WIREs Wiley Interdisciplinary Reviews

Article Title: Computational methods for ribosome profiling data analysis

Article Type: Focus Article

Authors:

Stephen J. Kiniry
[School of Biochemistry and Cell Biology, University College Cork, Cork, Ireland.
Stephen.kiniry@ucc.ie]

Audrey M. Michel
[School of Biochemistry and Cell Biology, University College Cork, Cork, Ireland.
A.Michel@ucc.ie]

Pavel V. Baranov
[School of Biochemistry and Cell Biology, University College Cork, Cork, Ireland.
P.baranov@ucc.ie]
[Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia.]

Abstract

Since the introduction of the ribosome profiling technique in 2009 its popularity has greatly increased. It is widely used for the comprehensive assessment of gene expression and for studying the mechanisms of regulation at the translational level. As the number of ribosome profiling datasets being produced continues to grow, so too does the need for reliable software that can provide answers to the biological questions it can address. This review describes the computational methods and tools that have been developed to analyse ribosome profiling data at the different stages of the process. It starts with initial routine processing of raw data and follows with more specific tasks such as the identification of translated open reading frames, differential gene expression analysis, or evaluation of local or global codon decoding rates. The review pinpoints challenges associated with each step and explains the ways in which they are currently addressed. In addition it provides a comprehensive, albeit incomplete, list of publicly available software applicable to each step, which may be a beneficial starting point to those unexposed to ribosome profiling analysis. The outline of current challenges in ribosome profiling data analysis may inspire computational biologists to search for novel, potentially superior, solutions that will improve and expand the bioinformaticians' toolbox for ribosome profiling data analysis.

Graphical/Visual Abstract and Caption

Introduction

Ribosome profiling or Ribo-Seq, involves the arrest of translating ribosomes (using translation inhibitors or other methods) as they traverse mRNA (Ingolia, Ghaemmaghami, Newman, & Weissman, 2009). A nuclease is then

used to break down any section of mRNA not being protected by a ribosome (Figure 1a). The remaining protected fragments of mRNA (footprints) can then be isolated, sequenced, and mapped to a reference transcriptome or genome. These footprints are approximately 30 nucleotides in length and when mapped can provide both quantitative as well as qualitative information on translation, see (Andreev et al., 2017; Brar & Weissman, 2015; Ingolia et al., 2014; Ingolia, Hussmann, & Weissman, 2018; Michel & Baranov, 2013) for reviews. Common applications of Ribo-Seq data analysis include translated Open Reading Frame (ORF) detection, ribosome stalling/pause site detection, and differential gene expression analysis (Figure 1b).

The detection of translated regions of a genome is a task for which ribosome profiling is particularly well suited. Translation can be identified even at ORFs consisting of only a start and a stop codon (Tanaka et al., 2016). Depending on the dataset this can be achieved at sub-codon resolution, meaning that even overlapping translated open reading frames (ORFs) can be detected (Michel et al., 2012). Even though the human genome has been sequenced a while ago, novel protein coding ORFs continue to be discovered, e.g. an upstream ORF (uORF) in the human *MIEF1* gene was predicted to code for a protein (Andreev, O'Connor, Fahey, et al., 2015) and was later found to be an assembly factor of mitochondrial ribosomes (Brown et al., 2017) and more recently characterized as the main product of *MIEF1* mRNA (Rathore et al., 2018).

Ribosome stalling/pause sites can also be characterized. A ribosome moving along an mRNA can pause or stall, blocking the path of other ribosomes, and thus regulate protein synthesis (Ivanov et al., 2018; Kurian, Palanimurugan, Godderz, & Dohmen, 2011; Yordanova et al., 2018) or trigger No-Go decay or Ribosome Quality Control pathways, see (Brandman & Hegde, 2016; Buskirk & Green, 2017; Inada, 2013) for reviews. Since ribosomes are more likely to occupy pause sites, more footprints are produced from these locations. Thus, the pause sites appear as local peaks of ribosome footprint density and can be detected computationally.

Another popular (though not unique) application of ribosome profiling is the quantitative characterization of differential gene expression, as it discriminates changes in mRNA translation from changes in mRNA levels. Translation regulation can also be assessed with polysome profiling where the levels of mRNA found in heavy polysome fractions are compared with total mRNA levels. The Ribo-Seq advantage over polysome profiling is that it provides information on the translation of a specific ORF (or ORFs) within an mRNA, however it has its own limitations, see (Gandin et al., 2016) for a comparison of the two approaches. Since ribosome profiling generates millions of sequencing reads the processing and analysis of the data requires intensive computation. The signal produced with ribosome profiling is far more complex and richer in potential applications than standard RNA-seq. Numerous computational approaches have been developed, see (Calviello & Ohler, 2017) for review. We structured this review by detailing the steps carried out for ribosome profiling data analysis and specific goals and overview software that has been developed for these tasks. The accession information for the software tools and/or its sources are provided in tables that are separated into categories. Many tools are multifunctional and could be placed in more than one category while some tools are unique. The selection of software for this review is based on published literature rather than on usability, since testing and benchmarking all published software is an onerous task that should be carried out separately.

Technical considerations when processing raw sequencing reads

Raw ribosome profiling data are usually single-end unprocessed sequencing reads in FASTQ format that need to be processed and mapped to a reference genome or transcriptome. Processing of the reads typically involves removal of adapter/linker sequences as well as removal of any reads aligning to ribosomal RNA (rRNA) and/or transfer RNA (tRNA). There are many freely available tools for both removing adapters and aligning short reads. For example, cutadapt (Martin, 2011) is commonly used to remove adapters, while bowtie (Langmead, Trapnell, Pop, & Salzberg, 2009) and STAR (Dobin et al., 2013) are commonly used for alignment. As these are not specific to ribosome profiling, they will not be discussed in detail here. However, there are several variable parameters involved in both processing and mapping which may significantly affect downstream analysis. Therefore, the initial read processing and alignment should be guided by how the data will be utilised downstream.

To reduce the mapping of non ribosome protected fragments, footprints whose lengths are below a certain threshold are usually discarded. This is done under the assumption that such shorter reads consist of RNA fragments other than those protected by the ribosome or of over-digested footprints. However, such length filtering needs to be applied with caution, because the length of footprints may depend on their sequence and

location, e.g. in bacteria, footprints derived from ribosomes bound to Shine Dalgarno sequences are longer (O'Connor, Li, Weissman, Atkins, & Baranov, 2013). Indeed, Allen Buskirk and colleagues have provided strong evidence suggesting that the earlier claim that Shine-Dalgarno sequences cause ribosome pauses in bacteria (G. W. Li, Oh, & Weissman, 2012) may be an artefact of the footprint length selection (Mohammad, Woolstenhulme, Green, & Buskirk, 2016). It is also important to note that the length of footprints varies considerably across datasets. Most ribosome footprints in eukaryotes are approximately 28-30 nucleotides and this corresponds to the length of mRNA fragments protected by the ribosome in a specific conformation when its A-site is occupied with a tRNA. Such a conformation is stabilized by certain translation inhibitors that bind to the E-site which is empty in the pretranslocational ribosome conformation. This includes cycloheximide which is by far the most widely used inhibitor in ribosome profiling studies. However, in a posttranslocational conformation, when the A-site is unoccupied, eukaryotic ribosomes protect shorter (20-22 nucleotides) fragments and such fragments could become predominant if different inhibitors are used, such as anisomycin which inhibits the peptidyl transferase reaction (Lareau, Hite, Hogan, & Brown, 2014; C. C. Wu, Zinshteyn, Wehner, & Green, 2019). Scanning ribosomes also leave footprints of varying length depending on their specific conformations (Archer, Shirokikh, Beilharz, & Preiss, 2016). The heterogeneity of ribosome footprint lengths is further exacerbated by suboptimal nuclease digestion which may lead to over or under-digestion of footprints.

The mapping of ribosome footprints to genomic sequences poses yet another problem, namely the mapping across exon-exon junctions. While this is an issue for most techniques involving the sequencing of RNA, it is particularly acute for ribosome profiling due to the short length of Ribo-Seq reads. This results in a systematic bias manifested in reduced unambiguous mappings at exon-exon junctions. This can be clearly seen in the GWIPS-viz browser multiregion view (Kiniry, Michel, & Baranov, 2018; Michel, Fox, et al., 2014). There are splice-aware aligners that are capable of mapping across exon-exon junctions, but the short length of ribosome footprints increases the chance of spurious mappings. One solution to this is to simply map ribosome footprints to transcriptome sequences, but this may not be desired when the accuracy of the transcriptome is in doubt or when its completeness is critical for downstream analysis, e.g. during the identification of novel translated regions.

PCR amplification of footprints during cDNA library generation is also a potential problem. While this bias is pertinent to many techniques requiring PCR amplification, it could be particularly acute for certain Ribo-Seq applications which rely on the accuracy of local footprint density measurements such as detection of ribosome pauses or estimation of codon decoding rates. Recent studies have started to solve this issue with the use of random barcodes introduced to cDNA during the first round of RT PCR reaction. Such barcodes are termed Unique Molecular Identifiers (UMI) and have been used in many applications (Islam et al., 2014; Kivioja et al., 2011). To our knowledge UMIs were first introduced to ribosome profiling by Miettinen and Björklund (Miettinen & Bjorklund, 2015) and now are part of standard ribosome profiling protocol (McGlincy & Ingolia, 2017). During data processing, reads with the same sequence that also share the same UMI are considered to be PCR duplicates and counted as one. This can be done with specific software such as UMI tools (Smith, Heger, & Sudbery, 2017). As the use of UMIs in Ribo-Seq studies is still relatively recent, it is difficult to assess how much of a problem is PCR duplication, though some studies suggest that with sufficient input material and low number of PCR cycles, PCR duplicates constitute only a small fraction of sequencing reads in Ribo-Seq data (Lecanda et al., 2016; McGlincy & Ingolia, 2017).

Accession information for software pipelines that can be used for data processing can be found in Table 1, however, certain software packages described later also contain pipelines for raw data processing and quality assessment.

Global assessment of the data quality

Assessing the quality of the data should be viewed as an obligatory requirement after initial pre-processing and mapping, as it saves wasted time trying to draw conclusions from poor quality data. Four relatively simple approaches are commonly utilised to achieve this; analysis of read length distributions, metagene profiles, a breakdown of regions to which Ribo-Seq reads align, and the triplet periodicity signal. Other more general approaches include assessing the correlation among replicates and the number of useful mapped reads. The implementation of these features have become "*de facto*" best practice and while they are indicative of quality they should not be viewed as definitive.

A typical ribosome profiling dataset obtained from eukaryotic cells is characterized by a sharp distribution of lengths with a predominant length around 28-30 nucleotides. The variation depends on the nuclease digestion conditions and the inhibitors used (see above). The distribution is wider for ribosome profiling datasets obtained from bacterial cells due to the read length distribution associated with Shine-Dalgarno interactions (O'Connor et al., 2013). The read length distribution can be analysed with a number of tools, for example, FastQC (Andrews, 2010), a general tool for assessing the sequence quality of reads obtained with high throughput sequencing. FastQC also can be used to evaluate the accuracy of base calls and to quantify positional nucleotide frequencies, GC content and over-represented sequences. These analyses can often uncover problematic features such as the frequent addition of untemplated nucleotides during reverse transcription, untrimmed adapter sequences, etc.

Another important way to assess the quality of the datasets is with a *metagene profile*. The metagene profile provides the frequency of footprints relative to all annotated start and stop codons. There are several ways to generate metagene profiles. One is to simply count the frequency of all footprints (using a single footprint position, i.e. the 5' or 3' end) at a specific coordinate relative to the annotated start codons (or stops) of all transcripts. The procedure for building a metagene profile relative to start codons could be represented as $D(i)=\Sigma_K(d_k(i+s_k))$, where D is a metagene footprint density, i is the coordinate of metagene profile, d_k is a footprint density at a transcript k from transcriptome K with s_k being the coordinate of the annotated start. A potential issue with such a representation is that highly expressed mRNAs could dramatically skew the metagene profile. To mitigate this issue, the frequency of footprints could be normalized across individual mRNAs, so that they have equal influence on the overall picture. It is also possible to normalize the CDS length and analyse the frequency of footprints of different lengths, producing a very informative translatome representation as has been done by Thomas Preiss and colleagues (Archer et al., 2016). A metagene profile of a high-quality ribosome profiling dataset is expected to have a sharp difference in footprint density at the start and stop codons, so that the density is higher downstream of starts and upstream of stops (Figure 2). For the generation of metagene profiles in bacteria it is important to exclude overlapping CDS regions as well as closely located CDS regions to avoid signal interference. In a similar vein, the generation of metagene profiles in higher eukaryotes necessitates selecting a single transcript isoform where multiple isoforms exist to avoid an artificial amplification of footprints counts by the number of splice isoforms. Ideally the translated transcript isoform(s) at a gene locus should be used. However, metagene profile generation is typically carried out early in the Ribo-Seq data analysis process and isoform delineation, if required, performed further downstream. Hence heuristic approaches are often used such as selecting "principal isoforms" from the APPRIS database (Rodriguez et al., 2018). Other heuristic approaches of a single representative transcript selection and their limitations are discussed later in relation to differential expression analysis.

