, this method is laborious and typically yields only a qualitative measure of protein synthesis.

In many cases, the ability to observe new protein synthesis globally and quantitatively provides insights that are not apparent from measurements of mRNA abundance. Bacterial operons provide a vivid example of the value of being able to directly measure rates of protein synthesis. As is the case for many protein complexes in bacteria, the eight different subunits of the F_0F_1 -ATP synthase are expressed from a single polycistronic mRNA, and thus measurements of mRNA levels would suggest that the subunits are all expressed at very similar levels. Ribosomal profiling, however, shows that the individual ORFs that encode the subunits of the F_0F_1 -ATP synthase

Fragment length organization similarity score (FLOSS) analysis

A metric for determining the probability that ribosome footprints over a given region (or set of regions) result from translation. This analysis involves comparing size distributions of footprints over a query region and over validated coding regions and is based on the concept that the biophysical properties of translating ribosomes result in characteristic signatures in ribosome footprint sizes.

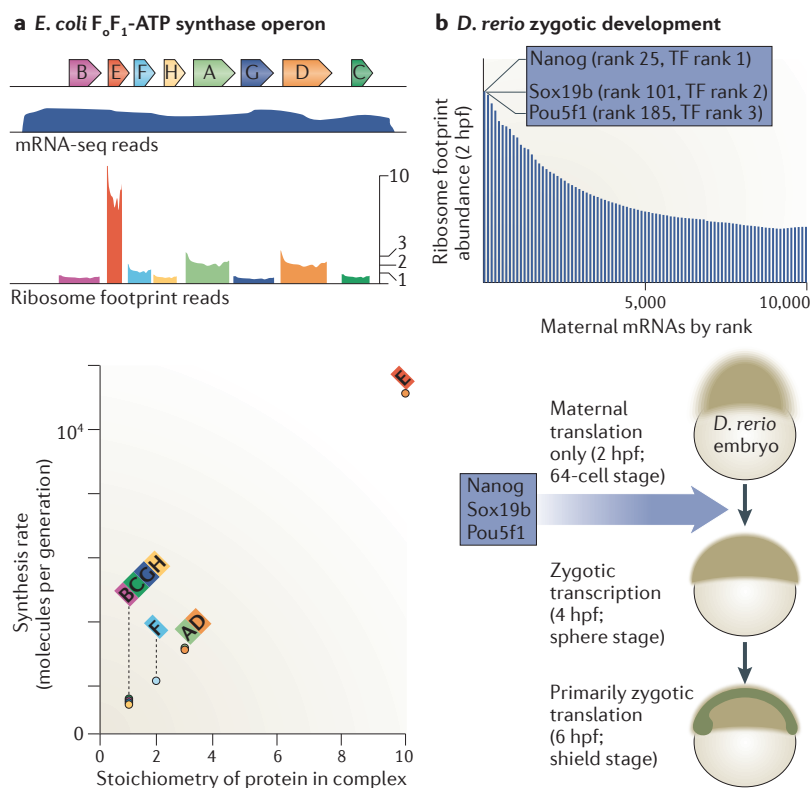


Figure 3 | Ribosome profiling facilitates quantitative proteomic discovery in diverse systems. **a** | Bacterial cells translate components of multi-member protein complexes in ratios that are proportional to their stoichiometry in these complexes. A notable example is the F_0F_1 -ATP synthase, which is composed of eight different proteins (A to H), translated from a single operon. mRNA abundance for each gene is thus similar, but ribosome profiling reveals intricate translational control. **b** | Zebrafish zygotic development requires the initiation of zygotic transcription 2 hours post-fertilization (hpf), although the specific transcription factors responsible for this transcription have been unclear. Ribosome profiling of embryos at 2 hpf showed that the three most highly translated transcription factors (TFs) from maternal mRNAs were Nanog, Sox19b and Pou5f1, and subsequent experiments confirmed that these three proteins drive zygotic activation. *D. rerio*, *Danio rerio*; *E. coli*, *Escherichia coli*; mRNA-seq, mRNA sequencing. Part **a** is modified with permission from REF. 51, Elsevier. Part **b** is modified from REF. 54, Nature Publishing Group.

operon are translated at a ratio of 1:1:1:1:2:3:3:10. Remarkably, these ratios precisely reflect the stoichiometry of these components in the ATP synthase^{51,52} (FIG. 3a). This property of proportional synthesis, by which subunits of multiprotein complexes are synthesized at rates that are proportional to their stoichiometry in the complex, turns out to be generally true for *Escherichia coli* and was also observed for some (but not all) complexes in budding yeast (*Saccharomyces cerevisiae*). Such measurements of instantaneous rates of protein synthesis may prove to be a general tool for exploring how proteins assemble and function together⁵¹.

Quantitative measurement of protein synthesis rates over multiple time points of a dynamic process can also provide information about specific gene function. For example, hierarchical clustering of patterns of new protein synthesis for each gene over the dynamic process of meiosis in budding yeast resulted in an intricate map of gene expression that provided highly detailed functional

information¹². In these data, the genes responsible for the complex, conserved and meiosis-specific processes of homologous recombination and synaptonemal complex assembly emerged as a single cluster of 46 genes. This observation was surprising, because these processes are known to be regulated extensively at the post-translational level, and also because the cluster included almost every gene that had been found through decades of intensive genetic and cytological screening focused on these processes. This cluster also included several uncharacterized genes, two of which (*GMC1* and *GMC2*) were subsequently shown to have roles in recombination and synaptonemal complex formation^{12,53}.

Another striking recent example of this type of analysis used ribosome profiling to identify the factors that are responsible for initiation of the zygotic developmental programme in zebrafish⁵⁴ (FIG. 3b). The initiation of zygotic development in vertebrates depends heavily on translational control, as maternal mRNAs provide the starting pool of material for translation. Zygotic activation then requires destruction of these maternal mRNAs and transfer of developmental control to the zygote itself. To determine the factors that mediate the first wave of zygotic transcription, ribosome profiling data were analysed for samples collected just before zygotic activation. This study identified Nanog, Sox19b and Pou5f1 as the three transcription factors that were most heavily translated, from the large pool of maternal mRNAs at this stage (FIG. 3b). Subsequent morpholino knockdown experiments showed that specifically blocking translation of these three factors resulted in a shutdown of the first wave of zygotic transcription and development, indicating that they are the key factors responsible for the initiation of the zygotic developmental programme⁵⁴.

Other recent studies in disparate systems — from the *Drosophila melanogaster* oocyte-to-embryo transition⁵⁵ to the Trypanosome life cycle⁵⁶ to the mammalian cell cycle⁵⁷ to plants under hypoxic conditions²⁷ — have used ribosome profiling to identify specific proteins that drive these complex processes. Cases in which ribosome profiling data provide markedly different information than can be obtained by traditional mRNA abundance measurements for gene expression tend to fall into two categories: systems in which transcriptional regulation is minimal^{26,54,55}; and dynamic cellular programmes^{11,12,27,35,57–59}. The latter category includes cellular differentiation, organismal development and dynamic responses to cellular stress, which are all cases in which the instantaneous and downstream gene expression measurements provided by ribosome profiling are particularly illuminating for understanding molecular control.