Triplet periodicity refers to the unequal distribution of read mappings relative to subcodon positions due to the triplet nature of the genetic code: elongating ribosomes move along mRNA in discrete steps of three nucleotides. The strength of the triplet periodicity can be assessed using the frequency with which a single footprint coordinate (e.g. 5' or 3' end) aligns to one of the three subcodon positions. See (Figure 3 (a,c)) for an example of a dataset with strong periodicity derived from (Calviello et al., 2016) and (Figure 3 (b,d)) for an example of a dataset with poor periodicity (Kirchner et al., 2017). Strong periodicity is a good indicator that the data is genuinely Ribo-Seq data and it can be used for the detection of translated reading frames (Michel et al., 2012), although it is not definitive as even RNA-seq data may indicate some periodicity due to crossover of sequencing biases and GC3 skew. However, the periodicity is dependent on the uniformity of the digestion position relative to the ribosome's decoding centre and thus varies depending on digestion conditions (Gerashchenko & Gladyshev, 2017) and specifics of the translation apparatus (see above). Thus, the absence of strong periodicity does not necessarily mean that the other useful features of the data are also poor. One way to express the periodicity quantitatively is to calculate the proportion of reads at the predominant subcodon position. Another is to assess the divergence from an equiprobable distribution using Shannon Entropy ($-\Sigma p_i log_2(p_i)$ where p_i is the relative frequency of footprints at the i subcodon position). The periodicity can also be detected with Fourier (Calviello et al., 2016; Chun, Rodriguez, Todd, & Mills, 2016) and wavelets transformations (Xu et al., 2018a). Both metagene profiles and triplet periodicity visualization plots can be produced by many different tools such as RibostreamR (Perkins, Mazzoni-Putman, Stepanova, Alonso, & Heber, 2019), Ribo-SeQC (Calviello, Sydow, Harnett, & Ohler, 2019), RiboGalaxy (Michel et al., 2016), Plastid (Dunn & Weissman, 2016), riboseqR (Chung et al., 2015) and mQC (Verbruggen & Menschaert, 2018).

Determining the position of the decoding center

An offset is typically applied to the sequence of ribosome footprints to infer the position of the A- or P-site of the ribosome that produced it. This is an integer which is added to the coordinate of the 5' end of a mapped read or, alternatively subtracted from the coordinate of its 3' end. The metagene profiles are often used to determine the offset, assuming that the first sharp increase in footprint density corresponds to the footprints of the ribosomes at the start codons. Since the start codons are recognised at the P-site, the distance between this increase and the first nucleotide of the start codon is used as the offset for determining positions of the P-sites, see (Figure 2(a)) for a metagene profile made using data from (Calviello et al., 2016). To determine the positions of the A-sites, 3 nucleotides are added if the metagene profile is based on the 5’ ends or subtracted if it is based on the 3’ ends. Typically, when reads are not stratified by read lengths, 5’ end mappings produce a greater triplet periodicity in eukaryotic organisms, while 3’ ends produce greater periodicity in bacteria (Woolstenhulme, Guydosh, Green, & Buskirk, 2015). This is most likely due to the asymmetric variability of read lengths relative to the decoding centre which in case of bacteria could be attributed to Shine-Dalgarno interactions with anti-Shine-Dalgarno (O'Connor et al., 2013). Applying a 'static' offset regardless of read length is often sufficient to determine positions of A- or P-sites with an accuracy that is satisfactory for numerous Ribo-Seq applications. However, the accurate determination of A or P-site positions is critical for certain applications such as the measurement of ribosome dwell times at specific codons (e.g. estimating codon decoding rates). The accuracy can be further improved with setting specific offsets for each read length, i.e. using separate metagene profiles made for each read length, see (Figure 2(c)) for a heatmap made using data from (Albert, Muzzey, Weissman, & Kruglyak, 2014).

This approach can provide more accurate inferred A- or P-site locations than a static offset and thus improve the periodicity signal. However, in any given dataset there may be read lengths that are not abundant compared to the predominant read length. These low-abundance read lengths are difficult to correctly assign an offset to. RiboWaltz (Lauria et al., 2018) aims to correct this by using offset values from abundant read lengths to infer the optimal offsets for less abundant read lengths. More sophisticated methods of offset determination have also been developed, (O'Connor, Andreev, & Baranov, 2016) proposed the determination of the offset that maximises the difference of the estimated dwell time between codons. This assumes that the A-site has a predominant role in influencing the decoding rate. Ribodeblur (Wang, McManus, & Kingsford, 2017), uses an expectation maximization-like procedure to obtain a more accurate estimate of A-sites. RiboproP (Zhao, Baez, Fredrick, & Bundschuh, 2018) is specifically designed to mitigate the sequence bias introduced from Ribo-Seq data generated with MNase , thus improving offsetting. See Table 1 for accession information to these tools.

Translated ORF detection

The detection of translated ORFs is an application for which ribosome profiling is uniquely well suited, particularly of short ORFs, whose products cannot be easily detected with proteomics techniques. Detecting translation using Ribo-Seq data is not straight forward as the presence of a footprint in a given genomic region does not necessarily mean that that region is being translated. In addition to the artefacts of mapping mentioned previously, not all sequences found in a ribosome profiling cDNA library derive from genuine ribosome protected fragments within the ribosome mRNA channel. In fact, most of the cDNA reads in any ribosome profiling library come from the ribosome itself as its rRNA gets digested during the procedure. Similarly fragments of other RNAs bound to the ribosome could contaminate the sample (fragments of tRNAs are also very abundant). Additional sources of contamination are fragments of RNAs from nucleoprotein complexes that could be co-isolated with ribosomal complexes. Thus, the difficult aspect of translated ORF detection is the discrimination of the signal obtained with genuine ribosome footprints from other RNA fragments.

Nonetheless potentially translated regions can often be easily recognized upon manual visual inspection of the corresponding sequence region. Several existing resources provide such functionality such as Svist4get (Egorov et al., 2019) SmProt (Hao et al., 2018), GWIPS-Viz (Michel, Kiniry, O'Connor, Mullan, & Baranov, 2018), Trips-Viz (Kiniry, O'Connor, Michel, & Baranov, 2018) HRPDViewer (W. S. Wu et al., 2018) and RiboViz (Carja, Xing, Wallace, Plotkin, & Shah, 2017). Many allow for viewing Ribo-Seq datasets from multiple studies simultaneously which can significantly boost the signal to noise ratio making translated regions easier to detect. Manual visual

detection is a simple and straightforward method of translated ORF detection, particularly when the translated ORF is highly translated and does not overlap with others. However, when several ORFs overlap or are nested within each other, their detection based purely on the density of footprints is difficult due to the heterogeneity of the signal within an ORF. Manual visual detection in these cases can be improved when footprints are discriminated based on the phase of their triplet periodicity. This could be done either by generating separate subcodon profiles or using differential colors for the reads depending on their phase relative to subcodon positions as in RiboSeqR (Chung et al., 2015), RiboGalaxy (Michel et al., 2016) or Trips-Viz (Kiniry, O'Connor, et al., 2018), see Table 2. The main disadvantage of manual identification of translated ORFs is the low throughput. Manual inspection of even a bacterial genome is impractical. Thus, numerous tools have been developed to enable automatic high throughput detection of translated ORFs using Ribo-Seq data.

They utilize different computational concepts including statistical tests as in Ribo-TISH (P. Zhang et al., 2017), linear regression as in ORF-RATER (Fields et al., 2015), robustness of triplet periodicity as in RiboTaper (Calviello et al., 2016) and RiboWave (Xu et al., 2018a), Hidden Markov models as in RiboHMM (Raj et al., 2016) as well as machine learning techniques, e. g. in REPARATION for bacterial genome reannotations (Ndah et al., 2017). An in-depth analysis of these approaches requires a separate dedicated review. Further we will summarize the most common features of translated ORFs that are often used by these tools to predict their translation.

The similarity between the patterns of footprints in mRNA 5’ leaders and lincRNAs observed in early mammalian datasets provoked a suggestion that translation takes place in RNA transcripts (and their parts) that were normally considered non-coding (Chew et al., 2013; Ingolia, Lareau, & Weissman, 2011). In response to this claim Gutman et al (Guttman, Russell, Ingolia, Weissman, & Lander, 2013) developed Ribosome Release Score (RRS) which measures the drop of ribosome footprint density downstream of ORF stop codons and have shown that a high RRS score is a signature of annotated protein coding ORFs, but not of ORFs found in 5’ leaders and lincRNAs. While RRS provides a useful metric for estimating the accuracy of translation termination, its use as a sole signature of translation is peculiar since it assumes that no re-initiation or leaky scanning takes place, while both phenomena are well documented in eukaryotic cells, see (Hinnebusch, 2014; Hinnebusch, Ivanov, & Sonenberg, 2016; Shirokikh & Preiss, 2018) for reviews. Re-initiation often takes place after termination at short ORFs and a large fraction of ribosome scanning complexes bypass start codons in a poor initiation context. This leads to a complex organization of short overlapping translated ORFs in the beginning of RNA transcripts where high ribosome density is observed both upstream and downstream of stop codons leading to low RRS scores. While RRS can indeed be used as a signature of ORF translation, since isolated ORFs are expected to exhibit high RRS scores, it is important to be aware of RRS limitations in detecting overlapping or closely located translated ORFs.

Indeed, a follow up study (Ingolia et al., 2014) developed another metric, fragment length organization similarity score (FLOSS) that is based on the similarity of length distributions of footprints across different transcripts and have shown that they can successfully discriminate RNA fragments mapped to genuine non-coding RNAs from those observed at translated ORFs providing further support to the initial claim that ribosomes do translate many short ORFs in 5’ leaders and RNA transcripts previously annotated as non-coding. Rfoot (Ji, 2018) uses the same principle of analysing read length distributions to identify non-ribosomal RNA footprints.
In addition to a characteristic distribution of read lengths in translated ORFs another feature that is strongly associated with translation is triplet periodicity, however, the detection of triplet periodicity is difficult when the ORF length is short due to the high heterogeneity of the signal. To mitigate this issue Calviello et al. (Calviello et al., 2016) designed RiboTaper which is based on the multitaper approach (Thomson, 1982) developed for signal processing that performs a spectral analysis on a signal that has been transformed in a number of different ways (tapers). SPECtre (Chun et al., 2016) is another tool for detecting periodicity based on spectral analysis of aligned Ribo-Seq data developed around the same time. More recently RiboWave (Xu et al., 2018a) was developed, which makes use of wavelet transformation to denoise ribosome profiling signal, and claims to outperform previously developed tools. Changes in ribosome footprint density can also be used as signature of translation. In addition to a drop of ribosome density at the ends of ORFs, many datasets exhibit characteristic patterns with elevated ribosome density at the beginning and the end of ORFs and this information can be taken into account when scoring potentially translated ORFs as in RiboHMM (Raj et al., 2016). The problem of this approach is that such changes in footprint density are often data specific and HMM emission probabilities obtained from the analysis of one dataset may not suit another dataset.

There are certain variations of ribosome profiling methods that enrich ribosomes at the starts of translation

initiation using specific translation inhibitors or their combinations (Gao et al., 2015; Ingolia et al., 2011).This information can also be utilized for the detection of translated ORFs as in Ribo-TISH (P. Zhang et al., 2017) and is especially useful for localisation of start codons at which ORF translation is initiated, as it is often more difficult than detection of translation itself since translation initiation often takes place at non-AUG codons (Ivanov, Firth, Michel, Atkins, & Baranov, 2011) especially when close to the 5' ends (Michel, Andreev, & Baranov, 2014) and sometimes multiple start codons are being used to initiate the same ORF as in *PTEN* (Tzani et al., 2016). Some tools such as Ribo-TISH (P. Zhang et al., 2017) and the recently developed DeepRibo (Clauwaert, Menschaert, & Waegeman, 2019) which uses neural networks to annotate bacterial genomes, are capable of utilising both elongating and initiating Ribo-Seq data.

While many tools for predicting translated ORFs exist (see Table 3), their predictions differ considerably. Moreover, it is difficult to make specific recommendations on what software to use in the absence of independent benchmarking studies. Such benchmarking is very difficult to carry out due to a lack of gold standard sets of translated ORFs and adequate methodology orthogonal to ribosome profiling. A set of annotated protein coding genes cannot be used as a gold standard dataset since it is biased towards long ORFs coding for functional proteins. Although mass spectrometry analysis (Van Damme, Gawron, Van Criekinge, & Menschaert, 2014; Vanderperre et al., 2013) and phylogenetic analysis (Andreev, O'Connor, Fahey, et al., 2015; Bazzini et al., 2014) are being used as orthogonal methodology, neither is truly adequate. Many of the short translated ORFs are unlikely to produce stable peptides that can be detected with mass spectrometry, though efforts have been made to combine proteomics and ribosome profiling evidence such as with Proteoformer (Crappe et al., 2015), Proteoformer 2 (Verbruggen et al., 2019) and OpenProt (Brunet, Brunelle, et al., 2018). Similarly, the signal obtained from phylogenetic conservation depends on the length of the ORF and the depth of its conservation. Translation of some ORFs may not affect fitness and would evolve neutrally. Also, a functional ORF was recently reported for which no evidence of evolutionary selection was found (C. Xie et al., 2019). Therefore, in the absence of benchmarking standards and appropriate orthogonal methodology, the software described in this section can be used for exploratory analysis only. Despite these limitations ribosome profiling has been used to successfully confirm novel translated regions (Castelo-Szekely et al., 2019; Chugunova et al., 2019; Hardy et al., 2019) and even to discover a novel mechanism of translation regulation (Yordanova et al., 2018).

Differential gene expression

Ribosome profiling analysis is probably most frequently used for the characterization of differential gene expression as part of a time series or control/treatment group. It is assumed that slowly and rapidly decoded codons are distributed somewhat equally and therefore the relative frequency with which footprints are mapped to a specific ORF should be proportional to the levels of RNA bearing this ORF and efficiency of translation initiation at this ORF. In other words, the ribosome profiling signal is reflective of the total protein synthesis which accounts for the RNA levels (synthesis and degradation) and the rate of RNA translation. Ribosome profiling experiments usually are carried out in parallel with RNA-seq experiments that allow determination of RNA levels. When RNA levels do not change, but the ribosome profiling signal changes, it is reasonable to attribute these changes to changes in translation efficiencies. Note, however, that changes in local densities could be also caused by ribosome pausing. In this case, the induction of a ribosome pause at a specific location may be misinterpreted as increased translation. For example, Lobanov et al 2017 have noticed that Euplotes mRNAs containing sites of ribosomal frameshifting have a higher ratio of Ribo-Seq to RNA-seq reads than mRNAs translated without frameshifting. They attributed this difference to ribosome pauses rather than higher translation rates. To mitigate the influence of ribosome pauses on the assessment of differential translation, coordinates with the highest peaks of density could be excluded from the analysis as has been done in Andreev et al 2015. Yet another alternative would be to do a bootstrap sampling of densities from random CDS coordinates. Inconsistencies in differential gene expression analysis revealed by such a bootstrapping procedure would indicate a potential problem associated with ribosome pausing.

Often attempts are made to measure differential translation even when RNA levels do change simply by dividing the number of ribosome footprints aligning to an ORF by the number of RNA-seq reads. Such a procedure has several problems. First, it results in ratios that, unlike countable data (footprints and RNA-seq reads), do not carry information on the statistical significance, e.g. the ratio of 2/4 equals the ratio 200/400. Second, the best fit for the distribution of such ratio values is believed to follow the Cauchy distribution that is hard to model

since both its mean and variance are undefinable. Finally, Ola Larsson and colleagues pointed out that spurious correlation between such ratios and their components (e.g. RNA levels) is necessitated mathematically (Larsson, Sonenberg, & Nadon, 2010).

In principle, differential translation could be defined as a miscorrelation between the RNA-seq and ribosome profiling signal and it can be detected with the tools designed for RNA-seq analysis such as DESeq2 (Love, Huber, & Anders, 2014) and EdgeR (Robinson, McCarthy, & Smyth, 2010) . Nonetheless, several standalone tools designed specifically for the characterization of differential translation efficiency from ribosome profiling data have been developed recently. Examples include babel (Olshen et al., 2013) , RiboDiff (Zhong et al., 2017), Riborex (W. Li, Wang, Uren, Penalva, & Smith, 2017), Xtail (Xiao, Zou, Liu, & Yang, 2016), RIVET (Ernlund, Schneider, & Ruggles, 2018) and Anota2Seq (Oertlin et al., 2019), see Table 4. Online databases such as Trips-Viz (Kiniry, O'Connor, et al., 2018) and TranslatomeDB (W. Liu, Xiang, Zheng, Jin, & Zhang, 2018) also provide functionalities for differential gene expression characterization with the former applying a simple Z-score transformation for this purpose (Andreev, O'Connor, Fahey, et al., 2015; Quackenbush, 2002). As the statistical frameworks of these tools differ, not surprisingly, the sets and the number of genes predicted by them as differentially regulated differ. The field is seemingly in need of objective and independent benchmarking. It is important to note that irrespective of the specific approaches used for the assessment of differential translation, the differences are relative and not absolute. Measurements of absolute changes in differential translation are not possible without spike-in controls allowing for the normalisation of the number of reads relative to the number of cells. Although attempts to introduce spike-in controls in ribosome profiling experiments have been made, e.g. Ingolia et al 2014, Andreev et al 2015, Iwasaki et al 2016, Popa et al 2016, and Gorochowski et al 2019, their suitability have not yet been rigorously assessed.

Besides difficulties in evaluating differential expression based on two countable signals, the task is exacerbated by the existence and translation of multiple RNA isoforms due to alternative splicing and transcription initiation in complex eukaryotes, such as mammals (Blencowe, 2006). By mapping ribosome footprints across exon-exon junctions of alternatively spliced isoforms it has been shown that alternative isoforms could indeed be simultaneously translated (Weatheritt, Sterne-Weiler, & Blencowe, 2016). However, when more than one RNA isoform is translated in the same sample, it is extremely difficult to compare their relative translation. Even when a certain cell type expresses only one predominant isoform, it is not apparent how to choose the one that will be used as a reference. In practice several heuristics are commonly applied to deal with this problem, each of which could lead to specific artefacts. One method is to use the longest isoform or the isoform with the longest annotated coding region. The rationale is that even if such an isoform differs from what is present in the cell, reads derived from the shorter isoform would align to the longer one allowing for measurement of expression differences. However, this can be problematic in cases where the shorter isoforms have coding exons that are missing in longer isoforms. This problem could be solved with creating a “union” of all transcripts by collapsing the genomic co-ordinates of all possible exons. While this is a sensible approach for the analysis of differential gene expression at the “gene level”, it may not be appropriate for the analysis of translated features within mRNA, e.g. uORFs, because such a union may disrupt such uORFs. When it is necessary to choose only a single transcript, so called “principal isoforms” could be used which are curated in the APPRIS database (Rodriguez et al., 2018). However, 5’ leaders/3’trailers are not taken into account here, meaning multiple isoforms that differ only in their noncoding regions would all be annotated as the principal isoform. In an attempt to move away from these heuristic approaches and their shortcomings, software has been developed which takes Ribo-Seq data into account to do transcript isoform level quantification, i.e. Ribomap (Wang, McManus, & Kingsford, 2016), ORQAS (Reixachs-Solé, Ruiz-Orera, Alba, & Eyras, 2019) and SaTann (Calviello, Hirsekorn, & Ohler, 2019), see Table 5. However, these tools assign footprints to different isoforms under the premise that their protein synthesis input is directly proportional to their RNA levels, i.e. they are translated with the same efficiency. This, however, may not always be the case, especially when different start codons are used in different isoforms. The information that can be used within the Ribo-Seq data itself is reads that uniquely align to a specific isoform (unique exons and exon-exon junctions), however, because this is often within short regions of mRNAs, the number of footprints mapped to them could be sensitive to differences in ribosome dwell times at these locations. On top of that Ribo-Seq data could not be used to discriminate between alternative isoforms that differ in exons that are not translated.

Pause detection

Elongation rates varies as the ribosome traverses an mRNA and ribosomes could pause or stall at certain locations. Ribosome stalling can be caused by factors such as the secondary structure (Pop et al., 2014; Somogyi, Jenner, Brierley, & Inglis, 1993; Tholstrup, Oddershede, & Sorensen, 2012), the interaction of the nascent peptide with the ribosome peptide channel (Becker, Oh, Weissman, Kramer, & Bukau, 2013; Tenson & Ehrenberg, 2002) and certain combinations of codons (Woolstenhulme et al., 2015). Pause sites have been shown to play important roles in translation in areas such as protein folding (Fluman, Navon, Bibi, & Pilpel, 2014; Tsai et al., 2008), and regulation of protein synthesis (Ivanov et al., 2018; Kurian et al., 2011; Yordanova et al., 2018). Pause sites are reflected in the Ribo-Seq data by high peaks relative to the surrounding region (Figure 1b). Like with translated ORF detection, pauses in Ribo-Seq data can be identified with manual visual inspection of ribosome footprint density profiles of individual mRNAs, but genome or transcriptome scale detection of pauses requires dedicated software.

PausePred (Kumari, Michel, & Baranov, 2018) is one such tool, available in both browser based and standalone versions that allows users to upload Ribo-Seq data and an optional annotation file. It then uses a sliding window approach to search for regions with high Ribo-Seq peaks relative to the background density. An accompanying tool, Rfeet then allows visualisation of the Ribo-Seq and (optionally) corresponding RNA-Seq data. Taking RNA-Seq into consideration is an important step when detecting pauses in Ribo-Seq data since it controls for the peaks caused by alignment artefacts. For example, when ambiguous mapping is allowed, a short region in a lowly expressed gene that shares sequence similarity with a highly expressed gene will appear as a peak in Ribo-Seq data. Similarly, if allowing only unambiguous alignments a short unique sequence surrounded by non-unique sequence will appear as a peak. Finally, a region with low sequence complexity that has many reads mapped just by chance can also appear as a peak in Ribo-Seq data. In all of these cases RNA-Seq data will also exhibit a pause at the same location, but not so in the case of a genuine ribosomal pause.

Prediction of footprint density

Ribo-Seq profiles are noticeably non-uniform, arising in part from differences in ribosome decoding rates in addition to the presence of sequencing biases occurring due to substrate sequence specificity of the enzymes used in generation and sequencing of cDNA libraries. Global assessment of footprint density allows for the magnitude of these biases to be estimated. A number of tools (see Table 6) have been developed to assess footprint density, including RUST (Ribo-Seq Unit Step Transformation) which allows for the measurement of how much various sequence features consistently influence the density of footprints at specific positions relative to the decoding center of the ribosome, .i.e. RUST would not detect features that only influence a small subset of unique locations (O'Connor et al., 2016). Metafootprint plots generated with RUST, (Figure 4), visualize these dependencies, Figure 4a is an example of a dataset with low sequencing bias from (Eichhorn et al., 2014) and Figure 4b is an example of a dataset with high sequencing bias from (Reid, Xu, Chen, Yang, & Sun, 2017). It is expected that the influence of the sequence at the decoding centre (i.e. A- and P-sites) should exceed that at the regions corresponding to read ends due to sequencing biases. Using the parameters of theses dependencies RUST can be used to predict ribosome profiling densities at the sequences for which no data exist with high accuracy.

Riboshape (T. Y. Liu & Song, 2016) is another tool which aims to understand the sequence features responsible for Ribo-Seqs non-uniformity. It does this using kernel smoothing to predict sequence features and then predicts the “shape” of ribosome profiles. The authors find that footprint density in *Saccharomyces cerevisiae* can be predicted with high accuracy. More recently developed tools utilise the power of deep learning to predict footprint density, such as ROSE (RibosOme Stalling Estimator) (S. Zhang et al., 2017) which is trained on transcripts with high ribosome profiling density to predict locations of ribosome pauses on transcripts with little to no signal. Finally iXnos (Tunney et al., 2018), which also uses neural networks, also aims to predict footprint densities. The authors compared the performance of iXnos to RUST and Riboshape and have shown that it outperforms them both on a single test dataset. They also demonstrated the utility of iXnos for optimizing the coding sequence to increase the translation efficiency.

Pipelines, libraries, environments and data resources

In the absence of dedicated software, the early ribosome profiling analysis was carried out with tools developed for other high throughput sequencing applications and with *ad hoc* computer scripts. Over the past decade a number of different pipelines, computational environments and data resources have been developed. Researchers now have a considerable choice of existing freely available software to suit their needs, platform preferences and style. There is a large number of pipelines written in different languages for processing raw ribosome profiling data (see Table 1), e. g. the ruby based pipeline RiboPip (Stefan, 2016) and the R package systemPipeR (H. Backman TW & Girke, 2016) that provide full workflows for Ribo-Seq and RNA-Seq data analysis as well as other techniques such as CHIP-Seq in the latter package. Since raw data processing is not specific to ribosome profiling (outside of the above considerations) many packages and pipelines developed specifically for ribosome profiling take processed aligned reads in BAM file as input and provide only additional functionality. An example is a rich and extensible python library Plastid (Dunn & Weissman, 2016). Likewise R packages Riboprofiling (Popa et al., 2016) and RiboseqR (Chung et al., 2015) also take alignment files as input and enable multifunctional downstream analysis.

The above software packages are operational through a command line and expect a certain familiarity with the Linux operating system. Moreover, setting up such software may require a certain effort and additional expertise for installing the software to a specific environment. The required skills and available time are often understandably lacking among wet lab researchers. The Galaxy Project (Afgan et al., 2018) offers a solution to this problem by providing a graphical web-based interface for data analysis, where workflows can be saved and rerun making the analysis reproducible. Software packages that do not have their own graphical interface could easily be integrated into Galaxy. Numerous specialized Galaxy servers have been created that provide the tools needed for a specific type of data. RiboGalaxy (Michel et al., 2016) is such an instance of Galaxy that provides several pipelines for the analysis of ribosome profiling data. RiboGalaxy is a part of the RiboSeq.Org collection (Figure 5).

Like other sequencing data, ribosome profiling data can be found in public databases and many journals make deposition of the data to public archives a prerequisite for publication. The availability of the raw data, however, does not mean that these data can be easily utilised. Processing the raw data requires software, computational power, time and most importantly a certain level of familiarity with ribosome profiling data and technical issues described in this review. To democratize the data and to make it available to a large biomedical community that could benefit from it, it is important to provide access not only to raw data, but also to processed alignments. The first such database was GWIPS-Viz (for Genome Wide Information on Protein Synthesis Visualized) (Michel, Fox, et al., 2014). It provides genomic alignments of uniformly processed ribosome footprints and corresponding RNA-seq fragments. The alignments can be visualized either individually for specific datasets or as aggregates. To date GWIPS-Viz hosts Ribo-Seq data from 23 organisms (Michel et al., 2018). See (Kiniry, Michel, et al., 2018; Michel, Ahern, Donohue, & Baranov, 2015) for tutorials on how to use GWIPS-viz. Another large database of processed and aligned ribosome profiling data is RPFdb (Wang et al., 2018; S. Q. Xie et al., 2016). While both GWIPS-viz and RPFdb are databases of genomic alignments of ribosome footprints, their functionality is markedly different, and their abilities overlap minimally. For example, RPFdb provides rich information on specific datasets such as raw counts and RPKM values for specific loci. Other databases such as the newly developed resource, Trips-Viz (for Transcriptome Information on Protein Synthesis Visualized) (Kiniry, O'Connor, et al., 2018) align data to a transcriptome as this has the advantage of eliminating the problem of mapping across exon-exon junctions. Trips-Viz is a web-based collaborative interactive environment for graphical computational analysis of publicly available Ribo-Seq data (although user generated data can also be uploaded). Several other databases have been developed that provide information derived from ribosome profiling analysis, see Table 2.

Conclusion

Over the past decade, many software modules, pipelines, visualization tools and data resources have been developed for Ribo-Seq analysis. As outlined in this review, more than one solution is now available for many tasks, from raw data processing to high-end applications such as the detection of translated ORFs. Subsequently researchers can choose tools to fit their specific computational backgrounds and styles. Nonetheless, despite the ample availability of resources, the field is far from saturation. We predict that it will continue to develop, perhaps, at an accelerated pace due to the following reasons.