How? Insights into the mechanism of translational control. The basic mechanism by which the ribosomal machinery reads codon information in mRNAs to create proteins is conserved, and many features of this process are well understood⁵⁰. Nonetheless, there are aspects of translational control that are not amenable to recapitulation *in vitro* and for which results from

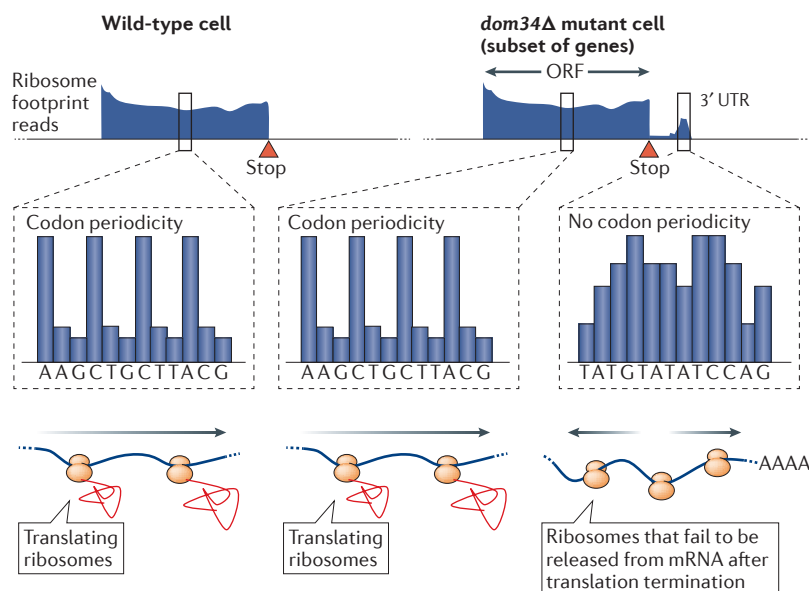


Figure 4 | Dom34 facilitates the release of 80S ribosomes from a subset of 3' untranslated regions (UTRs). Ribosome footprints indicative of assembled 80S ribosomes are seen in a subset of 3' UTRs in *dom34Δ* mutant cells. Unlike 80S ribosome footprints from open reading frames (ORFs), however, these do not show codon periodicity and represent ribosomes that have failed to properly release following translation termination.

genetic approaches alone may be difficult to interpret, owing to complex secondary effects that result from cellular adaptation to chronic abnormal protein synthesis. Furthermore, ribosome profiling facilitates the identification of translation mechanisms that vary across organisms, cellular state and individual transcripts, as well as the study of the roles of specific translation factors. Several important examples of discovery in translation mechanism have been highlighted in previous reviews^{22,60,61}. Here, we focus on just two recent studies in which ribosome profiling has illuminated important aspects of translation.

Dom34 (a homologue of eukaryotic release factor 1) has been shown to help to dissociate stalled ribosomes *in vitro*, but how and where it acted *in vivo* was unclear. Recent work explored the function of this protein through ribosome profiling of wild-type and *dom34Δ* budding yeast cells. The authors reasoned that if Dom34 was either dissociating ribosomes on truncated transcripts or causing multiple ribosomes to stack up owing to stalling, then the relevant footprints might be smaller or larger, respectively, in the absence of Dom34 (REF. 41) (FIG. 4). Indeed, in the case of the *HAC1* (homologous to *Atf/Creb1*) transcript, which was previously shown to exist in a truncated form in the cytosol⁶², ribosome profiling showed that *dom34Δ* budding yeast cells accumulated ribosomes with abnormal footprint sizes, indicating a defect in ribosome recycling at these sites⁴¹.

The largest effect that was revealed by ribosome profiling of *dom34Δ* cells — the presence of abundant ribosome footprints in 3' untranslated regions (3' UTRs) on a subset of mRNAs — was unexpected. In contrast to ribosome footprints in coding regions, the footprints

that mapped to 3' UTRs in the absence of Dom34 were not restricted to a single reading frame (FIGS. 1c, 4). This observation suggested that these footprints did not represent canonical translating ribosomes but were instead likely to result from a population of ribosomes that had failed to be released from mRNAs following translation termination (FIG. 4). Together, these data indicate that ribosomes are not always automatically released following stop codon recognition, and that Dom34 has a role in freeing ribosomes from truncated transcripts and 3' UTRs⁴¹.

Another important application of ribosome profiling has been the analysis of the mechanisms of drugs that target translation. Macrolides, for example, are a class of clinically important antibiotics that are known to bind in the nascent peptide exit channel of the ribosome. Macrolide activity has long been thought to cause early translational inhibition by blocking the egress of nascent peptides from the ribosome. However, this view has been overturned by recent ribosome profiling studies^{63–65}, which found that macrolides function primarily by selectively affecting the ability of the ribosome to form peptide bonds in specific sequence contexts. A key observation was that in bacteria that were treated with high doses of erythromycin or of telithromycin, a next-generation macrolide, not all protein synthesis was inhibited. In fact, telithromycin inhibited the translation of fewer proteins than erythromycin, despite being a more effective antibiotic⁶⁴.

The application of ribosome profiling to bacterial cells treated with erythromycin or telithromycin also showed that, even in cases of inhibited translation for a given mRNA, the ribosome did not always stop translating early in the transcript, as predicted by the classical model for macrolide action. Rather, ribosome footprint build-up, which is indicative of ribosome stalling, could be seen at various regions in the subset of mRNAs that were inhibited. The precise positional information that was obtained from these experiments made it possible to determine that these points of translation interruption were dependent on specific positively charged sequences ([R/K]X[R/K]) that were present in the peptidyl transfer centre of the ribosome. Macrolide-mediated inhibition of translation thus was not occurring primarily through obstruction of the peptide exit channel of the ribosome but instead was a result of ineffective peptide bond formation for certain amino acid sequences. This effect could be recapitulated precisely *in vitro* for some mRNAs, but poorly for others, suggesting that additional cellular factors might contribute to macrolide action⁶³. This improved understanding of macrolide mechanism has direct relevance to the development of newer, more effective antibiotics.

Where? Monitoring localized translation. A hallmark of eukaryotic cells is the presence of intricate subcellular structures that facilitate the compartmentalization of different biological processes. Localized protein synthesis has a crucial role in creating these subcellular structures by allowing proteins to be produced at their sites of action and in response to local cellular need (see REF. 66

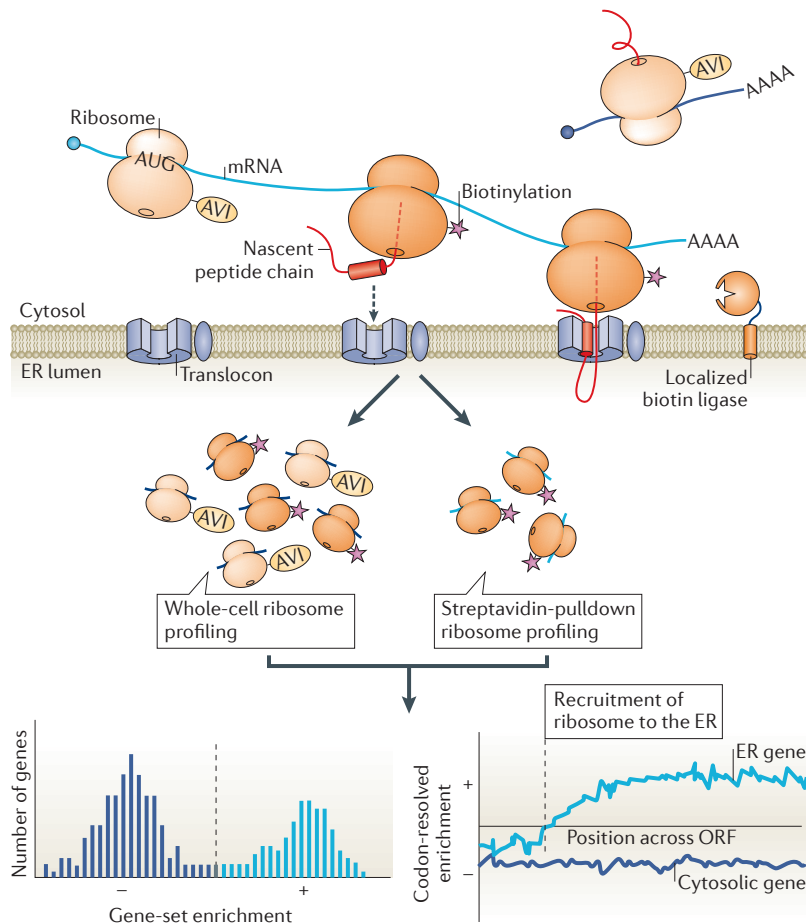


Figure 5 | Proximity-specific ribosome profiling at the endoplasmic reticulum (ER).