The ribosome profiling protocol itself continues to develop. Specific modifications of experimental procedures sometimes require development of new tools. Even more importantly, a number of issues in ribosome profiling data analysis remain unsolved, e.g. differential expression analysis of allelic variants. While parallel approaches exist for the same tasks (e.g. translated ORF detection, differential gene expression), the results obtained with these approaches often poorly converge. This is largely due to the lack of gold standards and reliable criteria for evaluating the performance of these tools. Development of benchmarking approaches is expected to lead to improvement of these tools by providing the means for their comparison and optimization.

Recent developments in the characterization of mRNA translation, largely fuelled by ribosome profiling, further revealed the complexity of the translational landscapes of individual mRNAs, especially of high eukaryotes and specifically human mRNAs. Translation initiation could take place on many codons in the same mRNA (Fritsch et al., 2012; Lee et al., 2012) , leading to the production of proteoforms with different N-termini (Ivanov et al., 2011; Menschaert et al., 2013). At the same time ribosomes reading through the stop codons lead to the generation of proteoforms with different C-termini (Jungreis et al., 2011; Loughran et al., 2018; Rajput, Pruitt, & Murphy, 2019; Schueren et al., 2014). On top of that a large proportion of mRNAs contain short translated ORFs (Andreev, O'Connor, Fahey, et al., 2015; Ji, Song, Regev, & Struhl, 2015; Johnstone, Bazzini, & Giraldez, 2016), some of which encoding functional proteins as in the human *MIEF1* mRNA (Andreev, O'Connor, Fahey, et al., 2015; Brown et al., 2017; Delcourt et al., 2018). The currently used data structures for representation of RNA transcripts based on a single reference transcript with a single CDS are not suited for the representation of this complexity (Brunet, Levesque, Hunting, Cohen, & Roucou, 2018). Thus, we envision the development of new, more adequate, data structures. The computational tools will need to adapt subsequently to these data structures.

Ribosome profiling allows for the quantitative assessment of only a single aspect of cellular activity, translation of its mRNAs. Often taking advantage of these data requires integration with other types of data (transcription initiation sites mapping, epitranscriptomics, mass spectrometry, etc.), and hence the tools for ribosome profiling data analysis need to provide such functionality either directly or through interoperability with the computational tools developed for the analysis of the data obtained with other techniques.

Yet another challenge is posed by the changes occurring in the analysis of big biodata in general. The volume of data in the Sequencing Data Archive doubles every 10-20 months (Langmead & Nellore, 2018) which is faster than the growth of computational power. While ribosome profiling data currently represents only a microscopic fraction of these data, it is unlikely that the volume of ribosome profiling data will be growing at a slower pace. Thus, the computational efficiency of the algorithms will become critical. As the data volumes increase their physical transfer between servers is becoming increasingly less practical. This necessitates a paradigm shift from data-to-tools to the tools-to-data which requires the development of dedicated cloud infrastructure (Langmead & Nellore, 2018). In the future tools will need to be adapted for these new environments.

Figures and Tables.

Table 1. Software environments for data processing, pipelines for quality assessment, offset detection and miscellaneous software.

Name	Notes	URL	Ref.
mQc	Quality assessment, part of PROTEOFORMER pipeline	https://github.com/Biobix/mQC	Verbruggen & Menschaert, 2018)
Plastid	Python based library	https://plastid.readthedocs.io/en/latest/	(Dunn & Weissman, 2016)
Rfoot	Inference of RNA-binding protein sites	https://github.com/zhejilab/Rfoot	(Ji, 2018)
Ribodeblur	Offset determination	https://github.com/Kingsford-Group/ribodeblur	(Wang et al., 2017)
RiboGalaxy	Galaxy based environment	https://ribogalaxy.ucc.ie/	(Michel et al., 2016)
Ribopip	Ruby based processing pipeline	https://github.com/stepf/RiboPip	(Stefan, 2016)
Riboprofiling	R based processing pipeline	http://bioconductor.org/packages/release/bioc/html/RiboProfiling.html	(Popa et al., 2016)
RiboProp	Offset determination	http://bioserv.mps.ohio-state.edu/RiboProP/	(Zhao et al., 2018)
RiboseqR	R based processing pipeline	http://bioconductor.org/packages/release/bioc/html/riboSeqR.html	(Chung et al., 2015)
RibostreamR	Web based analysis of user generated Ribo-Seq data	https://github.com/pjperki2/riboStreamR	(Perkins, Mazzoni-Putman, Stepanova, Alonso, & Heber, 2019),
RiboWaltz	Offset determination	https://github.com/LabTranslationalArchitectomics/riboWaltz	(Lauria et al., 2018)
Ribo-seQC	Quality assessment	https://github.com/ohlerlab/RiboseQC	(Calviello, Sydow, et al., 2019)
RRS	measures drop-off of ribosome footprint density at the end of ORFs	https://rdrr.io/github/JokingHero/ORFik/man/ribosomeReleaseScore.html	(Guttman et al., 2013)
SystemPipeR	R based processing pipeline	https://bioconductor.org/packages/release/bioc/html/systemPipeR.html	(H. Backman TW & Girke, 2016)
Trips-Viz	Web based analysis of public and user generated ribosome profiling data	https://trips.ucc.ie	(Kiniry, O'Connor, et al., 2018)

Table 2. Data resources and visualization environments.

GWIPS-viz	Genome browser for visualization of Ribo-Seq data aligned to genomes	https://gwips.ucc.ie	(Michel et al., 2014; Michel et al, 2015; Michel et al., 2018; Kiniry et al 2018)
HRPDViewer	A resource for visualization of Ribo-Seq data aligned to transcriptomes	http://cosbi4.ee.ncku.edu.tw/HRPDviewer/	(W. S. Wu et al., 2018)
Openprot	Database and viewer for exploration of Ribo-Seq and mass-spec data supporting translation of non-annotated ORFs	https://openprot.org/	(Brunet, Brunelle, et al., 2018)
RiboSeqDB	Repository of human and mouse ribosome profiling data	https://micro.biouml.org/bioumlweb/	(W. Liu et al., 2018)
RiboViz	Online tool for visualization of publicly available Ribo-Seq data	https://riboviz.org/	(Carja et al., 2017)
RPFdb	Database of ribosome profiling datasets rich in metainformation and their genomic alignments	http://sysbio.gzzoc.com/rpfdb/	(S. Q. Xie et al., 2016)
sORFs.org	Database of short ORFs whose translation is supported with Ribo-Seq data	http://sorfs.org	http://sorfs.org
svist4get	Command-line visualization tool	https://bitbucket.org/artegorov/svist4get/	(Egorov et al., 2019)
TranslatomeDB	On-line resource for visualization of public and user generated data	http://translatomedb.net/	(W. Liu, Xiang, Zheng, Jin, & Zhang, 2018)

Trips-Viz	On-line environment for graphical exploration of public and user generated ribosome profiling data aligned to transcriptomes.	https://trips.ucc.ie	(Kiniry, O'Connor, et al., 2018)

Table 3. Software tools for automatic detection of translated ORFs.

Name	Notes	URL	Ref.
DeepRibo	Detection of translated ORFs in bacterial genomes	https://github.com/Biobix/DeepRibo	(Clauwaert, Menschaert, & Waegeman, 2019)
orfRater	Detection of translated ORFs based on linear regression	https://github.com/alexfields/ORF-RATER	(Fields et al., 2015)
ORFScore	Scoring translated ORFs based on triplet periodicity	https://rdrr.io/bioc/ORFik/man/orfScore.html	(Bazzini et al., 2014)
PreTis	Detection of translation initiation starts based on linear regression	http://service.bioinformatik.uni-saarland.de/pretis/	(Reuter, Biehl, Koch, & Helms, 2016)
PRICE	Detection of translated ORFs using EM algorithm	https://github.com/erhard-lab/price	(Erhard et al., 2018)
Proteoformer	Detection of translated ORFs with support from mass-spec data	https://github.com/Biobix/proteoformer	(Crappe et al., 2015; Verbruggen et al., 2019)
REPARATION	Detection of translated ORFs in bacterial genomes	https://github.com/Biobix/REPARATION	(Ndah et al., 2017)
Ribocode	Detection of translated ORFs based on triplet periodicity	https://github.com/xryanglab/RiboCode	(Xiao et al., 2018)
riboHMM	HMM based detection of translated ORFs	https://github.com/rajanil/riboHMM	(Raj et al., 2016)
RibORF	SVM based identification of translated ORFs	https://github.com/zhejilab/RibORF	(Ji et al., 2015)
Ribosome profiling analysis framework	Detection of translated ORFs based on triplet periodicity	https://github.com/LUMC/ribosome-profiling-analysis-framework	(de Klerk et al., 2015)
RiboTaper	Detection of translated ORFs based on spectral analysis of Ribo-Seq signal using multitaper	https://ohlerlab.mdc-berlin.de/software/RiboTaper_126/	(Calviello et al., 2016)
Ribo-TISH	Is able to use Ribo-Seq data enriched at starts of initiation in addition to regular Ribo-Seq.	https://github.com/zhpn1024/ribotish	(P. Zhang et al., 2017)
RiboWave	Detection of translated ORFs based on spectral analysis of Ribo-Seq signal with Wavelet	https://github.com/lulab/Ribowave	(Xu et al., 2018b)

	transformation		
Rp-Bp	Bayesian approach for detecting translated ORFs.	https://github.com/dieterich-lab/rp-bp	(Malone et al., 2017)
SPECtre	Detection of translated ORFs based on spectral analysis of Ribo-Seq signal	https://github.com/mills-lab/spectre	(Chun et al., 2016)
uORF-seqr	Regression based detection of translated ORFs.	https://github.com/pspealman/uorfseqr	(Spealman et al., 2018)

Table 4. Software for the analysis of differential translation.

Name	URL	Ref.
Anota2Seq	https://bioconductor.org/packages/release/bioc/html/anota2seq.html	(Oertlin et al., 2019)
Babel	https://cran.r-project.org/web/packages/babel/index.html	(Olshen et al., 2013)
Ribodiff	https://github.com/ratschlab/RiboDiff	(P. Zhang et al., 2017)
Riborex	https://github.com/smithlabcode/riborex	(W. Li et al., 2017)
Rivet	https://ruggleslab.github.io/rivet/	(Ernlund et al., 2018)
Xtail	https://github.com/xryanglab/xtail	(Xiao et al., 2016)

Table 5. Software for the analysis of specific isoforms.

Name	URL	Ref.
Orqas	http://www.cs.cmu.edu/~ckingsf/software/ribomap/	(Reixachs-Solé et al., 2019)
Ribomap	https://github.com/lcalviell/SaTAnn	(Wang, McManus, & Kingsford, 2016)
SaTann	https://github.com/comprna/ORQAS	(Calviello, Hirsekorn, et al., 2019)

Table 6. Software for the analysis of local footprint densities.

Name	Notes	URL	Ref.
iXnos	Neural network based model of local densities. Can be used to predict local densities and for sequence optimization for increased expression	https://github.com/lareaulab/iXnos	(Tunney et al., 2018)
Pausepred	Detection of local peaks	https://pausepred.ucc.ie/	(Kumari et al., 2018)
Riboshape	A kernel-smoothing model enabling prediction of local densities.	https://sourceforge.net/projects/riboshape/	(T. Y. Liu & Song, 2016)

Rose	An approach for predicting ribosome stalling sites using deep convolutional network.	https://github.com/mlcb-thu/rose	(S. Zhang et al., 2017)
RUST	Unit-step based normalization of footprint densities for the analysis of sequence features effecting footprint densities, can be used to predict local densities.	https://lapti.ucc.ie/rust/	(O'Connor et al., 2016)

Acknowledgments

The authors wish to thank Patrick O'Connor for his constructive criticism of the manuscript. The authors wish to acknowledge support by SFI-HRB-Wellcome Trust Biomedical Research Partnership (Investigator Award in Science) [210692/Z/18/Z to P.V.B.] and Irish Research Council to A.M.M. and S.J.K.

Conflict of interests

AMM and PVB are founders of RiboMaps Ltd, a company providing ribosome profiling as a service.