A ribosome subunit is fused to a biotin-acceptor (AVI) tag and BirA biotin ligase is fused to a localization element that spatially restricts its activity, for example, to the ER. Only ribosomes that orient AVI towards the ER surface, as seen during their close association with the ER membrane during protein translocation, are biotinylated when a controlled pulse of biotin is applied to cells. Cells are then frozen and ribosomes are collected. Ribosome profiling is carried out on all ribosomes and also specifically on ribosomes pulled down with streptavidin. The pulldown-enriched mRNA population (light blue) represents genes that are greatly enriched for translation at the ER.

The positional data from these analyses also reveals the point in the message at which a translating ribosome is recruited to the ER. ORF, open reading frame. Modified from Jan, C. H., Williams, C. C. & Weissman, J. S. Principles of ER cotranslational translocation revealed by proximity-specific ribosome profiling. *Science* **346**, 1257521 (2014).

Reprinted with permission from AAAS.

for a review). As translation is an important amplification step, localization of a single mRNA molecule can allow for correctly localized synthesis of hundreds of protein molecules. In addition, such local synthesis prevents potentially toxic effects of proteins being present — even if only during transit time — in an inappropriate cellular compartment. Finally, localized translation allows for the regulation of protein synthesis on the basis of a proximal stimulus, such as that seen in dendrites in response to neuronal stimulation, which is thought to contribute to the learning process⁶⁶.

Despite the broad importance of localized translation, few gene expression analysis tools are available that faithfully preserve spatial information. Until recently, global approaches for studying subcellular control of protein

synthesis have been limited to bulk interrogations that cannot uniquely identify proteins or that require careful biochemical fractionation of the compartment of interest, which limits both the location and the resolution of analyses. Proximity-specific ribosome profiling now enables *in vivo* measurement of localized translation within cells. The basis of proximity-specific ribosome profiling is selective biotinylation of ribosomes in a manner that depends on their subcellular location in intact, unperturbed cells (FIG. 5). The use of *in vivo* labelling allows the recovery of ribosomes from defined locations, including those that cannot be purified by classical cell fractionation techniques. Combining this purification strategy with ribosome profiling provides a tool for the identification of locally translated transcripts and sub-codon monitoring of translation at the site of interest.

So far, proximity-specific ribosome profiling has been used to probe two processes, translocation into mitochondria and into the endoplasmic reticulum (ER), with both studies yielding unexpected results^{67,68}. In the case of mitochondria, the approach provided insight into a long-standing question: do mitochondrial proteins begin translocation co-translationally, or is the predominant route of mitochondrial translocation post-translational? Proximity-specific ribosome profiling showed that the majority of mitochondrial inner membrane proteins — but not proteins targeted to other mitochondrial sites — were co-translationally targeted⁶⁷. These studies also revealed exquisite specificity in protein trafficking, with the vast majority of translocated proteins that were identified being targeted exclusively to either the ER or the mitochondria. A prominent exception was the fumarate reductase Osm1; follow-up studies showed that dual targeting of this protein resulted from the translation of alternative isoforms with distinct targeting signals⁶⁷.

Monitoring of translation on the ER surface determined several principles that are used by cells to coordinate translation and ER targeting⁶⁸ (FIG. 5). First, this work showed that co-translational targeting to the ER is pervasive and is principally determined by the location of the hydrophobic targeting sequence within the protein. The observation that co-translationally targeted mRNAs can be translated at the translocon immediately after or even before translation of their targeting sequence suggested a crucial role for polysomes in retaining mRNAs at the ER. In addition, distinct translocon complexes engage nascent chains at different points during synthesis. ER-targeted nascent chains typically undergo a conformational rearrangement within the translocon that results in a 'looped' conformation of the nascent chains, with their amino termini facing the cytosol. However, proximity-specific ribosome profiling revealed that a subset of proteins, the targeting of which requires the translocon-associated factor secretory 66 (Sec66), engage the translocon only after 120 amino acids have been synthesized, which facilitates the direct adoption of the looped conformation. Finally, monitoring the fate of ER-associated ribosomes following translation termination using pulsed biotinylation experiments showed that any given ribosome can exchange readily between the ER and the cytosol, as ribosomes labelled on the ER are

Translocon

The proteinaceous tunnel through which nascent proteins cross the endoplasmic reticulum membrane.

able to access the full pool of cytosolic mRNAs following at most a few rounds of translation at the ER⁶⁸.

In principle, proximity-specific ribosome profiling could be applied to any subcellular location for which it is possible to target biotin ligase activity. It can also be combined with approaches that analyse different poly-some fractions^{55,69} or with the translating ribosome affinity capture (TRAP)^{70–76} strategy. Together, these techniques could make it possible to explore regulated localized translation in specific neuronal subtypes in response to learning programmes.

What is being made? Defining translation events. Perhaps the most surprising emergent area of discovery that has been facilitated by ribosome profiling results from the ability of the method to identify, in a systematic manner, the full set of ribosome-translated polypeptides in a cell. Algorithm-based analyses of the genomic sequence of an organism alone can direct identification of probable coding sequences. Such strategies, however, are based on assumptions about what a coding region should look like, including start and stop codon identity, splice junction cues, conservation and the total codon length of an ORF. Such approaches for identifying protein-coding genes could miss functional coding sequences, particularly those that are short and/or species specific⁷⁷. These approaches might also miss coding regions that result from translational frame-shifting or stop codon read-through. Furthermore, translation and protein synthesis have effects beyond the production of stable proteins with discrete molecular functions. Polypeptide products from all cellular translation must be degraded, and non-canonical translation products yield unanticipated antigens that may have roles in viral detection or in autoimmunity^{39,78}. Finally, the process of translation can affect the stability of the template message by triggering co-translational decay pathways including nonsense-mediated decay⁷⁹. Thus, knowing which transcripts are translated has important implications for the fate of the mRNA, the ribosome and the cell. Ribosome profiling provides a unique opportunity to experimentally address this question in a given biological system or cell state of interest.

Ribosome profiling data from many organisms have generally provided experimental evidence for the translation of ORFs that had already been computationally predicted to encode proteins. These data have also suggested a diverse set of translated areas outside canonical coding regions (reviewed in REFS 60,80). These include, in some cases, ribosome footprints that are not clearly organized within ORFs, most commonly in 5' leader regions and mammalian long non-coding RNAs. The importance of translation of these regions remains an open question, although the unusual patterns of ribosome footprints that are often observed suggest that they may not reflect regions that are translated into canonical peptide products. In some cases, the translation that produces these footprints may mediate translational regulation, as is the case for translation of regulatory uORFs. Alternatively, some such cases may reflect translation that is used to regulate mRNA stability⁸¹.