References

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Cech, M., . . . Blankenberg, D. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res, 46*(W1), W537-W544. doi:10.1093/nar/gky379

Albert, F. W., Muzzey, D., Weissman, J. S., & Kruglyak, L. (2014). Genetic influences on translation in yeast. *PLoS Genet, 10*(10), e1004692. doi:10.1371/journal.pgen.1004692

Andreev, D. E., O'Connor, P. B., Fahey, C., Kenny, E. M., Terenin, I. M., Dmitriev, S. E., . . . Baranov, P. V. (2015). Translation of 5' leaders is pervasive in genes resistant to eIF2 repression. *Elife, 4*, e03971. doi:10.7554/eLife.03971

Andreev, D. E., O'Connor, P. B., Loughran, G., Dmitriev, S. E., Baranov, P. V., & Shatsky, I. N. (2017). Insights into the mechanisms of eukaryotic translation gained with ribosome profiling. *Nucleic Acids Res, 45*(2), 513-526. doi:10.1093/nar/gkw1190

Andreev, D. E., O'Connor, P. B., Zhdanov, A. V., Dmitriev, R. I., Shatsky, I. N., Papkovsky, D. B., & Baranov, P. V. (2015). Oxygen and glucose deprivation induces widespread alterations in mRNA translation within 20 minutes. *Genome Biol, 16*, 90. doi:10.1186/s13059-015-0651-z

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Archer, S. K., Shirokikh, N. E., Beilharz, T. H., & Preiss, T. (2016). Dynamics of ribosome scanning and recycling revealed by translation complex profiling. *Nature, 535*(7613), 570-574. doi:10.1038/nature18647

Bazzini, A. A., Johnstone, T. G., Christiano, R., Mackowiak, S. D., Obermayer, B., Fleming, E. S., . . . Giraldez, A. J. (2014). Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO J, 33*(9), 981-993. doi:10.1002/embj.201488411

Becker, A. H., Oh, E., Weissman, J. S., Kramer, G., & Bukau, B. (2013). Selective ribosome profiling as a tool for studying the interaction of chaperones and targeting factors with nascent polypeptide chains and ribosomes. *Nat Protoc, 8*(11), 2212-2239. doi:10.1038/nprot.2013.133

Blencowe, B. J. (2006). Alternative splicing: new insights from global analyses. *Cell, 126*(1), 37-47. doi:10.1016/j.cell.2006.06.023

Brandman, O., & Hegde, R. S. (2016). Ribosome-associated protein quality control. *Nat Struct Mol Biol, 23*(1), 7-15. doi:10.1038/nsmb.3147

Brar, G. A., & Weissman, J. S. (2015). Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nat Rev Mol Cell Biol, 16*(11), 651-664. doi:10.1038/nrm4069

Brown, A., Rathore, S., Kimanius, D., Aibara, S., Bai, X. C., Rorbach, J., . . . Ramakrishnan, V. (2017). Structures of the human mitochondrial ribosome in native states of assembly. *Nat Struct Mol Biol, 24*(10), 866-869. doi:10.1038/nsmb.3464

Brunet, M. A., Brunelle, M., Lucier, J. F., Delcourt, V., Levesque, M., Grenier, F., . . . Roucou, X. (2018). OpenProt: a more comprehensive guide to explore eukaryotic coding potential and proteomes. *Nucleic Acids Res*. doi:10.1093/nar/gky936

Brunet, M. A., Levesque, S. A., Hunting, D. J., Cohen, A. A., & Roucou, X. (2018). Recognition of the polycistronic nature of human genes is critical to understanding the genotype-phenotype relationship. *Genome Res, 28*(5), 609-624. doi:10.1101/gr.230938.117

Buskirk, A. R., & Green, R. (2017). Ribosome pausing, arrest and rescue in bacteria and eukaryotes. *Philos Trans R Soc Lond B Biol Sci, 372*(1716). doi:10.1098/rstb.2016.0183

Calviello, L., Hirsekorn, A., & Ohler, U. (2019). SaTAnn quantifies translation on the functionally heterogeneous transcriptome. doi:10.1101/608794

Calviello, L., Mukherjee, N., Wyler, E., Zauber, H., Hirsekorn, A., Selbach, M., . . . Ohler, U. (2016). Detecting actively translated open reading frames in ribosome profiling data. *Nat Methods, 13*(2), 165-170. doi:10.1038/nmeth.3688

Calviello, L., & Ohler, U. (2017). Beyond Read-Counts: Ribo-seq Data Analysis to Understand the Functions of the Transcriptome. *Trends Genet, 33*(10), 728-744. doi:10.1016/j.tig.2017.08.003

Calviello, L., Sydow, D., Harnett, D., & Ohler, U. (2019). doi:10.1101/601468

Carja, O., Xing, T., Wallace, E. W. J., Plotkin, J. B., & Shah, P. (2017). riboviz: analysis and visualization of ribosome profiling datasets. *BMC Bioinformatics, 18*(1), 461. doi:10.1186/s12859-017-1873-8

Castelo-Szekely, V., De Matos, M., Tusup, M., Pascolo, S., Ule, J., & Gatfield, D. (2019). Charting DENR-dependent translation reinitiation uncovers predictive uORF features and links to circadian timekeeping via Clock. *Nucleic Acids Res, 47*(10), 5193-5209. doi:10.1093/nar/gkz261

Chew, G. L., Pauli, A., Rinn, J. L., Regev, A., Schier, A. F., & Valen, E. (2013). Ribosome profiling reveals resemblance between long non-coding RNAs and 5' leaders of coding RNAs. *Development, 140*(13), 2828-2834. doi:10.1242/dev.098343

Chugunova, A., Loseva, E., Mazin, P., Mitina, A., Navalayeu, T., Bilan, D., . . . Dontsova, O. (2019). LINC00116 codes for a mitochondrial peptide linking respiration and lipid metabolism. *Proc Natl Acad Sci U S A, 116*(11), 4940-4945. doi:10.1073/pnas.1809105116

Chun, S. Y., Rodriguez, C. M., Todd, P. K., & Mills, R. E. (2016). SPECtre: a spectral coherence--based classifier of actively translated transcripts from ribosome profiling sequence data. *BMC Bioinformatics, 17*(1), 482. doi:10.1186/s12859-016-1355-4

Chung, B. Y., Hardcastle, T. J., Jones, J. D., Irigoyen, N., Firth, A. E., Baulcombe, D. C., & Brierley, I. (2015). The use of duplex-specific nuclease in ribosome profiling and a user-friendly software package for Ribo-seq data analysis. *RNA, 21*(10), 1731-1745. doi:10.1261/rna.052548.115

Clauwaert, J., Menschaert, G., & Waegeman, W. (2019). DeepRibo: a neural network for precise gene annotation of prokaryotes by combining ribosome profiling signal and binding site patterns. *Nucleic Acids Res*. doi:10.1093/nar/gkz061

Crappe, J., Ndah, E., Koch, A., Steyaert, S., Gawron, D., De Keulenaer, S., . . . Menschaert, G. (2015). PROTEOFORMER: deep proteome coverage through ribosome profiling and MS integration. *Nucleic Acids Res, 43*(5), e29. doi:10.1093/nar/gku1283

de Klerk, E., Fokkema, I. F., Thiadens, K. A., Goeman, J. J., Palmblad, M., den Dunnen, J. T., . . . t Hoen, P. A. (2015). Assessing the translational landscape of myogenic differentiation by ribosome profiling. *Nucleic Acids Res, 43*(9), 4408-4428. doi:10.1093/nar/gkv281

Delcourt, V., Brunelle, M., Roy, A. V., Jacques, J. F., Salzet, M., Fournier, I., & Roucou, X. (2018). The Protein Coded by a Short Open Reading Frame, Not by the Annotated Coding Sequence, Is the Main Gene Product of the Dual-Coding Gene MIEF1. *Mol Cell Proteomics, 17*(12), 2402-2411. doi:10.1074/mcp.RA118.000593

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., . . . Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics, 29*(1), 15-21. doi:10.1093/bioinformatics/bts635

Dunn, J. G., & Weissman, J. S. (2016). Plastid: nucleotide-resolution analysis of next-generation sequencing and genomics data. *BMC Genomics, 17*(1), 958. doi:10.1186/s12864-016-3278-x

Egorov, A. A., Sakharova, E. A., Anisimova, A. S., Dmitriev, S. E., Gladyshev, V. N., & Kulakovskiy, I. V. (2019). svist4get: a simple visualization tool for genomic tracks from sequencing experiments. *BMC Bioinformatics, 20*(1), 113. doi:10.1186/s12859-019-2706-8

Eichhorn, S. W., Guo, H., McGeary, S. E., Rodriguez-Mias, R. A., Shin, C., Baek, D., . . . Bartel, D. P. (2014). mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues. *Mol Cell, 56*(1), 104-115. doi:10.1016/j.molcel.2014.08.028

Erhard, F., Halenius, A., Zimmermann, C., L'Hernault, A., Kowalewski, D. J., Weekes, M. P., . . . Dolken, L. (2018). Improved Ribo-seq enables identification of cryptic translation events. *Nat Methods, 15*(5), 363-366. doi:10.1038/nmeth.4631

Ernlund, A. W., Schneider, R. J., & Ruggles, K. V. (2018). RIVET: comprehensive graphic user interface for analysis and exploration of genome-wide translatomics data. *BMC Genomics, 19*(1), 809. doi:10.1186/s12864-018-5166-z

Fields, A. P., Rodriguez, E. H., Jovanovic, M., Stern-Ginossar, N., Haas, B. J., Mertins, P., . . . Weissman, J. S. (2015). A Regression-Based Analysis of Ribosome-Profiling Data Reveals a Conserved Complexity to Mammalian Translation. *Mol Cell, 60*(5), 816-827. doi:10.1016/j.molcel.2015.11.013

Fluman, N., Navon, S., Bibi, E., & Pilpel, Y. (2014). mRNA-programmed translation pauses in the targeting of E. coli membrane proteins. *Elife, 3*. doi:10.7554/eLife.03440

Fritsch, C., Herrmann, A., Nothnagel, M., Szafranski, K., Huse, K., Schumann, F., . . . Brosch, M. (2012). Genome-wide search for novel human uORFs and N-terminal protein extensions using ribosomal footprinting. *Genome Res, 22*(11), 2208-2218. doi:10.1101/gr.139568.112

Gandin, V., Masvidal, L., Hulea, L., Gravel, S. P., Cargnello, M., McLaughlan, S., . . . Topisirovic, I. (2016). nanoCAGE reveals 5' UTR features that define specific modes of translation of functionally related MTOR-sensitive mRNAs. *Genome Res, 26*(5), 636-648. doi:10.1101/gr.197566.115

Gao, X., Wan, J., Liu, B., Ma, M., Shen, B., & Qian, S. B. (2015). Quantitative profiling of initiating ribosomes in vivo. *Nat Methods, 12*(2), 147-153. doi:10.1038/nmeth.3208

Gerashchenko, M. V., & Gladyshev, V. N. (2017). Ribonuclease selection for ribosome profiling. *Nucleic Acids Res, 45*(2), e6. doi:10.1093/nar/gkw822

Gorochowski, T. E., Chelysheva, I., Eriksen, M., Nair, P., Pedersen, S., & Ignatova, Z. (2019). Absolute quantification of translational regulation and burden using combined sequencing approaches. *Mol Syst Biol, 15*(5), e8719. doi:10.15252/msb.20188719

Guttman, M., Russell, P., Ingolia, N. T., Weissman, J. S., & Lander, E. S. (2013). Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell, 154*(1), 240-251. doi:10.1016/j.cell.2013.06.009

H. Backman TW, & Girke, T. (2016). systemPipeR: NGS workflow and report generation environment. *BMC Bioinformatics, 17*, 388. doi:10.1186/s12859-016-1241-0

Hao, Y., Zhang, L., Niu, Y., Cai, T., Luo, J., He, S., . . . Chen, R. (2018). SmProt: a database of small proteins encoded by annotated coding and non-coding RNA loci. *Brief Bioinform, 19*(4), 636-643. doi:10.1093/bib/bbx005

Hardy, S., Kostantin, E., Wang, S. J., Hristova, T., Galicia-Vazquez, G., Baranov, P. V., . . . Tremblay, M. L. (2019). Magnesium-sensitive upstream ORF controls PRL phosphatase expression to mediate energy metabolism. *Proc Natl Acad Sci U S A, 116*(8), 2925-2934. doi:10.1073/pnas.1815361116

Hinnebusch, A. G. (2014). The scanning mechanism of eukaryotic translation initiation. *Annu Rev Biochem, 83*, 779-812. doi:10.1146/annurev-biochem-060713-035802

Hinnebusch, A. G., Ivanov, I. P., & Sonenberg, N. (2016). Translational control by 5'-untranslated regions of eukaryotic mRNAs. *Science, 352*(6292), 1413-1416. doi:10.1126/science.aad9868

Inada, T. (2013). Quality control systems for aberrant mRNAs induced by aberrant translation elongation and termination. *Biochim Biophys Acta, 1829*(6-7), 634-642. doi:10.1016/j.bbagrm.2013.02.004

Ingolia, N. T., Brar, G. A., Stern-Ginossar, N., Harris, M. S., Talhouarne, G. J., Jackson, S. E., . . . Weissman, J. S. (2014). Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell Rep, 8*(5), 1365-1379. doi:10.1016/j.celrep.2014.07.045

Ingolia, N. T., Ghaemmaghami, S., Newman, J. R., & Weissman, J. S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science, 324*(5924), 218-223. doi:10.1126/science.1168978

Ingolia, N. T., Hussmann, J. A., & Weissman, J. S. (2018). Ribosome Profiling: Global Views of Translation. *Cold Spring Harb Perspect Biol*. doi:10.1101/cshperspect.a032698

Ingolia, N. T., Lareau, L. F., & Weissman, J. S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell, 147*(4), 789-802. doi:10.1016/j.cell.2011.10.002

Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., . . . Linnarsson, S. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat Methods, 11*(2), 163-166. doi:10.1038/nmeth.2772

Ivanov, I. P., Firth, A. E., Michel, A. M., Atkins, J. F., & Baranov, P. V. (2011). Identification of evolutionarily conserved non-AUG-initiated N-terminal extensions in human coding sequences. *Nucleic Acids Res, 39*(10), 4220-4234. doi:10.1093/nar/gkr007

Ivanov, I. P., Shin, B. S., Loughran, G., Tzani, I., Young-Baird, S. K., Cao, C., . . . Dever, T. E. (2018). Polyamine Control of Translation Elongation Regulates Start Site Selection on Antizyme Inhibitor mRNA via Ribosome Queuing. *Mol Cell, 70*(2), 254-264 e256. doi:10.1016/j.molcel.2018.03.015

Iwasaki, S., Floor, S. N., & Ingolia, N. T. (2016). Rocaglates convert DEAD-box protein eIF4A into a sequence-selective translational repressor. *Nature, 534*(7608), 558-561. doi:10.1038/nature17978

Ji, Z. (2018). Rfoot: Transcriptome-Scale Identification of RNA-Protein Complexes from Ribosome Profiling Data. *Curr Protoc Mol Biol, 124*(1), e66. doi:10.1002/cpmb.66

Ji, Z., Song, R., Regev, A., & Struhl, K. (2015). Many lncRNAs, 5'UTRs, and pseudogenes are translated and some are likely to express functional proteins. *Elife, 4*, e08890. doi:10.7554/eLife.08890

Johnstone, T. G., Bazzini, A. A., & Giraldez, A. J. (2016). Upstream ORFs are prevalent translational repressors in vertebrates. *EMBO J, 35*(7), 706-723. doi:10.15252/embj.201592759

Jungreis, I., Lin, M. F., Spokony, R., Chan, C. S., Negre, N., Victorsen, A., . . . Kellis, M. (2011). Evidence of abundant stop codon readthrough in Drosophila and other metazoa. *Genome Res, 21*(12), 2096-2113. doi:10.1101/gr.119974.110

Kiniry, S. J., Michel, A. M., & Baranov, P. V. (2018). The GWIPS-viz Browser. *Curr Protoc Bioinformatics, 62*(1), e50. doi:10.1002/cpbi.50

Kiniry, S. J., O'Connor, P. B. F., Michel, A. M., & Baranov, P. V. (2018). Trips-Viz: a transcriptome browser for exploring Ribo-Seq data. *Nucleic Acids Res*. doi:10.1093/nar/gky842

Kirchner, S., Cai, Z., Rauscher, R., Kastelic, N., Anding, M., Czech, A., . . . Ignatova, Z. (2017). Alteration of protein function by a silent polymorphism linked to tRNA abundance. *PLoS Biol, 15*(5), e2000779. doi:10.1371/journal.pbio.2000779

Kivioja, T., Vaharautio, A., Karlsson, K., Bonke, M., Enge, M., Linnarsson, S., & Taipale, J. (2011). Counting absolute numbers of molecules using unique molecular identifiers. *Nat Methods, 9*(1), 72-74. doi:10.1038/nmeth.1778

Kumari, R., Michel, A. M., & Baranov, P. V. (2018). PausePred and Rfeet: webtools for inferring ribosome pauses and visualizing footprint density from ribosome profiling data. *RNA, 24*(10), 1297-1304. doi:10.1261/rna.065235.117

Kurian, L., Palanimurugan, R., Godderz, D., & Dohmen, R. J. (2011). Polyamine sensing by nascent ornithine decarboxylase antizyme stimulates decoding of its mRNA. *Nature, 477*(7365), 490-494. doi:10.1038/nature10393

Langmead, B., & Nellore, A. (2018). Cloud computing for genomic data analysis and collaboration. *Nat Rev Genet, 19*(5), 325. doi:10.1038/nrg.2018.8

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol, 10*(3), R25. doi:10.1186/gb-2009-10-3-r25

Lareau, L. F., Hite, D. H., Hogan, G. J., & Brown, P. O. (2014). Distinct stages of the translation elongation cycle revealed by sequencing ribosome-protected mRNA fragments. *Elife, 3*, e01257. doi:10.7554/eLife.01257

Larsson, O., Sonenberg, N., & Nadon, R. (2010). Identification of differential translation in genome wide studies. *Proc Natl Acad Sci U S A, 107*(50), 21487-21492. doi:10.1073/pnas.1006821107

Lauria, F., Tebaldi, T., Bernabo, P., Groen, E. J. N., Gillingwater, T. H., & Viero, G. (2018). riboWaltz: Optimization of ribosome P-site positioning in ribosome profiling data. *PLoS Comput Biol, 14*(8), e1006169. doi:10.1371/journal.pcbi.1006169

Lecanda, A., Nilges, B. S., Sharma, P., Nedialkova, D. D., Schwarz, J., Vaquerizas, J. M., & Leidel, S. A. (2016). Dual randomization of oligonucleotides to reduce the bias in ribosome-profiling libraries. *Methods, 107*, 89-97. doi:10.1016/j.ymeth.2016.07.011

Lee, S., Liu, B., Lee, S., Huang, S. X., Shen, B., & Qian, S. B. (2012). Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc Natl Acad Sci U S A, 109*(37), E2424-2432. doi:10.1073/pnas.1207846109

Li, G. W., Oh, E., & Weissman, J. S. (2012). The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature, 484*(7395), 538-541. doi:10.1038/nature10965

Li, W., Wang, W., Uren, P. J., Penalva, L. O. F., & Smith, A. D. (2017). Riborex: fast and flexible identification of differential translation from Ribo-seq data. *Bioinformatics, 33*(11), 1735-1737. doi:10.1093/bioinformatics/btx047

Liu, T. Y., & Song, Y. S. (2016). Prediction of ribosome footprint profile shapes from transcript sequences. *Bioinformatics, 32*(12), i183-i191. doi:10.1093/bioinformatics/btw253

Liu, W., Xiang, L., Zheng, T., Jin, J., & Zhang, G. (2018). TranslatomeDB: a comprehensive database and cloud-based analysis platform for translatome sequencing data. *Nucleic Acids Res, 46*(D1), D206-D212. doi:10.1093/nar/gkx1034

Lobanov, A. V., Heaphy, S. M., Turanov, A. A., Gerashchenko, M. V., Pucciarelli, S., Devaraj, R. R., . . . Gladyshev, V. N. (2017). Position-dependent termination and widespread obligatory frameshifting in Euplotes translation. *Nat Struct Mol Biol, 24*(1), 61-68. doi:10.1038/nsmb.3330

Loughran, G., Jungreis, I., Tzani, I., Power, M., Dmitriev, R. I., Ivanov, I. P., . . . Atkins, J. F. (2018). Stop codon readthrough generates a C-terminally extended variant of the human vitamin D receptor with reduced calcitriol response. *J Biol Chem, 293*(12), 4434-4444. doi:10.1074/jbc.M117.818526

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol, 15*(12), 550. doi:10.1186/s13059-014-0550-8

Malone, B., Atanassov, I., Aeschimann, F., Li, X., Grosshans, H., & Dieterich, C. (2017). Bayesian prediction of RNA translation from ribosome profiling. *Nucleic Acids Res, 45*(6), 2960-2972. doi:10.1093/nar/gkw1350

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet, 17*(1).

McGlincy, N. J., & Ingolia, N. T. (2017). Transcriptome-wide measurement of translation by ribosome profiling. *Methods, 126*, 112-129. doi:10.1016/j.ymeth.2017.05.028

Menschaert, G., Van Criekinge, W., Notelaers, T., Koch, A., Crappe, J., Gevaert, K., & Van Damme, P. (2013). 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*(7), 1780-1790. doi:10.1074/mcp.M113.027540

Michel, A. M., Ahern, A. M., Donohue, C. A., & Baranov, P. V. (2015). GWIPS-viz as a tool for exploring ribosome profiling evidence supporting the synthesis of alternative proteoforms. *Proteomics, 15*(14), 2410-2416. doi:10.1002/pmic.201400603

Michel, A. M., Andreev, D. E., & Baranov, P. V. (2014). Computational approach for calculating the probability of eukaryotic translation initiation from ribo-seq data that takes into account leaky scanning. *BMC Bioinformatics, 15*, 380. doi:10.1186/s12859-014-0380-4

Michel, A. M., & Baranov, P. V. (2013). Ribosome profiling: a Hi-Def monitor for protein synthesis at the genome-wide scale. *Wiley Interdiscip Rev RNA, 4*(5), 473-490. doi:10.1002/wrna.1172

Michel, A. M., Choudhury, K. R., Firth, A. E., Ingolia, N. T., Atkins, J. F., & Baranov, P. V. (2012). Observation of dually decoded regions of the human genome using ribosome profiling data. *Genome Res, 22*(11), 2219-2229. doi:10.1101/gr.133249.111

Michel, A. M., Fox, G., A, M. K., De Bo, C., O'Connor, P. B., Heaphy, S. M., . . . Baranov, P. V. (2014). GWIPS-viz: development of a ribo-seq genome browser. *Nucleic Acids Res, 42*(Database issue), D859-864. doi:10.1093/nar/gkt1035

Michel, A. M., Kiniry, S. J., O'Connor, P. B. F., Mullan, J. P., & Baranov, P. V. (2018). GWIPS-viz: 2018 update. *Nucleic Acids Res, 46*(D1), D823-D830. doi:10.1093/nar/gkx790

Michel, A. M., Mullan, J. P., Velayudhan, V., O'Connor, P. B., Donohue, C. A., & Baranov, P. V. (2016). RiboGalaxy: A browser based platform for the alignment, analysis and visualization of ribosome profiling data. *RNA Biol, 13*(3), 316-319. doi:10.1080/15476286.2016.1141862

Miettinen, T. P., & Bjorklund, M. (2015). Modified ribosome profiling reveals high abundance of ribosome protected mRNA fragments derived from 3' untranslated regions. *Nucleic Acids Res, 43*(2), 1019-1034. doi:10.1093/nar/gku1310

Mohammad, F., Woolstenhulme, C. J., Green, R., & Buskirk, A. R. (2016). Clarifying the Translational Pausing Landscape in Bacteria by Ribosome Profiling. *Cell Rep, 14*(4), 686-694. doi:10.1016/j.celrep.2015.12.073

Ndah, E., Jonckheere, V., Giess, A., Valen, E., Menschaert, G., & Van Damme, P. (2017). REPARATION: ribosome profiling assisted (re-)annotation of bacterial genomes. *Nucleic Acids Res, 45*(20), e168. doi:10.1093/nar/gkx758

O'Connor, P. B., Andreev, D. E., & Baranov, P. V. (2016). Comparative survey of the relative impact of mRNA features on local ribosome profiling read density. *Nat Commun, 7*, 12915. doi:10.1038/ncomms12915

O'Connor, P. B., Li, G. W., Weissman, J. S., Atkins, J. F., & Baranov, P. V. (2013). rRNA:mRNA pairing alters the length and the symmetry of mRNA-protected fragments in ribosome profiling experiments. *Bioinformatics, 29*(12), 1488-1491. doi:10.1093/bioinformatics/btt184

Oertlin, C., Lorent, J., Murie, C., Furic, L., Topisirovic, I., & Larsson, O. (2019). Generally applicable transcriptome-wide analysis of translation using anota2seq. *Nucleic Acids Res*. doi:10.1093/nar/gkz223

Olexiouk, V., Van Criekinge, W., & Menschaert, G. (2018). An update on sORFs.org: a repository of small ORFs identified by ribosome profiling. *Nucleic Acids Res, 46*(D1), D497-D502. doi:10.1093/nar/gkx1130

Olshen, A. B., Hsieh, A. C., Stumpf, C. R., Olshen, R. A., Ruggero, D., & Taylor, B. S. (2013). Assessing gene-level translational control from ribosome profiling. *Bioinformatics, 29*(23), 2995-3002. doi:10.1093/bioinformatics/btt533

Perkins, P., Mazzoni-Putman, S., Stepanova, A., Alonso, J., & Heber, S. (2019). RiboStreamR: a web application for quality control, analysis, and visualization of Ribo-seq data. *BMC Genomics, 20*(Suppl 5), 422. doi:10.1186/s12864-019-5700-7

Pop, C., Rouskin, S., Ingolia, N. T., Han, L., Phizicky, E. M., Weissman, J. S., & Koller, D. (2014). Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation. *Mol Syst Biol, 10*, 770. doi:10.15252/msb.20145524

Popa, A., Lebrigand, K., Paquet, A., Nottet, N., Robbe-Sermesant, K., Waldmann, R., & Barbry, P. (2016). RiboProfiling: a Bioconductor package for standard Ribo-seq pipeline processing. *F1000Res, 5*, 1309. doi:10.12688/f1000research.8964.1

Quackenbush, J. (2002). Microarray data normalization and transformation. *Nature Genetics, 32*, 496. doi:10.1038/ng1032

Raj, A., Wang, S. H., Shim, H., Harpak, A., Li, Y. I., Engelmann, B., . . . Pritchard, J. K. (2016). Thousands of novel translated open reading frames in humans inferred by ribosome footprint profiling. *Elife, 5*. doi:10.7554/eLife.13328

Rajput, B., Pruitt, K. D., & Murphy, T. D. (2019). RefSeq curation and annotation of stop codon recoding in vertebrates. *Nucleic Acids Res, 47*(2), 594-606. doi:10.1093/nar/gky1234

Rathore, A., Chu, Q., Tan, D., Martinez, T. F., Donaldson, C. J., Diedrich, J. K., . . . Saghatelian, A. (2018). MIEF1 Microprotein Regulates Mitochondrial Translation. *Biochemistry, 57*(38), 5564-5575. doi:10.1021/acs.biochem.8b00726

Reid, D. W., Xu, D., Chen, P., Yang, H., & Sun, L. (2017). Integrative analyses of translatome and transcriptome reveal important translational controls in brown and white adipose regulated by microRNAs. *Sci Rep, 7*(1), 5681. doi:10.1038/s41598-017-06077-3

Reixachs-Solé, M., Ruiz-Orera, J., Alba, M. M., & Eyras, E. (2019). Ribosome profiling at isoform level reveals an evolutionary conserved impact of differential splicing on the proteome. doi:10.1101/582031

Reuter, K., Biehl, A., Koch, L., & Helms, V. (2016). PreTIS: A Tool to Predict Non-canonical 5' UTR Translational Initiation Sites in Human and Mouse. *PLoS Comput Biol, 12*(10), e1005170. doi:10.1371/journal.pcbi.1005170

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics, 26*(1), 139-140. doi:10.1093/bioinformatics/btp616

Rodriguez, J. M., Rodriguez-Rivas, J., Di Domenico, T., Vazquez, J., Valencia, A., & Tress, M. L. (2018). APPRIS 2017: principal isoforms for multiple gene sets. *Nucleic Acids Res, 46*(D1), D213-D217. doi:10.1093/nar/gkx997

Schueren, F., Lingner, T., George, R., Hofhuis, J., Dickel, C., Gartner, J., & Thoms, S. (2014). Peroxisomal lactate dehydrogenase is generated by translational readthrough in mammals. *Elife, 3*, e03640. doi:10.7554/eLife.03640

Sharipov, R. N. Y., I.S; Kondrakhin, Y.V; Volkova, O.A. (2014). RiboSeqDB – a repository of selected human and mouse ribosome footprint and RNA-seq data. *Virtual Biology, 2*

Shirokikh, N. E., & Preiss, T. (2018). Translation initiation by cap-dependent ribosome recruitment: Recent insights and open questions. *Wiley Interdiscip Rev RNA, 9*(4), e1473. doi:10.1002/wrna.1473

Smith, T., Heger, A., & Sudbery, I. (2017). UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res, 27*(3), 491-499. doi:10.1101/gr.209601.116

Somogyi, P., Jenner, A. J., Brierley, I., & Inglis, S. C. (1993). Ribosomal pausing during translation of an RNA pseudoknot. *Mol Cell Biol, 13*(11), 6931-6940. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/8413285

Spealman, P., Naik, A. W., May, G. E., Kuersten, S., Freeberg, L., Murphy, R. F., & McManus, J. (2018). Conserved non-AUG uORFs revealed by a novel regression analysis of ribosome profiling data. *Genome Res, 28*(2), 214-222. doi:10.1101/gr.221507.117

Stefan, D. (2016). RiboPip. Retrieved from https://github.com/stepf/RiboPip

Tanaka, M., Sotta, N., Yamazumi, Y., Yamashita, Y., Miwa, K., Murota, K., . . . Fujiwara, T. (2016). The Minimum Open Reading Frame, AUG-Stop, Induces Boron-Dependent Ribosome Stalling and mRNA Degradation. *Plant Cell, 28*(11), 2830-2849. doi:10.1105/tpc.16.00481

Tenson, T., & Ehrenberg, M. (2002). Regulatory nascent peptides in the ribosomal tunnel. *Cell, 108*(5), 591-594. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/11893330

Tholstrup, J., Oddershede, L. B., & Sorensen, M. A. (2012). mRNA pseudoknot structures can act as ribosomal roadblocks. *Nucleic Acids Res, 40*(1), 303-313. doi:10.1093/nar/gkr686

Thomson, D. J. (1982). Spectrum estimation and harmonic analysis. *Proceedings of the IEEE, 70*(9), 1055-1096. doi:10.1109/PROC.1982.12433

Tsai, C. J., Sauna, Z. E., Kimchi-Sarfaty, C., Ambudkar, S. V., Gottesman, M. M., & Nussinov, R. (2008). Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima. *J Mol Biol, 383*(2), 281-291. doi:10.1016/j.jmb.2008.08.012

Tunney, R., McGlincy, N. J., Graham, M. E., Naddaf, N., Pachter, L., & Lareau, L. F. (2018). Accurate design of translational output by a neural network model of ribosome distribution. *Nat Struct Mol Biol, 25*(7), 577-582. doi:10.1038/s41594-018-0080-2

Tzani, I., Ivanov, I. P., Andreev, D. E., Dmitriev, R. I., Dean, K. A., Baranov, P. V., . . . Loughran, G. (2016). Systematic analysis of the PTEN 5' leader identifies a major AUU initiated proteoform. *Open Biol, 6*(5). doi:10.1098/rsob.150203

Van Damme, P., Gawron, D., Van Criekinge, W., & Menschaert, G. (2014). N-terminal proteomics and ribosome profiling provide a comprehensive view of the alternative translation initiation landscape in mice and men. *Mol Cell Proteomics, 13*(5), 1245-1261. doi:10.1074/mcp.M113.036442

Vanderperre, B., Lucier, J. F., Bissonnette, C., Motard, J., Tremblay, G., Vanderperre, S., . . . Roucou, X. (2013). Direct detection of alternative open reading frames translation products in human significantly expands the proteome. *PLoS One, 8*(8), e70698. doi:10.1371/journal.pone.0070698

Verbruggen, S., & Menschaert, G. (2018). mQC: A post-mapping data exploration tool for ribosome profiling. *Comput Methods Programs Biomed*. doi:10.1016/j.cmpb.2018.10.018

Verbruggen, S., Ndah, E., Van Criekinge, W., Gessulat, S., Kuster, B., Wilhelm, M., . . . Menschaert, G. (2019). PROTEOFORMER 2.0: further developments in the ribosome profiling-assisted proteogenomic hunt for new proteoforms. *Mol Cell Proteomics*. doi:10.1074/mcp.RA118.001218

Wang, H., McManus, J., & Kingsford, C. (2016). Isoform-level ribosome occupancy estimation guided by transcript abundance with Ribomap. *Bioinformatics, 32*(12), 1880-1882. doi:10.1093/bioinformatics/btw085

Wang, H., McManus, J., & Kingsford, C. (2017). Accurate Recovery of Ribosome Positions Reveals Slow Translation of Wobble-Pairing Codons in Yeast. *J Comput Biol, 24*(6), 486-500. doi:10.1089/cmb.2016.0147

Wang, H., Yang, L., Wang, Y., Chen, L., Li, H., & Xie, Z. (2018). RPFdb v2.0: an updated database for genome-wide information of translated mRNA generated from ribosome profiling. *Nucleic Acids Res*. doi:10.1093/nar/gky978

Weatheritt, R. J., Sterne-Weiler, T., & Blencowe, B. J. (2016). The ribosome-engaged landscape of alternative splicing. *Nat Struct Mol Biol, 23*(12), 1117-1123. doi:10.1038/nsmb.3317

Woolstenhulme, C. J., Guydosh, N. R., Green, R., & Buskirk, A. R. (2015). High-precision analysis of translational pausing by ribosome profiling in bacteria lacking EFP. *Cell Rep, 11*(1), 13-21. doi:10.1016/j.celrep.2015.03.014

Wu, C. C., Zinshteyn, B., Wehner, K. A., & Green, R. (2019). High-Resolution Ribosome Profiling Defines Discrete Ribosome Elongation States and Translational Regulation during Cellular Stress. *Mol Cell*. doi:10.1016/j.molcel.2018.12.009

Wu, W. S., Jiang, Y. X., Chang, J. W., Chu, Y. H., Chiu, Y. H., Tsao, Y. H., . . . Tseng, J. T. (2018). HRPDviewer: human ribosome profiling data viewer. *Database (Oxford), 2018*. doi:10.1093/database/bay074

Xiao, Z., Huang, R., Xing, X., Chen, Y., Deng, H., & Yang, X. (2018). De novo annotation and characterization of the translatome with ribosome profiling data. *Nucleic Acids Res, 46*(10), e61. doi:10.1093/nar/gky179

Xiao, Z., Zou, Q., Liu, Y., & Yang, X. (2016). Genome-wide assessment of differential translations with ribosome profiling data. *Nat Commun, 7*, 11194. doi:10.1038/ncomms11194

Xie, C., Bekpen, C., Kunzel, S., Keshavarz, M., Krebs-Wheaton, R., Skrabar, N., . . . Tautz, D. (2019). A de novo evolved gene in the house mouse regulates female pregnancy cycles. *Elife, 8*. doi:10.7554/eLife.44392

Xie, S. Q., Nie, P., Wang, Y., Wang, H., Li, H., Yang, Z., . . . Xie, Z. (2016). RPFdb: a database for genome wide information of translated mRNA generated from ribosome profiling. *Nucleic Acids Res, 44*(D1), D254-258. doi:10.1093/nar/gkv972

Xu, Z., Hu, L., Shi, B., Geng, S., Xu, L., Wang, D., & Lu, Z. J. (2018a). Ribosome elongating footprints denoised by wavelet transform comprehensively characterize dynamic cellular translation events. *Nucleic Acids Res, 46*(18), e109. doi:10.1093/nar/gky533

Xu, Z., Hu, L., Shi, B., Geng, S., Xu, L., Wang, D., & Lu, Z. J. (2018b). Ribosome elongating footprints denoised by wavelet transform comprehensively characterize dynamic cellular translation events. *Nucleic Acids Res*. doi:10.1093/nar/gky533

Yordanova, M. M., Loughran, G., Zhdanov, A. V., Mariotti, M., Kiniry, S. J., O'Connor, P. B. F., . . . Baranov, P. V. (2018). AMD1 mRNA employs ribosome stalling as a mechanism for molecular memory formation. *Nature, 553*(7688), 356-360. doi:10.1038/nature25174

Zhang, P., He, D., Xu, Y., Hou, J., Pan, B. F., Wang, Y., . . . Chen, Y. (2017). Genome-wide identification and differential analysis of translational initiation. *Nat Commun, 8*(1), 1749. doi:10.1038/s41467-017-01981-8

Zhang, S., Hu, H., Zhou, J., He, X., Jiang, T., & Zeng, J. (2017). Analysis of Ribosome Stalling and Translation Elongation Dynamics by Deep Learning. *Cell Syst, 5*(3), 212-220 e216. doi:10.1016/j.cels.2017.08.004

Zhao, D., Baez, W., Fredrick, K., & Bundschuh, R. (2018). RiboProP: A Probabilistic Ribosome Positioning Algorithm for Ribosome Profiling. *Bioinformatics*. doi:10.1093/bioinformatics/bty854

Zhong, Y., Karaletsos, T., Drewe, P., Sreedharan, V. T., Kuo, D., Singh, K., . . . Ratsch, G. (2017). RiboDiff: detecting changes of mRNA translation efficiency from ribosome footprints. *Bioinformatics, 33*(1), 139-141. doi:10.1093/bioinformatics/btw585

Further Reading

[Please insert any further reading/resources here]

Figure Legends:

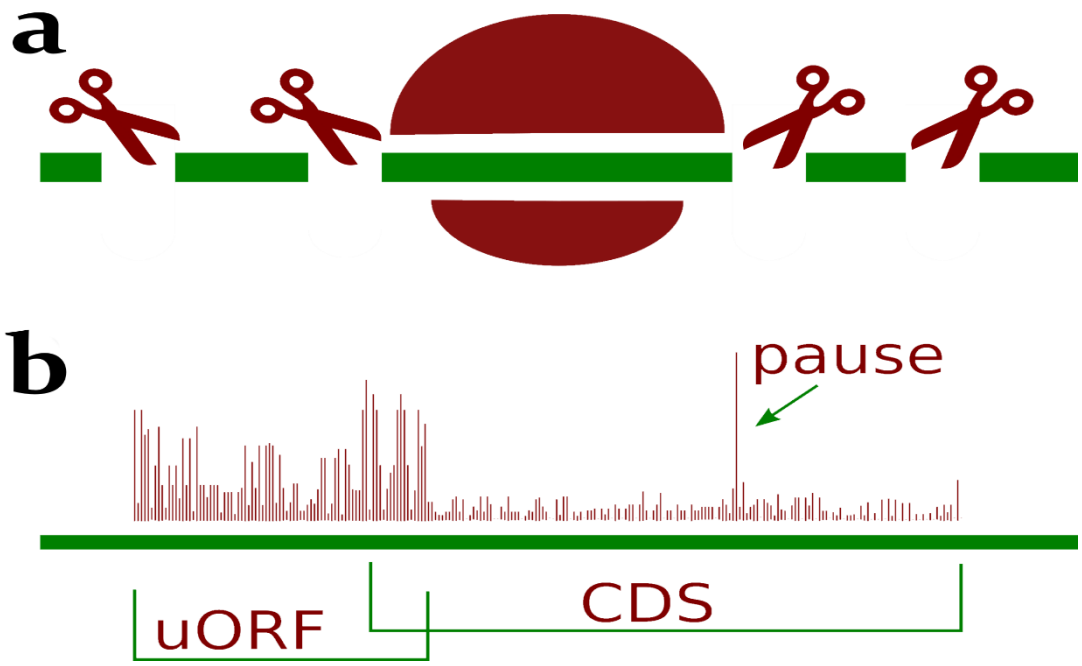

FIGURE 1| The principle of ribosome profiling. (a) The ribosome protects mRNA from nuclease digestion. The sequences of the protected fragments (footprints) constitute ribosome profiling data. (b) A schematic example of a ribosome footprints density plot (ribosome profile). It shows positions of ribosome decoding centres (brown columns) inferred from sequences of ribosome footprints along an RNA transcript (green bar). The height of the columns reflects the number of footprints matching the corresponding mRNA position. The density suggests the efficient translation of an upstream Open Reading Frame (uORF) overlapping the annotated protein coding region (CDS) and the presence of a ribosome pause site in the CDS.

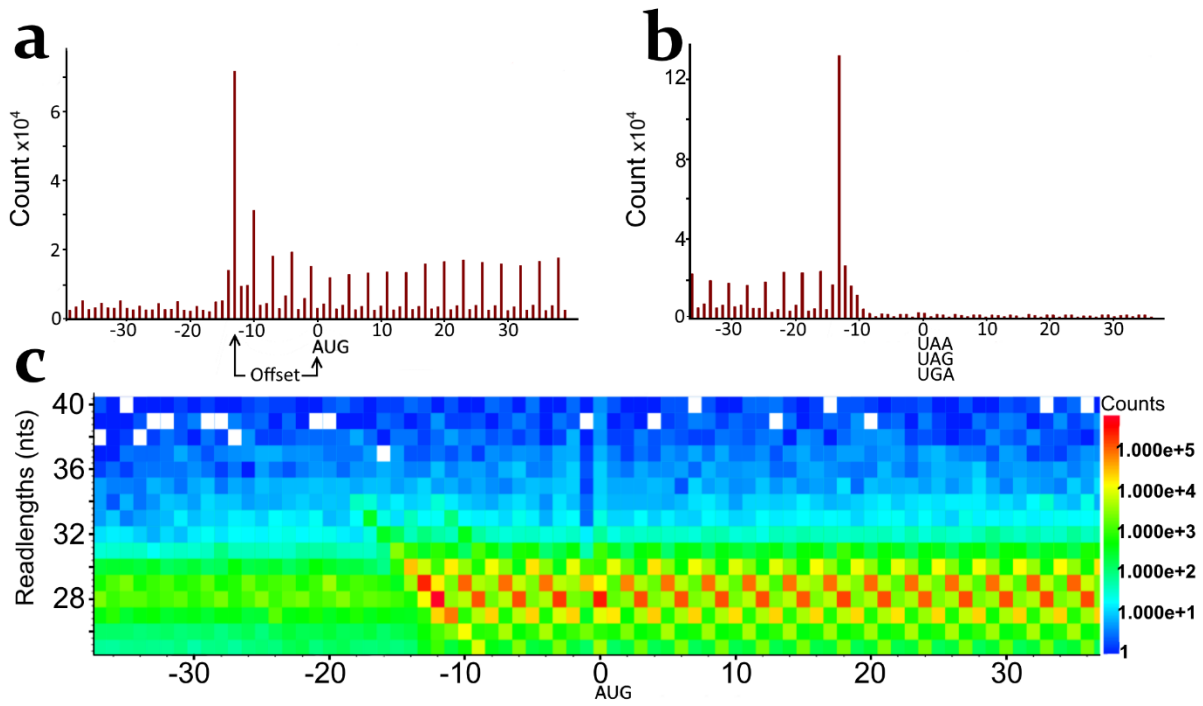

FIGURE 2| Examples of metagene profiles. (a) The profile was created by aggregating Ribo-Seq counts from a region surrounding the annotated start codon (zero coordinate) of every gene for a single read length. This example shows the positions of footprint 5' ends, but 3' ends may also be used. Since initiation is slower than elongation, a peak of footprint density is expected at the start codon. Thus the location of the 5' end peak density indicates the distance between footprints 5' ends and ribosome P-site codon where tRNA-Met_i is being incorporated (offset). (b) Same as (a) but relative to annotated stop codons (zero coordinate). A drop of footprint density is observed upstream of the stop. (c) A start codon metagene profile constructed as a heatmap has the advantage of displaying multiple read lengths simultaneously. It can be seen that the distance between 5' ends and P-site codons vary depending on read lengths suggesting that different offsets should be applied to the reads depending on their length.

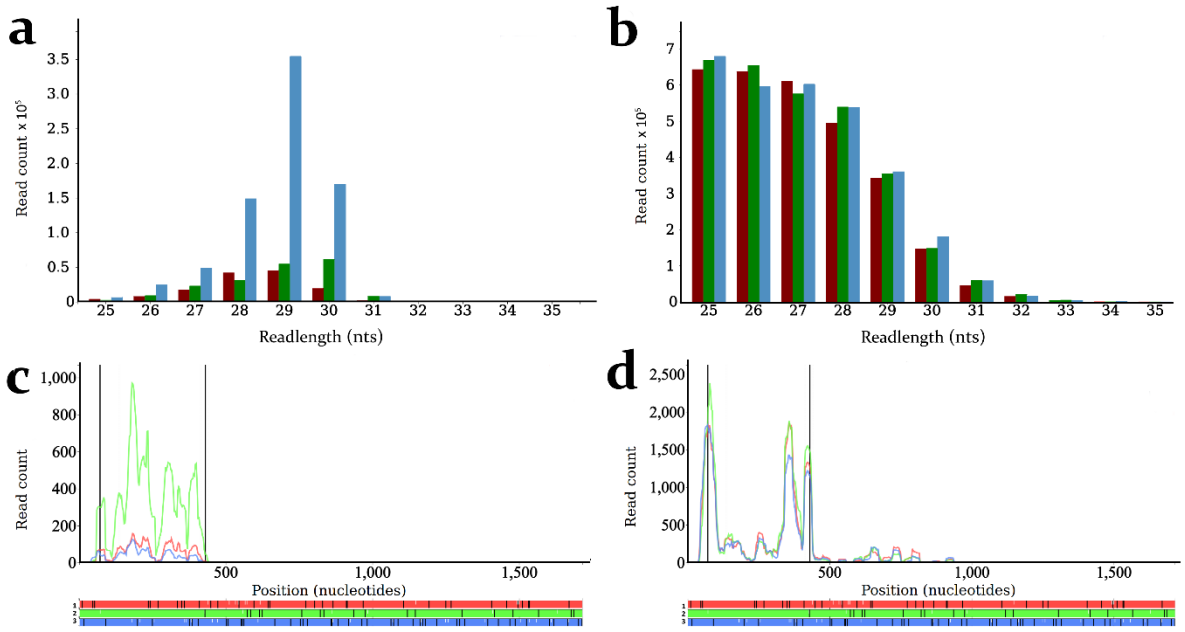

FIGURE 3| Assessment of ribosome profiling data quality (a, b) Triplet periodicity plots that show the number of footprints aligning to one of the three subcodon positions (differentially colored) for each subcodon position. (a) An example of good quality data showing strong periodicity and desirable read length distribution. (b) An example of data showing no triplet periodicity and an unexpected read length distribution. (c, d) Sub-codon ribosome profile of an ENSEMBL transcript expressed from the human *B2M* locus visualized with Trips-Viz. The ORF plot at the bottom shows three reading frames (differentially colored) with white dashes for AUG codons and black dashes for stops. The annotated CDS is demarked by the vertical black lines in the main plot and corresponds to the second reading frame. The footprint density is shown separately depending on the sub-codon phase of the aligned reads as curves that are colored to match the color of the supported reading frames. The reading frame detection is possible in (c), but not in (d) which correspond to (a) and (b) respectively. In addition in (c) the vast majority of reads map entirely within the CDS, while in (d) there are reads which map to the 3' trailer region that are unlikely to be derived from translating ribosomes. For the source of the data see text.

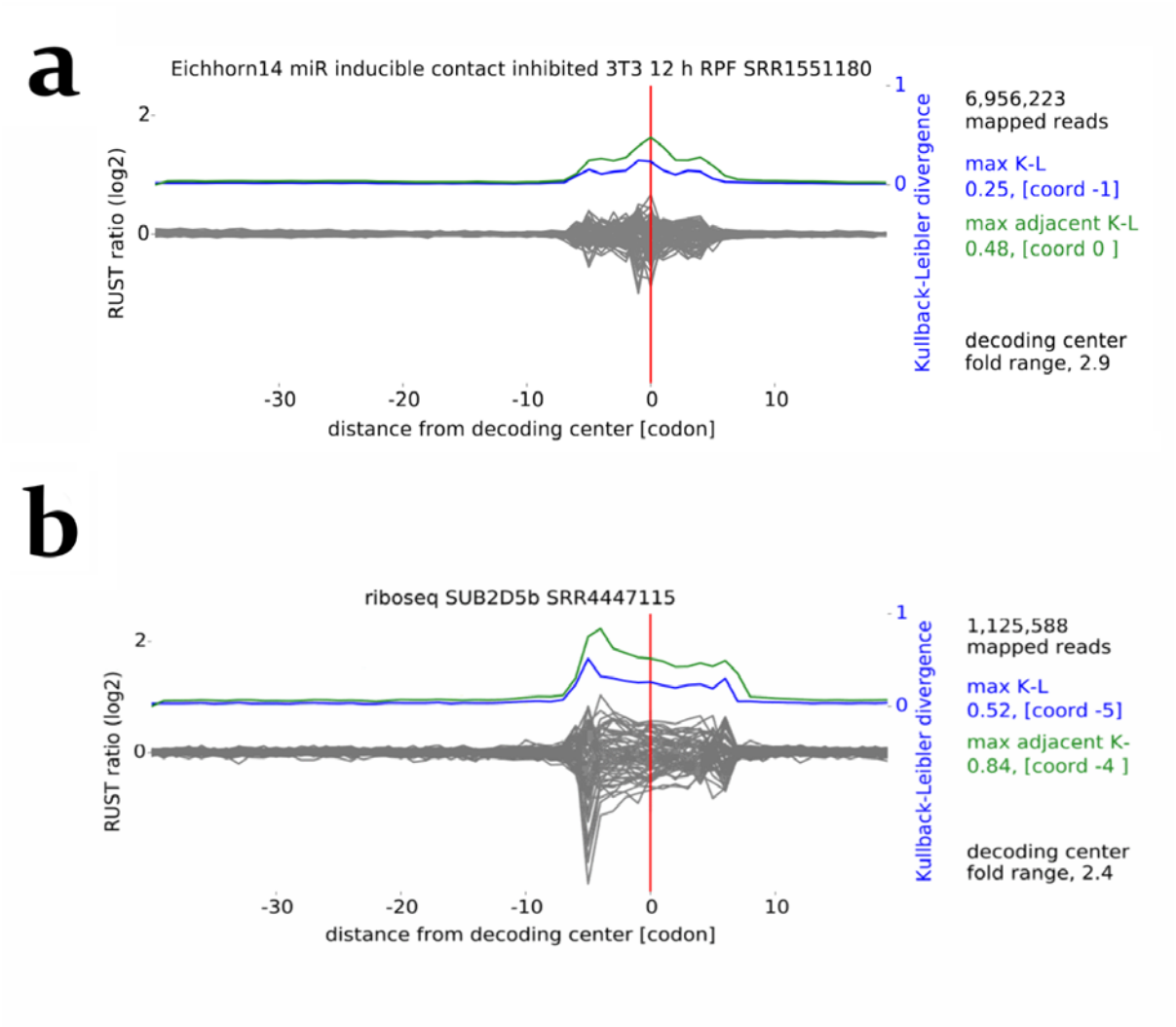

FIGURE 4| RUST metafootprint profiles that can be used for the assessment of sequencing biases that are manifested by high relative entropy (measured as Kullback-Leibler divergence) at the ends of footprints. The decoding center of the ribosome (A-site) is denoted by the vertical red line. The blue line represents Kullback-Leibler divergence at an individual codon level. The green line represents Kullback-Leibler divergence for adjacent codons. In the absence of sequencing biases the Kullback-Leibler divergence is expected to be the highest at the decoding center. (a) A dataset with low sequencing bias. (b) A dataset with high sequencing bias at the 5' ends of footprints. For the data sources see the text.

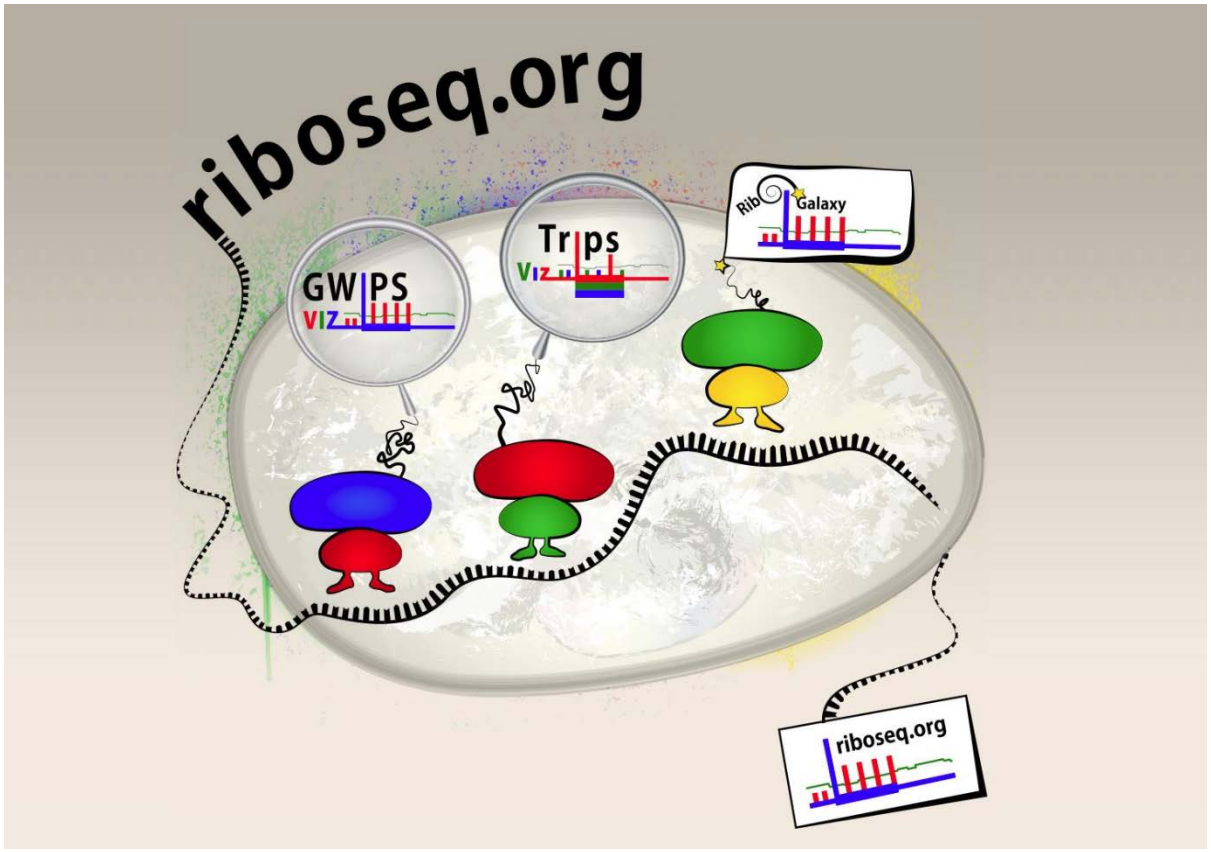

FIGURE 5| The RiboSeq.Org web portal serves as an entry point to GWIPS-viz, Trips-Viz and RiboGalaxy. GWIPS-viz provides visualizations of publicly available ribosome footprints mapped to several genomes. Trips-Viz offers rich functionality for the analysis of public and user generated data aligned to transcriptomes. RiboGalaxy provides cross-platform graphical interface for the tools initially written as command line software.