However, in diverse organisms and conditions, ribosome footprints are seen that are organized within ORFs that were not previously known to encode proteins, in a manner that resembles those in canonical coding regions (as in FIG. 1c). This indicates that there is greater coding-region diversity and flexibility than had previously been recognized^{10–13}. The translated ORFs that have been defined by such ribosome footprints fall into two broad categories: translated short ORFs (sORFs) in predicted intergenic regions, often on RNAs that had been provisionally characterized as non-coding; and translated ORFs encoding alternative isoforms of known proteins. Both categories could represent major emergent areas of biological importance.

How pervasive is sORF translation? Algorithms for predicting protein-coding regions typically rely on assumptions about translated ORF length. The minimum ORF length of 100 codons that is used by most computational annotation approaches was chosen both to minimize the number of false positive gene calls and to reflect the predicted biophysical folding stability of 100-amino-acid proteins relative to shorter amino acid strings. Recently, however, several short peptides have been shown to be translated and to have crucial intracellular and extracellular roles in metazoans^{14,82–84}. Concomitant with these findings, ribosome profiling data in several systems, including mouse embryonic stem cells, meiotic yeast cells, hypoxic plants and virus-infected human fibroblasts, have identified many ribosome footprints that fall outside canonical coding regions but that cover short and discrete regions between an AUG and a stop codon^{10–12,16,27,85}. These observations suggest that canonical protein-coding sequences may be only a subset of the sequences that are translated in cells.

There are, however, some features of the newly identified translated sORFs that have led to doubts about their authenticity. First, some are present on RNAs that were thought to be non-coding^{10–12,82,83,86}. In many cases, these sORFs are not well conserved^{13,87,88}. They also sometimes seem to be translated in overlapping reading frames^{10–12,87,89}, a feature that has been thought to be unusual among typical eukaryotic genes (although ribosome profiling data have recently been used to identify such cases among canonical genes, as well⁹⁰). Finally, translated sORF products are difficult to detect systematically using mass spectrometry approaches. The validation or exclusion of these regions as examples of biologically relevant translation has been a major recent focus of interest.

Several analytical approaches to ribosome profiling data allow rigorous testing of the degree to which ribosome footprints over newly predicted translated sORFs match those that are seen for traditional protein-coding sequences (TABLE 1). These analyses often examine whether ORFs that are predicted to be translated by ribosome profiling show footprint organization that is consistent with the canonical mechanism for translation, such as sharp footprint-abundance transitions at known start codons and stop codons, and codon periodicity^{12,13,16,38,85,87,91} (FIG. 1c; TABLE 1) (discussed in REF. 80).

Translating ribosome affinity capture (TRAP). A method that allows identification of translated mRNAs on the basis of their *in vivo* association with a tagged ribosomal subunit that is expressed in a cell type-specific manner. This method is a valuable tool for assaying tissue-specific translation in animal and plant systems.

Nonsense-mediated decay mRNA degradation, which has traditionally been thought to result from stop codons that terminate translation more 5' than is usual on an mRNA.

Short ORFs (sORFs). Open reading frames of fewer than 100 codons on mRNAs that are not known to encode a canonical (long) protein. sORFs are a class of ORF that have not traditionally been thought to be frequently translated, although ribosome profiling and other approaches have recently validated the translation of thousands of sORFs in a range of organisms.

ORFs encoding alternative isoforms of known proteins Open reading frames (ORFs) that differ from another ORF at the same locus in either the start codon or the stop codon position but share the same reading frame. Translation of these ORFs may result in, for example, different subcellular targeting for a similar protein.

Table 1 | **Novel translated ORFs compared with characterized translated ORFs by diverse metrics**

Metric	Characterized sequences		Novel translated ORFs identified by ribosome profiling		Refs
	Non-coding RNAs (such as snoRNAs, tRNAs, <i>XIST</i> and <i>HOTAIR</i>)	Protein-coding sequences	sORFs	uORFs	
Association between footprint arrangement and putative start codons*	No general association	Footprint-covered regions usually start precisely at AUG codons, occasionally at near-cognate codons	Footprint-covered regions usually start precisely at AUG codons, occasionally at near-cognate codons	Footprint-covered regions often start precisely at AUG and at near-cognate codons	7,11,12
Association between footprint arrangement and putative stop codons*	No general association	Footprint-covered regions stop precisely at canonical stop codons	Footprint-covered regions stop precisely at canonical stop codons	Footprint-covered regions stop precisely at canonical stop codons	7
Footprint abundance relative to mRNA abundance	Very low (especially for properly sized footprints)	Low to high, depending on translation efficiency	Low to high, depending on translation efficiency	Low to high, depending on translation efficiency	7,10,13,38
Codon periodicity of footprints*	No	Yes	Yes	Often unclear owing to generally short length	7,13,41,90
Signatures of protein-coding conservation	No	Often	Sometimes, difficult to assess for very short regions	Unclear, primarily owing to short length	13,38,77,80,98
Identification of protein product by mass spectrometry	No	Often	Sometimes (dependent on length and peptide properties)	Sometimes (dependent on length and peptide properties)	11,92–97
Stable physical association of transcript with ribosomes	Not generally, but may occur in specific cases (for example, tRNAs)	Yes	Yes	Yes	17,39
Sensitivity of footprints to translation inhibitors	No	Yes	Yes	Yes	39
FLOSS (fragment length organization similarity score)*	High	Low	Low	Low	39
% putative ORF covered by footprints*	Low	High	High	Difficult to assess owing to frequent uORF overlap	38
Inside/out ratio (local enrichment of footprints within putative ORF)*	Low	High	High, difficult to assess when translated sORFs overlap	Difficult to assess owing to frequent uORF overlap	38
Ratio of footprints at putative start codons to footprints at immediately prior codons*	Low	High	High	High	12
RRS (ribosome release score)	Poor	Good	Sometimes high, but particularly poor in cases of translated sORF overlap	Frequent overlap in uORF translation leads to poor scores, difficult to assess	87
Cellular function determined by genetic or molecular analyses	Sometimes	Sometimes	Rarely, thus far, but important examples exist	Not assayed in many cases, but a subset are regulatory for translation of other ORFs	14,82–84
Summary					
Likelihood on the basis of the above metrics that regions encode functional proteins or peptides	Low	High	High for a subset, but unclear how generally functional the peptide products are. Likely to be a heterogeneous population with diverse roles.	Unclear; uORF regions predicted to be translated by ribosome profiling probably represent true translation, but resultant peptides may not be stable.	

HOTAIR, HOX transcript antisense RNA; ORF, open reading frame; snoRNA, small nucleolar RNA; sORF, short ORF; uORF, upstream ORF; *XIST*, X inactive specific transcript. *See the glossary terms, FIG. 1c, FIG. 2 and [Supplementary information S1](#) (figure) for class definitions and examples.

Most of these approaches provide support for the predicted widespread translation of short and alternative ORFs^{11–13,15,16,38,67,85,91} (TABLE 1). Nevertheless, even with ribosome profiling data, reliably identifying the full set of translated ORFs remains a challenge, especially in cases in which protein-coding sequences overlap.

Numerous complementary experimental approaches have aimed to further probe the degree to which newly predicted protein-coding sequences represent true cellular translation (TABLE 1). So far, these approaches generally confirm that the reads that are detected in regions predicted to be translated by ribosome profiling experiments represent translating 80S ribosomes. For example, ribosome footprints over putative translated sORFs tend to respond to translation inhibitors in a manner comparable to benchmarked translating ribosomes³⁹. Translated mRNA regions predicted from mouse ribosome profiling data immunoprecipitate with tagged 60S ribosomal subunits in a specific manner, similar to that seen for characterized translated ORFs³⁹. This finding suggests that true translating ribosomes produce the footprints that are detected by ribosome profiling over ORFs not previously annotated as being translated, rather than these mRNA fragments being artefacts resulting from the protection of mRNA by scanning translation initiation complexes or alternative RNA–protein complexes. An important open question is whether these translated regions produce stable peptides. Suggesting that they may, sORFs identified as being translated by ribosome profiling that have been carboxy-terminally tagged in yeast and in human cytomegalovirus (HCMV)-infected cells can be seen to accumulate in a regulated manner that mirrors predictions from ribosome profiling data^{11,39}. Meanwhile, specialized mass spectrometry approaches continue to identify a subset of peptides resulting from such sORFs in several systems^{11,92–97}, suggesting that at least some of these sORFs do encode abundant, stable peptides.

Most convincingly, a few sORFs that were predicted to be translated from polysome association and ribosome profiling data have now been shown to have biological function^{14,39,83}. In *D. melanogaster*, the peptides encoded by two such translated sORFs contained in the sarcolamban locus have been shown to directly bind to a calcium transporter in heart cells and thus regulate normal heart function⁸³ (FIG. 6a). In zebrafish, the short protein Toddler was found to drive gastrulation by functioning as a secreted developmental signal¹⁴. In mammals, a prominent example is the several translated sORFs, predicted on the basis of ribosome profiling of HCMV-infected human foreskin fibroblasts, that reside on the $\beta 2.7$ RNA, which has traditionally been defined as non-coding¹¹. Peptides resulting from the translation of two of these sORFs have been shown by mass spectrometry to accumulate during HCMV infection. In addition, analysis of serum samples from HCMV-positive and HCMV-negative blood bank donors showed a robust immune response to the peptides produced from several of these $\beta 2.7$ sORFs, specifically in the HCMV-positive individuals³⁹ (FIG. 6b). This result suggests that the ribosome-occupied sORFs are translated,

and that their products are processed and presented on MHC molecules as functional antigens in humans, thus expanding the range of epitopes displayed during viral infections. The condition-specific translation of many sORFs suggests that they could similarly be used to distinguish cancer cells from normal cells, with important implications for immunomodulatory therapies.

The translation of some sORFs could also help to fuel the evolution of new proteins⁸⁸. It is possible that transcriptional noise, together with the propensity of the ribosome to translate capped cytosolic RNAs, may allow novel transcripts to engage the ribosome and allow translational sampling of new, short motifs. Initially these sORFs may evolve under neutral selection. However, a subset could provide a small fitness advantage, resulting in positive selection and possible stabilization through lengthening over time, until they resemble canonical long protein-coding genes (FIG. 6c). Such regions would not necessarily be initially expected to show signatures of protein-coding conservation (as in REF. 98), and many might not produce a robust mutant phenotype when disrupted, making their study challenging.

How plastic is translation? Alternative isoforms abound.

The results of ribosome profiling in yeast and in mammals have indicated that many genes may yield two or more protein variants independently of splicing, which indicates that there may be surprising flexibility in both where translation starts and where it stops in eukaryotes. Such alternative isoforms have been seen and characterized previously; in budding yeast, for example, both alanyl-tRNA synthetase 1 (*ALA1*) and glycyl-tRNA synthetase 1 (*GRS1*) have been shown to exist in two isoforms, providing populations of the protein that are either cytosolic or mitochondrial, depending on the presence or absence of an N-terminal in-frame extension^{99,100}. These examples are also detected by ribosome profiling¹² and seem to be just a few of many^{10,12,67,101}, supporting a model in which diverse but targeted localization might be achieved for many proteins through sometimes small alterations in the site of translation initiation^{18,91,101}. Conversely, ribosome profiling of several yeast species, and of *D. melanogaster* embryos and cultured cells, revealed extensive heterogeneity in translation termination sites^{15,102,103}, resulting from regulated read-through of hundreds of genes. As with the N-terminal-extension isoforms, many of these C-terminal extensions are predicted to confer new subcellular localizations to the protein products^{15,104}.

Use of ribosome profiling has also facilitated the identification of interesting examples of regulated truncated protein isoforms^{10–12,89}. In human cells, a recent study identified a shortened alternative isoform of mitochondrial antiviral signalling protein (MAVS), which is an important player in innate immune signalling⁸⁹. The alternative MAVS isoform results from translation initiation downstream of the canonical start site to create an in-frame truncation, which the authors term ‘mini-MAVS’. Whereas full-length MAVS induces interferon production, miniMAVS antagonizes MAVS function by interfering with such production.

Signatures of protein-coding conservation

Purifying evolutionary selection results in higher levels of synonymous than nonsynonymous substitutions, specifically among homologous coding sequences. The pattern of nonsynonymous to synonymous differences among homologous regions compared in a phylogenetic group can be used to predict the likelihood that a genomic locus encodes a translated open reading frame (ORF).

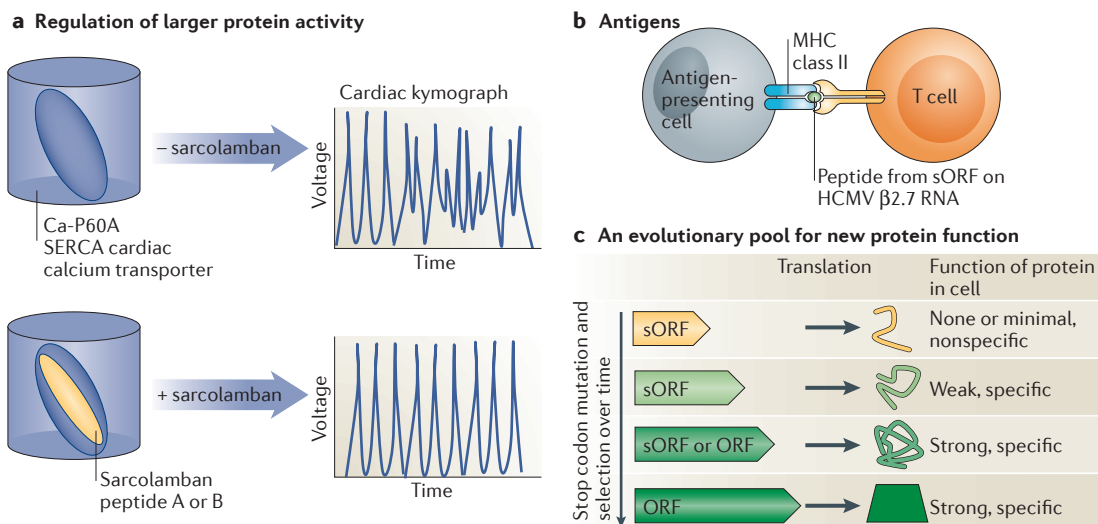


Figure 6 | Proposed cellular roles for the peptide products of translated short open reading frames (sORFs) identified by ribosome profiling. **a** | The two sarcolamban peptides are 28 and 29 amino acids in length, are conserved from fruit flies to human and regulate normal heart function in flies through direct binding to the sarcoendoplasmic reticulum Ca^{2+} ATPase (SERCA) calcium transporter Ca-P60A in cardiac tissue. **b** | Sera from human cytomegalovirus (HCMV)-positive blood donors were used to identify a specific and robust immune response against multiple short peptides translated from the $\beta 2.7$ RNA, which was previously thought to be a non-coding RNA. **c** | Spurious translation of short regions may produce a pool of peptides with weak or no cellular function. New protein domains may evolve through selection for maintenance of peptides with weak cellular function, followed by stop codon mutation and further selection for increasingly specific and important cellular function over time. Part **a** from Magny, E. G. *et al.* Conserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames. *Science* **341**, 1116–1120 (2013). Redrawn with permission from AAAS.

The large and diverse set of unconventional regions of translation suggested by ribosome profiling shows that there is considerably more to translational regulation and cellular content than was previously known. Some of these regions are likely to be translated into functional proteins, but it is likely that others will not produce stable protein products that are similar to those from traditional genes. Rather, subsets of these newly identified regions of translation may have regulatory, immune or currently neutral cellular roles. Unravelling the set of functions that are carried out by translated genomic regions poses a fascinating and daunting challenge.

Perspective

Protein synthesis consumes a large proportion of cellular resources and is central to almost every function of a cell. Ribosome profiling allows, for the first time, *in vivo* and global measurement of translation, providing a precise and quantitative account of what cells are translating, how this translation is regulated, and when and where translation happens. The rich and quantitative nature of ribosome profiling data provide an unprecedented opportunity to explore and model complex cellular processes.

Although it has long been known that translational regulation has important roles in development, in cellular responses to stimuli and in disease, the limited number of well-studied examples of regulation at the level of protein synthesis have generally been identified in an *ad hoc* manner. When paired with RNA-seq measurements of mRNA levels, ribosome profiling now allows

instantaneous measurement of all translational control in a given system, providing a tool for broad discovery of the underlying biology of a cellular process or state of choice. Furthermore, the detailed information that is yielded by this method provides valuable insight into fundamental aspects of how translation works. Despite the conserved nature of much of the translation machinery, important open questions about the mechanism of protein synthesis remain, including the basis for most specificity of translation among different mRNAs and the connections between translation and nascent protein folding.

Finally, owing to the precise genomic positional information provided by ribosome profiling, the protein-coding capacity of genomes can now be defined experimentally. This has led to the identification of a broad range of non-canonical translation events, including the translation of novel sORFs and alternative forms of previously annotated proteins, thereby challenging traditional views of protein-coding regions and gene diversity. Analytical advances that facilitate more comprehensive identification of other non-canonical translation events, such as those resulting from frame-shifting and stop codon read-through will continue to expand our understanding of the protein-coding capacity of complex genomes. The functions of the many novel short and alternatively translated regions that have been identified so far by ribosome profiling remain an intriguing and largely open question, the answer to which could fundamentally change the way we think about the encoding of information in genomes. Newly available CRISPR-based

methods¹⁰⁵ now make it possible to shut down the expression of any transcript^{106–109} or to introduce nonsense mutations into any ORF. As such, these approaches provide a central tool for efforts to define the functional roles of this broad array of newly identified translation products.

Specialized alterations to ribosome profiling that will advance its use in complex systems include the analysis of subsets of ribosomes, either those associated with

specific factors or protein modifications, or those in increasingly specific cell types or subcellular locations. Further transformative advances are likely to emerge from progressively more sophisticated and creative analysis of the rich data sets that are generated from ribosome profiling experiments, allowing major surprises to be revealed, even in systems that were thought to be well characterized.

1. McCann, K. L. & Baserga, S. J. Mysterious ribosomopathies. *Science* **341**, 849–850 (2013).
2. Cleary, J. D. & Ranum, L. P. W. Repeat-associated non-ATG (RAN) translation in neurological disease. *Hum. Mol. Genet.* **22**, R45–R51 (2013).
3. Ellis, S. R. Nucleolar stress in Diamond Blackfan anemia pathophysiology. *Biochim. Biophys. Acta* **1842**, 765–768 (2014).
4. Trainor, P. A. & Merrill, A. E. Ribosome biogenesis in skeletal development and the pathogenesis of skeletal disorders. *Biochim. Biophys. Acta* **1842**, 769–778 (2014).
5. Bolze, A. *et al.* Ribosomal protein SA haploinsufficiency in humans with isolated congenital asplenia. *Science* **340**, 976–978 (2013).
6. Kondrashov, N. *et al.* Ribosome-mediated specificity in *Hox* mRNA translation and vertebrate tissue patterning. *Cell* **145**, 383–397 (2011).
7. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S. & Weissman, J. S. Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223 (2009).
This work defines the ribosome profiling method and details its specificity, precision and utility.
8. Wolin, S. L. & Walter, P. Ribosome pausing and stacking during translation of a eukaryotic mRNA. *EMBO J.* **7**, 3559–3569 (1988).
9. Steitz, J. A. Polypeptide chain initiation: nucleotide sequences of the three ribosomal binding sites in bacteriophage R17 RNA. *Nature* **224**, 957–964 (1969).
10. 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).
11. Stern-Ginossar, N. *et al.* Decoding human cytomegalovirus. *Science* **338**, 1088–1093 (2012).
12. Brar, G. A. *et al.* High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science* **335**, 552–557 (2012).
References 11 and 12 describe the application of ribosome profiling to physiological dynamic cellular processes, the HCMV infection cycle in human cells and meiosis in budding yeast, respectively. In these disparate systems, both studies identified many new examples of translational control, uORF translation and the translation of many sORFs and alternative ORFs in genomic regions that were thought to be non-coding.
13. Bazzini, A. A. *et al.* Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO J.* **33**, 981–993 (2014).
14. Pauli, A. *et al.* Toddler: an embryonic signal that promotes cell movement via Apelin receptors. *Science* **343**, 1248636 (2014).
15. Dunn, J. G., Foo, C. K., Belletier, N. G., Gavis, E. R. & Weissman, J. S. Ribosome profiling reveals pervasive and regulated stop codon readthrough in *Drosophila melanogaster*. *eLife* **2**, e01179 (2013).
16. Aspden, J. L. *et al.* Extensive translation of small open reading frames revealed by Poly-Ribo-Seq. *eLife* **3**, e03528 (2014).
17. Smith, J. E. *et al.* Translation of small open reading frames within unannotated RNA transcripts in *Saccharomyces cerevisiae*. *Cell Rep.* **7**, 1858–1866 (2014).
18. Lee, S. *et al.* Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc. Natl Acad. Sci. USA* **109**, E2424–E2432 (2012).
19. Andreev, D. E. *et al.* Translation of 5' leaders is pervasive in genes resistant to eIF2 repression. *eLife* **4**, e03971 (2014).
20. Van Dijk, E. L., Auger, H., Jaszczyszyn, Y. & Thermes, C. Ten years of next-generation sequencing technology. *Trends Genet.* **30**, 418–426 (2014).
21. 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).
22. Kuersten, S., Radek, A., Vogel, C. & Penalva, L. O. Translation regulation gets its 'omics' moment. *Wiley Interdiscip. Rev. RNA* **4**, 617–630 (2013).
23. Oh, E. *et al.* Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor *in vivo*. *Cell* **147**, 1295–1308 (2011).
24. Arias, C. *et al.* KSHV 2.0: a comprehensive annotation of the Kaposi's sarcoma-associated herpesvirus genome using next-generation sequencing reveals novel genomic and functional features. *PLoS Pathog.* **10**, e1003847 (2014).
25. Stadler, M. & Fire, A. Wobble base-pairing slows *in vivo* translation elongation in metazoans. *RNA* **17**, 2063–2073 (2011).
26. Bazzini, A. A., Lee, M. T. & Giraldez, A. J. Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science* **336**, 233–237 (2012).
27. Juntawong, P., Girke, T., Bazin, J. & Bailey-Serres, J. Translational dynamics revealed by genome-wide profiling of ribosome footprints in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **111**, E203–E212 (2014).
28. Jensen, B. C. *et al.* Extensive stage-regulation of translation revealed by ribosome profiling of *Trypanosoma brucei*. *BMC Genomics* **15**, 911 (2014).
29. Caro, F., Ahlyong, V., Betegon, M. & DeRisi, J. L. Genome-wide regulatory dynamics of translation in the *Plasmodium falciparum* asexual blood stages. *eLife* **3**, e04106 (2014).
30. Schafer, S. *et al.* Translational regulation shapes the molecular landscape of complex disease phenotypes. *Nat. Commun.* **6**, 7200 (2015).
31. Rooijers, K., Loayza-Puch, F., Nijtmans, L. G. & Agami, R. Ribosome profiling reveals features of normal and disease-associated mitochondrial translation. *Nat. Commun.* **4**, 2886 (2013).
32. Zoschke, R., Watkins, K. P. & Barkan, A. A rapid ribosome profiling method elucidates chloroplast ribosome behavior *in vivo*. *Plant Cell* **25**, 2265–2275 (2013).
33. Michel, A. M. *et al.* GWIPS-viz: development of a ribosome genome browser. *Nucleic Acids Res.* **42**, D859–D864 (2014).
34. Liu, X., Jiang, H., Gu, Z. & Roberts, J. W. High-resolution view of bacteriophage lambda gene expression by ribosome profiling. *Proc. Natl Acad. Sci. USA* **110**, 11928–11933 (2013).
35. Gerashchenko, M. V., Lobanov, A. V. & Gladyshev, V. N. Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress. *Proc. Natl Acad. Sci. USA* **109**, 17394–17399 (2012).
36. Li, G.-W., Oh, E. & Weissman, J. S. The anti-Shine–Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* **484**, 538–541 (2012).
37. Woolstenhulme, C. J., Guydosh, N. R., Green, R. & Buskirk, A. R. High-precision analysis of translational pausing by ribosome profiling in bacteria lacking EFP. *Cell Rep.* **11**, 13–21 (2015).
38. Chew, G.-L. *et al.* Ribosome profiling reveals resemblance between long non-coding RNAs and 5' leaders of coding RNAs. *Development* **140**, 2828–2834 (2013).
39. Ingolia, N. T. *et al.* Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell Rep.* **8**, 1365–1379 (2014).
40. Andreev, D. E. *et al.* Oxygen and glucose deprivation induces widespread alterations in mRNA translation within 20 minutes. *Genome Biol.* **16**, 90 (2015).
41. Guydosh, N. R. & Green, R. Dom34 rescues ribosomes in 3' untranslated regions. *Cell* **156**, 950–962 (2014).
42. Shalgi, R. *et al.* Widespread regulation of translation by elongation pausing in heat shock. *Mol. Cell* **49**, 439–452 (2013).
43. Han, Y. *et al.* Ribosome profiling reveals sequence-independent post-initiation pausing as a signature of translation. *Cell Res.* **24**, 842–851 (2014).
44. Liu, B., Han, Y. & Qian, S.-B. Cotranslational response to proteotoxic stress by elongation pausing of ribosomes. *Mol. Cell* **49**, 453–463 (2013).
45. Subramaniam, A. R., Zid, B. M. & O'Shea, E. K. An integrated approach reveals regulatory controls on bacterial translation elongation. *Cell* **159**, 1200–1211 (2014).
This work probes position-specific changes in ribosome distribution among various cellular conditions, concluding that tRNA abundances do not account for elongation rates for most codons, and that pausing of ribosomes during starvation may result in translation abortion.
46. Lareau, L. F., Hite, D. H., Hogan, G. J. & Brown, P. O. Distinct stages of the translation elongation cycle revealed by sequencing ribosome-protected mRNA fragments. *eLife* **3**, e01257 (2014).
This work identifies a class of short ribosome footprints that may be enriched by treatment with translation elongation inhibitors and that are likely to represent a distinct conformation of the ribosome at a specific stage of the elongation cycle.
47. Siegel, A. F., van den Engh, G., Hood, L., Trask, B. & Roach, J. C. Modeling the feasibility of whole genome shotgun sequencing using a pairwise end strategy. *Genomics* **68**, 237–246 (2000).
48. Roberts, A., Schaeffer, L. & Pachter, L. Updating RNA-Seq analyses after re-annotation. *Bioinformatics* **29**, 1631–1637 (2013).
49. Saliba, A.-E., Westermann, A. J., Gorski, S. A. & Vogel, J. Single-cell RNA-seq: advances and future challenges. *Nucleic Acids Res.* **42**, 8845–8860 (2014).
50. Green & R. & Noller, H. F. Ribosomes and translation. *Annu. Rev. Biochem.* **66**, 679–716 (1997).
51. Li, G.-W., Burkhardt, D., Gross, C. & Weissman, J. S. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* **157**, 624–635 (2014).
52. Aris, J. P., Klionsky, D. J. & Simoni, R. D. The F₁ subunits of the *Escherichia coli* F₁F₀-ATP synthase are sufficient to form a functional proton pore. *J. Biol. Chem.* **260**, 11207–11215 (1985).
53. Humphries, N. *et al.* The Ecm11–Gmc2 complex promotes synaptonemal complex formation through assembly of transverse filaments in budding yeast. *PLoS Genet.* **9**, e1003194 (2013).
54. Lee, M. T. *et al.* Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. *Nature* **503**, 360–364 (2013).
55. Kronja, I. *et al.* Widespread changes in the posttranscriptional landscape at the *Drosophila* oocyte-to-embryo transition. *Cell Rep.* **7**, 1495–1508 (2014).
56. Vasquez, J.-J., Hon, C.-C., Vanselow, J. T., Schlosser, A. & Siegel, T. N. Comparative ribosome profiling reveals extensive translational complexity in different *Trypanosoma brucei* life cycle stages. *Nucleic Acids Res.* **42**, 3623–3637 (2014).

57. Stumpf, C. R., Moreno, M. V., Olshen, A. B., Taylor, B. S. & Ruggero, D. The translational landscape of the mammalian cell cycle. *Mol. Cell* **52**, 574–582 (2013).
58. Stadler, M. & Fire, A. Conserved translatome remodeling in nematode species executing a shared developmental transition. *PLoS Genet.* **9**, e1003739 (2013).
59. Liu, B. & Qian, S.-B. Translational reprogramming in cellular stress response. *Wiley Interdiscip. Rev. RNA* **5**, 301–305 (2014).
60. Ingolia, N. T. Ribosome profiling: new views of translation, from single codons to genome scale. *Nat. Rev. Genet.* **15**, 205–213 (2014).
61. Michel, A. M. & Baranov, P. V. Ribosome profiling: a Hi-Def monitor for protein synthesis at the genome-wide scale. *Wiley Interdiscip. Rev. RNA* **4**, 473–490 (2013).
62. Cox, J. S. & Walter, P. A novel mechanism for regulating activity of a transcription factor that controls the unfolded protein response. *Cell* **87**, 391–404 (1996).
63. Kannan, K. *et al.* The general mode of translation inhibition by macrolide antibiotics. *Proc. Natl Acad. Sci. USA* **111**, 15958–15963 (2014).
64. Kannan, K., Vázquez-Laslop, N. & Mankin, A. S. Selective protein synthesis by ribosomes with a drug-obstructed exit tunnel. *Cell* **151**, 508–520 (2012).
65. Davis, A. R., Gohara, D. W. & Yap, M.-N. F. Sequence selectivity of macrolide-induced translational attenuation. *Proc. Natl Acad. Sci. USA* **111**, 15379–15384 (2014).
66. Jung, H., Gkogkas, C. G., Sonenberg, N. & Holt, C. E. Remote control of gene function by local translation. *Cell* **157**, 26–40 (2014).
67. Williams, C. C., Jan, C. H. & Weissman, J. S. Targeting and plasticity of mitochondrial proteins revealed by proximity-specific ribosome profiling. *Science* **346**, 748–751 (2014).
68. Jan, C. H., Williams, C. C. & Weissman, J. S. Principles of ER cotranslational translocation revealed by proximity-specific ribosome profiling. *Science* **346**, 1257521 (2014).
69. Arava, Y. *et al.* Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **100**, 3889–3894 (2003).
70. Heiman, M., Kulicke, R., Fenster, R. J., Greengard, P. & Heintz, N. Cell type-specific mRNA purification by translating ribosome affinity purification (TRAP). *Nat. Protoc.* **9**, 1282–1291 (2014).
71. Heiman, M. *et al.* A translational profiling approach for the molecular characterization of CNS cell types. *Cell* **135**, 738–748 (2008).
72. Inada, T. *et al.* One-step affinity purification of the yeast ribosome and its associated proteins and mRNAs. *RNA* **8**, 948–958 (2002).
73. Zanetti, M. E., Chang, I.-F., Gong, F., Galbraith, D. W. & Bailey-Serres, J. Immunopurification of polyribosomal complexes of *Arabidopsis* for global analysis of gene expression. *Plant Physiol.* **138**, 624–635 (2005).
74. Mustroph, A., Zanetti, M. E., Girke, T. & Bailey-Serres, J. Isolation and analysis of mRNAs from specific cell types of plants by ribosome immunopurification. *Methods Mol. Biol.* **959**, 277–302 (2013).
75. Thomas, A. *et al.* A versatile method for cell-specific profiling of translated mRNAs in *Drosophila*. *PLoS ONE* **7**, e40276 (2012).
76. Housley, M. P. *et al.* Translational profiling through biotinylation of tagged ribosomes in zebrafish. *Development* **141**, 3988–3993 (2014).
77. Reinhardt, J. A. & Jones, C. D. Two rapidly evolving genes contribute to male fitness in *Drosophila*. *J. Mol. Evol.* **77**, 246–259 (2013).
78. Starck, S. R. *et al.* Leucine-tRNA initiates at CUG start codons for protein synthesis and presentation by MHC class I. *Science* **336**, 1719–1723 (2012).
79. Rebbapragada, I. & Lykke-Andersen, J. Execution of nonsense-mediated mRNA decay: what defines a substrate? *Curr. Opin. Cell Biol.* **21**, 394–402 (2009).
80. Pauli, A., Valen, E. & Schier, A. F. Identifying (non-) coding RNAs and small peptides: challenges and opportunities. *BioEssays* **37**, 103–112 (2015).
81. Ulitsky, I. & Bartel, D. P. lincRNAs: genomics, evolution, and mechanisms. *Cell* **154**, 26–46 (2013).
82. Kondo, T. *et al.* Small peptides switch the transcriptional activity of Shavenbaby during *Drosophila* embryogenesis. *Science* **329**, 336–339 (2010). **This paper identifies key roles in fly development for several short peptides (from 11 to 32 amino acids) translated from sORFs on a transcript that was previously thought to be non-coding.**
83. Magny, E. G. *et al.* Conserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames. *Science* **341**, 1116–1120 (2013).
84. Anderson, D. M. *et al.* A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell* **160**, 595–606 (2015).
85. Ruiz-Orera, J., Messegue, X., Subirana, J. A. & Alba, M. M. Long non-coding RNAs as a source of new peptides. *eLife* **3**, e03523 (2014).
86. Xu, Y. & Ganem, D. Making sense of antisense: seemingly noncoding RNAs antisense to the master regulator of Kaposi's sarcoma-associated herpesvirus lytic replication do not regulate that transcript but serve as mRNAs encoding small peptides. *J. Virol.* **84**, 5465–5475 (2010).
87. Guttman, M., Russell, P., Ingolia, N. T., Weissman, J. S. & Lander, E. S. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* **154**, 240–251 (2013).
88. Carvunis, A.-R. *et al.* Proto-genes and *de novo* gene birth. *Nature* **487**, 370–374 (2012). **In this work, the authors present evidence to support the protogene hypothesis, according to which new proteins can evolve through the selection and elongation of ORFs encoding peptides translated from putative intergenic transcripts.**
89. Brubaker, S. W., Gauthier, A. E., Mills, E. W., Ingolia, N. T. & Kagan, J. C. A bicistronic MAVS transcript highlights a class of truncated variants in antiviral immunity. *Cell* **156**, 800–811 (2014).
90. Michel, A. M. *et al.* Observation of dually decoded regions of the human genome using ribosome profiling data. *Genome Res.* **22**, 2219–2229 (2012).
91. Noderer, W. L. *et al.* Quantitative analysis of mammalian translation initiation sites by FACS-seq. *Mol. Syst. Biol.* **10**, 748 (2014).
92. Schwaib, A. G. *et al.* Chemoproteomic discovery of cysteine-containing human short open reading frames. *J. Am. Chem. Soc.* **135**, 16750–16753 (2013).
93. Slavoff, S. A. *et al.* Peptidomic discovery of short open reading frame-encoded peptides in human cells. *Nat. Chem. Biol.* **9**, 59–64 (2013).
94. Ma, J. *et al.* Discovery of human sORF-encoded polypeptides (SEPs) in cell lines and tissue. *J. Proteome Res.* **13**, 1757–1765 (2014).
95. Crappé, J. *et al.* Combining *in silico* prediction and ribosome profiling in a genome-wide search for novel putatively coding sORFs. *BMC Genomics* **14**, 648 (2013).
96. Menschaert, G. *et al.* Deep proteome coverage based on ribosome profiling aids mass spectrometry-based protein and peptide discovery and provides evidence of alternative translation products and near-cognate translation initiation events. *Mol. Cell. Proteomics* **12**, 1780–1790 (2013).
97. Vanderperre, B. *et al.* Direct detection of alternative open reading frames translation products in human significantly expands the proteome. *PLoS ONE* **8**, e70698 (2013).
98. Lin, M. F., Jungreis, I. & Kellis, M. PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics* **27**, i275–i282 (2011).
99. Huang, H.-Y., Tang, H.-L., Chao, H.-Y., Yeh, L.-S. & Wang, C.-C. An unusual pattern of protein expression and localization of yeast alanyl-tRNA synthetase isoforms. *Mol. Microbiol.* **60**, 189–198 (2006).
100. Chang, K.-J. & Wang, C.-C. Translation initiation from a naturally occurring non-AUG codon in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **279**, 13778–13785 (2004).
101. Wan, J. & Qian, S.-B. TISdb: a database for alternative translation initiation in mammalian cells. *Nucleic Acids Res.* **42**, D845–D850 (2014). **This work presents a database of alternative translation initiation sites that have been identified by ribosome profiling in mammalian cells.**
102. Artieri, C. G. & Fraser, H. B. Evolution at two levels of gene expression in yeast. *Genome Res.* **24**, 411–421 (2014).
103. Jungreis, I. *et al.* Evidence of abundant stop codon readthrough in *Drosophila* and other metazoa. *Genome Res.* **21**, 2096–2113 (2011).
104. Schueren, F. *et al.* Peroxisomal lactate dehydrogenase is generated by translational readthrough in mammals. *eLife* **3**, e03640 (2014).
105. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
106. Zalatan, J. G. *et al.* Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* **160**, 339–350 (2015).
107. Gilbert, L. A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451 (2013).
108. Gilbert, L. A. *et al.* Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* **159**, 647–661 (2014).
109. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).

Acknowledgements

The authors wish to thank C. Jan and E. Ünal for helpful comments on this manuscript and N. Ingolia for development of the original ribosome profiling protocol and helpful discussions. This work was partially supported by the Winkler Family Biological Sciences Award to G.A.B. and Howard Hughes Medical Institute and Center for RNA Systems Biology funding to J.S.W.

Competing interests statement

The authors declare **competing interests**: see Web version for details.

SUPPLEMENTARY INFORMATION

See online article: S1 (figure)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF