Review

Alexander Bartholomäus, Cristian Del Campo and Zoya Ignatova*

Mapping the non-standardized biases of ribosome profiling

DOI 10.1515/hsz-2015-0197
Received June 20, 2015; accepted August 25, 2015; previously published online August 29, 2015

Abstract: Ribosome profiling is a new emerging technology that uses massively parallel amplification of ribosome-protected fragments and next-generation sequencing to monitor translation *in vivo* with codon resolution. Studies using this approach provide insightful views on the regulation of translation on a global cell-wide level. In this review, we compare different experimental set-ups and current protocols for sequencing data analysis. Specifically, we review the pitfalls at some experimental steps and highlight the importance of standardized protocol for sample preparation and data processing pipeline, at least for mapping and normalization.

Keywords: deep sequencing; mapping; normalization; nucleases; ribosome footprints; translation.

Introduction

At any given time, the amounts and types of proteins reflect the functional status of the cell. The protein composition is a balance between protein synthesis and degradation. On the synthesis side, protein production is controlled at the level of transcription and translation and the messenger RNA (mRNA) is the connecting entity between these two processes. Moreover, emerging evidence suggests that the mRNA open-reading frame bears far more information than just the amino acid sequence of the synthesized protein. Codon choice to encode one amino acid (Plotkin and Kudla, 2011), tRNA modifications (Nedialkova and Leidel, 2015; Tyagi and Pedrioli, 2015) or secondary structures (Wen et al., 2008; Chen et al., 2013) modulate the local speed at which mRNA is translated and link it to protein biogenesis or stress response. Recent developments in the next-generation sequencing (NGS) technologies revealed additional layers embedded in the mRNA to regulate its translatability and consequently the downstream processes in protein biogenesis including cotranslational folding, insertion into membranes and interactions with auxiliary factors (Kramer et al., 2009; Zhang and Ignatova, 2011; Pechmann et al., 2014). Specifically, a recent twist of the NGS technologies to capture translating ribosomes, named ribosome profiling (Ingolia et al., 2009), has significantly advanced our understanding on translation regulation in various organisms [reviewed in (Ingolia, 2014)]. Ribosome profiling is based on high-throughput sequencing of ribosome-protected RNA fragments, or ribosomal 'footprints', which specifically report on the position of the translating ribosomes with a nucleotide resolution (Ingolia et al., 2009). A growing body of published literature illustrates the power of this approach to unravel new aspects on translation regulation, for example identification of extensive upstream initiation at non-AUG codons in eukaryotes (Ingolia et al., 2009, 2011; Fritsch et al., 2012; Lee et al., 2012) and specific regulation of the stress response at translation level (Liu et al., 2013; Shalgi et al., 2013; Andreev et al., 2015). Further development of the profiling technology to isolate a fraction of ribosomes that are involved in specific cellular processes revealed new insights into the localized protein synthesis in yeast (Jan et al., 2014) or the interaction with a trigger factor, an auxiliary factor facilitating cotranslational folding in bacteria (Oh et al., 2011).

Without doubt, ribosome profiling is a powerful technology to address various aspects of translation regulation on a genome-wide scale, and several excellent reviews summarize the power of this technology (Morris, 2009; Kuersten et al., 2013; Michel and Baranov, 2013; Ingolia, 2014). However, this approach is relatively young, with steadily evolving experimental protocol and a non-standardized platform for data analysis. The pace of

***Corresponding author: Zoya Ignatova,** Biochemistry and Molecular Biology, Department of Chemistry, University of Hamburg, Martin-Luther-King-Platz 6, D-20146 Hamburg, Germany, e-mail: zoya.ignatova@chemie.uni-hamburg.de
Alexander Bartholomäus and Cristian Del Campo: Biochemistry and Molecular Biology, Department of Chemistry, University of Hamburg, Martin-Luther-King-Platz 6, D-20146 Hamburg, Germany

exploration creates some difficulties in comparing results produced in different laboratories. In addition, different approaches to analyze the data disclose variations in their interpretation (Gerashchenko and Gladyshev, 2014). Here, we focus on the ribosome profiling procedure and data analyses and critically review the biases of the various steps in the profiling protocol as a potential source of variation. We also provide examples on how variations in the ribosome profiling procedure put restrictions on the downstream analysis and determine the information that can be extracted from the data. We suggest standardizing ribosome profiling protocol and adjusting only a step (or few steps) depending on the specific scientific question.

Isolation of intact translating ribosomes

At the core of ribosome profiling is a nuclease digestion of mRNA unprotected by the ribosome and recovering ribosome-protected mRNA fragments (i.e. ribosome footprints) (Steitz, 1969) and their conversion into a DNA library that is further analyzed by deep sequencing (Ingolia et al., 2009) (Figure 1). Thus, this approach maps the position of the translating ribosomes on each mRNA and provides a snap-shot of translation.

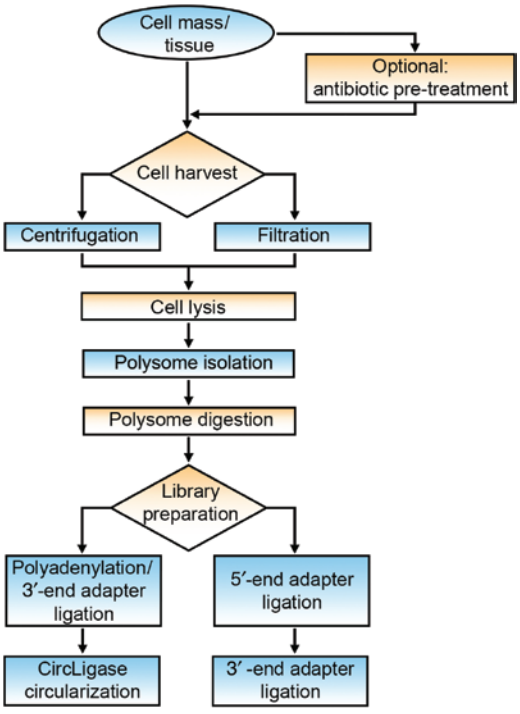

Figure 1: Flow-chart of isolation of intact ribosome-mRNA complexes and library preparation for the ribosome profiling experiment.
Crucial steps at which specific decisions need to be taken are color-coded in orange. Detailed knowledge of the bias of each of those procedures is essential for the careful interpretation of the sequencing data.

Harvesting the cells and antibiotic pretreatment

The most delicate step in the sample preparation is the isolation of intact ribosome-mRNA complexes. Ideally, the isolation procedure should faithfully freeze the translating ribosomes and avoid conditions that stimulate ribosomal drop-off and, most importantly, ribosome relocation on the mRNA during the sample processing.

Early in the development of the ribosome profiling approach, cells were pre-incubated with elongation inhibitors (mainly chloramphenicol for bacteria and cychlohexamide for eukaryotes) to inhibit further movement of the elongating ribosomes along the mRNA (Ingolia et al., 2009). The antibiotic treatment markedly affects the coverage profiles and introduces some bias in the results; the elongation inhibitors do not uniformly stall elongating ribosomes but rather show a codon-dependent mode of action (Orelle et al., 2013). Cycloheximide also allows one complete translocation cycle before blocking the ribosome (Pestova and Hellen, 2003; Schneider-Poetsch et al., 2010) and thus diffuses the read-out when determining codon-dependent stalling (Nedialkova and Leidel, 2015), while non-antibiotic treated cells deliver much sharper pause sites corresponding to rare codons (Pelechano et al., 2015).

In addition, a broad cumulative peak downstream of the start codon has been seen in the earlier profiling papers that use elongation inhibitors and interpreted as slow initiation (Ingolia et al., 2009). The initial peak, albeit still present, significantly decreases when cell mass is flash-frozen and elongation inhibiters are omitted (Guydosh and Green, 2014; Lareau et al., 2014).

The disproportionately high accumulation of reads at initiation is rather an artifact of the antibiotic pretreatment (Becker et al., 2013) and results from inhibition of translation elongation with ongoing initiation (Ingolia et al., 2011). The antibiotic does not immediately reach the threshold of complete inhibition of elongation; instead its concentration increases gradually in the cell (Gerashchenko and Gladyshev, 2014). Hence, upon treatment, some initiating ribosomes continue into the elongation cycle until they encounter the drug, which results in an excess of ribosomal footprints over the first five to ten

codons from the coding sequence (Gerashchenko and Gladyshev, 2014). Additionally, an 80S ribosome stalled in the proximity of the start codon will prevent any subsequent scanning ribosome from reaching the initiation codon, which may result in an apparent stalling at an upstream open-reading frame (uORF). Thus, an initiation site with mediocre context in uORF will be occupied because of the highly efficient but blocked downstream start site (Jackson and Standart, 2015), which may lead to an erroneous interpretation of alternative uORF-induced initiation.

Careful consideration of the effect of antibiotics on ribosome coverage offers little support that the large number of genes with uORFs is involved in shaping the resistance to oxidative stress (Gerashchenko and Gladyshev, 2014). Ribosome profiling without antibiotics prior to cell harvesting revealed that translation of only a small fraction of uORF-bearing mRNA was refractory to oxidative stress (Andreev et al., 2015). Elongation inhibitors added prior to harvesting the ribosome-mRNA complexes alter the distribution of reads in the cumulative ribosome profiles (namely, the aligned and averaged profiles of many genes). For example, emetine-stalled elongating ribosomes give slightly longer fragments than those isolated from cycloheximide-treated mammalian cells, suggesting that various antibiotics stabilize different ribosome conformation (Ingolia et al., 2011). Conversely, drug pretreatment may eliminate some features of biological importance in the cumulative ribosome profiles. For example, antibiotic pretreatment in mammalian cells eliminates the ribosomal peak at the end of the open-reading frames, which is observed in untreated cells (Ingolia et al., 2011).

The most widely applied cell harvesting procedure involves rapid cooling of the cell suspension and centrifugation (Becker et al., 2013) (Figure 1). Bacteria are cooled by pouring the cell suspension over crushed ice, while eukaryotic (mammalian) cells cultured in monolayer are re-suspended in ice-cold PBS supplemented with elongation inhibitor and immediately pelleted by centrifugation (Guo et al., 2010). Tissues are usually flash-frozen and grinded in the lysis buffer supplemented with elongation inhibitor (Gonzalez et al., 2014). An alternative approach for harvesting of cells growing in suspension is a rapid filtration of the cells in a prewarmed glass nitrocellulose filtration system and flash-freezing the membrane with the cells (Figure 1). So far, this filtration approach has been mainly used in unicellular organisms (yeast and *Escherichia coli*, for example) (Ingolia et al., 2009; Oh et al., 2011; Li et al., 2012). Both harvesting protocols show good reproducibility between biological replicates (r=0.99, Pearson correlation coefficient) (Becker et al., 2013). Importantly, however, the RPF accumulation at native stalling sites, e.g. SecM and TnaC, is higher using the filtration harvesting (Becker et al., 2013). Most likely, the filtration approach compared to the centrifugation is less susceptible to variations and faithfully halts the translating ribosomes. Still, harvesting by centrifugation might be the only option for cells that cannot be rapidly filtered. However, it is important to perform it as quickly as possible using pre-chilled devices.

In summary, the procedure for isolation of ribosome-mRNA complexes is of crucial importance. While drug pretreatment may not influence differential expression analysis, as the expression of each gene is compared under two different conditions with an otherwise uniform protocol, the use of elongation inhibitors or the harvesting procedure may alter the interpretation of position-specific information.

Cell lysis

Similar to the cell harvesting procedure, the aim at this step is to recover the ribosome-mRNA complexes with minimal losses from ribosomal dissociation (or drop-off) and mRNA degradation. The composition of the lysis buffer is optimized to stabilize the ribosome-mRNA complexes with high concentration of magnesium (between 5 and 20 mM) and an additional salt, such as KCl or NaCl and NH4Cl.

The isolation of intact polysomes is a procedure established in early ribosome research and is still applied today almost unchanged (Wettstein et al., 1963; Dresden and Hoagland, 1965). The composition of the lysis buffer underwent several variations. However, some components of the lysis buffer, if overdosed, may distort the ribosome profiles. For example, high NaCl concentration decreases the monosome peak and enhances the fraction of dissociated ribosomal subunits (Becker et al., 2013); high salt concentration increases the fraction of vacant ribosomes that are not engaged in translation (Blobel and Sabatini, 1971) and consequently decreases the number of RPF. Magnesium stabilizes the translating ribosomes (Ron et al., 1968) and at high concentrations freezes the conformational changes in the bacterial ribosome (Blanchard et al., 2004). Moreover, high magnesium concentration induces folding of the mRNA, which hinders the subsequent nucleolytic digestion (Andreev et al., 2015). Lowering the magnesium concentration from 15 mM to 5 mM greatly improves the codon positioning of the footprints

and the resolution of the ribosome profiling (Ingolia et al., 2012). Also, low magnesium conditions permit conformational flexibility of the ribosome and create heterogeneity in the length ribosomal footprints (Lareau et al., 2014); the variant ribosomal footprints are informative on distinct stages of the translating ribosome during the elongation cycle.

The lysis buffer also contains an elongation inhibitor to additionally stabilize the ribosome-mRNA complexes during sample processing. The binding kinetics of the antibiotic when present in the cell lysis is rapid compared to the diffusion-driven process of antibiotic enrichment in intact cells during the pretreatment procedure. Generation of cell extracts from *Saccharomyces* cells in the cycloheximide-containing lysis buffer faithfully halted the ribosomes along the mRNA with no distortion (Guydosh and Green, 2014). Cycloheximide should be preferred over alternative substances that stabilize eukaryotic ribosome-mRNA complexes, e.g. the non-hydrolyzable GTP analog GMP-PNP, as they slightly increase the size of the ribosome footprints (Guydosh and Green, 2014). Although such studies with bacterial elongation inhibitors are missing, it can be expected that their mode of action will be similar to that of the cycloheximide when added to the lysis buffer.

Along with variations in the composition of the lysis buffer, the lysis procedure also varies. In general, despite the presence of components stabilizing the ribosome-mRNA complexes (e.g. elongation inhibitors, magnesium) to avoid ribosomal reallocation or dissociation, lysis is usually carried out at low temperatures by either adding frozen drops of lysis buffer to a frozen cell powder or flash-freezing with the cell mass. When this is not applicable, i.e. by ribosome profiling of tissues, the lysis buffer is generally added to the sample ice-cold (Gonzalez et al., 2014).

Eukaryotic cells are lysed on ice by repeated micropipetting or homogenization (Guo et al., 2010; Becker et al., 2013; Chew et al., 2013). Pulverized bacteria or monocellular eukaryotes are homogenized in a mill with liquid nitrogen (Oh et al., 2011; Guydosh and Green, 2014; Woolstenhulme et al., 2015). This method is transferrable to any cell type and frozen tissue and should be the preferred lysis approach as it allows treatment of the sample at very low temperatures. During the homogenization, local temperature fluctuations in the sample should be avoided by careful choice of the conditions, i.e. short homogenization pulses and pre-cooling the grinder jar before and after each homogenization cycle (Oh et al., 2011; Guydosh and Green, 2014; Woolstenhulme et al., 2015).

Nucleolytic generation of ribosomal footprints

The clarified lysate is then digested with a nuclease to generate monosomes (Figure 1). RNase I has been exclusively used in eukaryotic ribosome profiling (Ingolia et al., 2012) and micrococcal nuclease (MNase) from *Staphylococcus aureus* in bacteria; RNase I is inactive in bacteria (Datta and Burma, 1972). MNase can also be used in eukaryotic lysates (Reid and Nicchitta, 2012; Dunn et al., 2013), and, in fact, it leads to a reduced amount of ribosomal RNA (rRNA) contamination compared to RNase I treatment (Oh et al., 2011; Miettinen and Bjorklund, 2015). The activity of the MNase is modulated by calcium ions. A disadvantage of MNase is its preferential cleavage at A or T nucleotides (Dingwall et al., 1981) and consequently, the MNase-generated ribosome footprints might be enriched in A or T nucleotides at their 5′ ends. Compared to fragments derived from yeast lysates treated with RNase I, the MNase-generated footprints are more heterogeneous in length (Becker et al., 2013) due to steric effects and less precise 5′ cleavage (Woolstenhulme et al., 2015). In contrast, MNase cleaves precisely at the 3′ end contour of the ribosome, thus the calibration of the reads in bacterial system should be preferably done using the 3′ ends of the reads (Woolstenhulme et al., 2015) (see section 'Analysis of the sequencing data'). RNase I cleavages are precise at both 5′ and 3′ ends, enabling calibration using both termini. Conversely, RNase I-treated samples show a slight bias towards enrichment of short genes (Miettinen and Bjorklund, 2015), although the reason for this remains unclear.

Contamination with rRNA fragments released by the nucleolytic digestion substantially decreases the amount of informative sequencing data. Importantly, the rRNA fragments generated during the nucleolysis of the polysomes are species-specific, but are limited to only few fragments and can be efficiently removed to near completeness by using few complementary oligonucleotides. Thus, in setting up a protocol for ribosome profiling in a new cell line or species, it is recommendable by to perform a pioneer sequencing run to identify the contaminant rRNA species and design specific oligonuclotides for the depletion of rRNA-derived fragements.

Finally, the amount of each nuclease needs a careful determination; enhanced nuclease activity (caused either by large amounts of enzyme, pH variations or long digestion times) leads primarily to an increased contamination of the ribosome footprint libraries with rRNA fragments. By contrast, insufficient amount of MNase causes less stringent cleavage of the mRNA and results in longer

fragments which migrate outside of the range selected for ribosomal fragments during the gel purification procedure. Consequently, it will yield lower depth and coverage of the mRNAs and it will decrease the accuracy in determining ribosome positions along mRNAs (Becker et al., 2013).

Generation of the deep-sequencing library

The preparation of libraries for deep sequencing involves fusion of adapters to the generated small DNA or RNA fragments. This process also contains biases and a detailed knowledge is of crucial importance to avoid erroneous interpretation of the data. A recent review summarizes the critical caveats in each step of library preparation (van Dijk et al., 2014). Here, we only compare various methods for adaptor ligations to the ribosomal footprints, which are unique to the ribosome profiling procedure. In principle, after nucleolytic digestion the ribosome profiling follows the typical steps of library preparation in the micro RNA-Seq methodology (Guo et al., 2010), including sequential adaptor ligation, reverse transcription of the RNA fragments and PCR amplification of the transcribed DNA. The earliest approach uses circularization of the fragments to fuse adaptors at both ends (Ingolia et al., 2009). Prior to this, each fragment is polyadenylated at its 3′ ends with poly(A)-polymerase (Ingolia et al., 2009), which serves as a priming site for the reverse transcription. Polyadenylation was also introduced to produce uniform 3′ ends of all fragments and to reduce the bias in the ligation (Ingolia, 2010), however the sequenced fragments are enriched in adenines at their 3′ termini (Artieri and Fraser, 2014). Furthermore, in the circularization procedure, an additional preference for adenine at the first 5′-position is observed (Lamm et al., 2011; Artieri and Fraser, 2014): it does not depend on the polyA-tails of the fragments and the origin of this bias is unknown.

Later developments in the library preparation of ribosomal footprints use ligation approaches established in the sequencing of miRNAs, in which 3′ and 5′ adaptors are ligated sequentially to the fragments without circularization (Guo et al., 2010). This allowed capture of low-abundance fragments and omitted the sequence bias (i.e. the preference for adenines at 5′ and 3′ positions). Note that direct ligation of a 3′ adapter might be applied also as an alternative to polyA-tailing, preceding the circularization approach. However, some sequences in the libraries generated with sequential adaptor ligation were overrepresented compared to a sequencing in which the adaptors were ligated using the circularization protocol. The overrepresented fragments are a consequence of local secondary structure preferences (Hafner et al., 2011; Zhuang et al., 2012) and their propensity to co-fold with the adaptor sequences (Jackson et al., 2014). Using truncated T4 RNA Ligase 2 instead of the previously used full-length, non-truncated version decreased the amount of those fragments by a half (Jackson et al., 2014). Introducing short (2–4 nt) randomized sequences at the 5′ and 3′ ends of the adaptors also reduced the adaptor ligation bias (Jayaprakash et al., 2011; Sorefan et al., 2012; Zhang et al., 2013).

Analysis of the sequencing results

The ribosomal footprints are very short (25–35 nt dependent on the organism, nucleolytic digestion protocol and manually excised region of the gel) and are usually sequenced by a single-end sequencing approach. The maximum number of total reads coming from a sequencing machine vary between sequencing samples (Mortazavi et al., 2008; Garber et al., 2011): for example, our experience with various organisms (bacteria, mouse cell lines and tissues, plants and human samples) for which we performed ribosome profiling on a Illumina HiSeq2000 (Illumina, San Diego, USA), have generated 40–195 million reads per sequencing lane. The final amount of reads correlates with the quality and quantity of the input material. The first step in the data processing undergoes an initial quality and adaptor trimming (Figure 2). There is no uniform quality cut-off score and most ribosome profiling

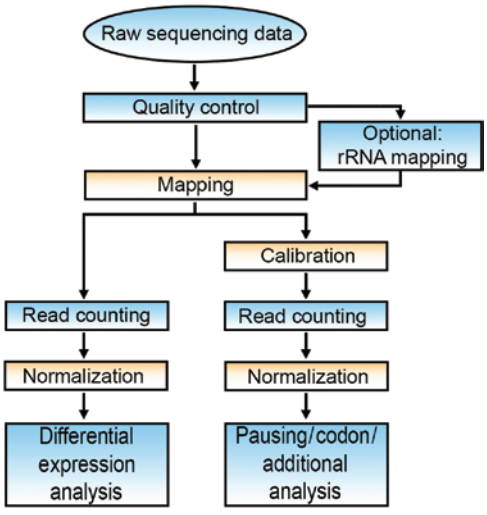

Figure 2: Flow-chart of data analysis in ribosome profiling. Crucial steps are color-coded in orange.

data are processed with a Phred score in the range ~20–30 or with 99.0–99.9% base accuracy (Ingolia et al., 2012; Zhang et al., 2012). In NGS data sets the quality drops towards the 3′ end of the reads (Dohm et al., 2008) which is also mirrored in the ribosome profiling libraries despite the short length of the fragments. Most of the tools used for this initial data processing (https://code.google.com/p/cutadapt/; http://hannonlab.cshl.edu/fastx_toolkit) (Lindgreen, 2012; Bolger et al., 2014) also offer removal of reads with length shorter than expected upon adaptor cutting.

Read mapping

Read mapping is the most crucial procedure. Although principally the ribosomal footprints are in their core an RNA-Seq data set, there is no standardized pipeline with recommended mapping parameters. Mapping can be performed to genomes or transcriptomes, but the short single-end reads generated in the ribosome profiling experiment cannot be used for *de novo* assembly of genomes or transcriptomes (Simpson and Pop, 2015). Mapping to the genome should be preferred as it is unbiased towards known exon and intron annotations and allows for discovery of previously undescribed ORFs (Andreev et al., 2015). Genome mapping usually gives greater coverage than mapping to transcriptomes (the loss of reads on exon junctions is minor) (Oshlack et al., 2010). Furthermore, genomes are better defined than transcriptomes, which are constructed in several different ways (reviewed in Garber et al., 2011). Also, mapping to genomes is less computationally intense and thus faster.

A prerequisite to good results is complete genome annotation, i.e. the availability of the gene coordinates. Genome annotation is a subject of intensive and constant improvement. For example the *E.coli* genome hosted on the NCBI server (Freddolino et al., 2012) is updated daily and the number of genes constantly changes. Although this fast adjustment makes new findings immediately available, it creates a gap with the hand-curated databases, some of which may offer more precise annotation of additional features. For example, *RegulonDB* (Salgado et al., 2013) offers more information on additional features than the NCBI annotations, including genes organized in operons, 5′ and 3′ UTRs. For eukaryotes the development is equally fast with frequently updated versions of genomes and their annotations. Three important webservers host various eukaryotic genomes: NCBI reference sequences, RefSeq (Pruitt et al., 2007), *ensembl* (Cunningham et al., 2015) and UCSC (Kent et al., 2002). The genome annotation choice may significantly influence the downstream quantification of expression and differential analysis (Zhao and Zhang, 2015), although a simple advice on which database to use is not possible and should be driven by the purpose of the analysis. For research aiming at reproducible and robust gene expression estimates, RefSeq might be preferred (Wu et al., 2013). More exploratory questions may rely on more complex annotations, e.g. *ensembl*.

The mapping tools can be classified into two major groups: hash-table based (Li et al., 2008; Homer et al., 2009); or Burrows-Wheeler Transform (BWT) algorithms (Langmead et al., 2009; Li and Durbin, 2009). While BWT-based approaches are faster and less computationally demanding, the hash-table-based algorithms are more flexible in aligning reads with non-perfect matches. Also the efficiency of BWT-based mapping approaches inversely correlates with the number of mismatches [reviewed in (Li and Homer, 2010; Garber et al., 2011)]. Comparison of the tools is not trivial and differs depending on the data set, thus only few objective investigations have been performed so far (Giannoulatou et al., 2014). In the majority of ribosome profiling experiments (Ingolia et al., 2009, 2011; Guo et al., 2010; Gerashchenko et al., 2012; Li et al., 2012; Chew et al., 2013; Guttman et al., 2013; Aspden et al., 2014; Baudin-Baillieu et al., 2014; Bazzini et al., 2014; Subramaniam et al., 2014), *Bowtie* (Langmead et al., 2009) is used as a BWT-based mapping program. *Bowtie* offers two ways of mapping a read to a reference sequence: seed- (parameter n) and mismatch-based approach (parameter v, Table 1). The seed approach aligns first a seed (or core) of a read and then extends the alignment further along the read length. Thereby, the mismatches in the seed count stronger than those in the extensions. Mostly, default *Bowtie* parameters (parameter n for the seed-based approach) are used (Guo et al., 2010; Li et al., 2012; Baudin-Baillieu et al., 2014; Subramaniam et al., 2014). Some studies apply the mismatch approach (Ingolia et al., 2011; Gerashchenko et al., 2012) which scores every base of each read equally. As the default seed length of 28 nt remains unchanged when using the default parameter settings, the seed-based strategy effectively works as a mismatch approach.

A general drawback of *Bowtie* is its inability to map splice junctions. One commonly used tool to align short reads across junctions is *TopHat* (Trapnell et al., 2009; Kim et al., 2013) which can also find junctions *de novo*. First, the *TopHat* pipeline maps to all reads to a reference genome using *Bowtie* and allows reporting more than one alignment of a read (i.e. m=inf k=20 [translated to *Bowtie* parameters]). *TopHat* then assembles the mapped reads using the assembly module in Maq (Li et al., 2008) in

Table 1: *Bowtie* and *TopHat* mapping parameters and their effects.

Bowtie		*TopHat*		Description	Comments
Parameter	**Default**	**Parameter**	**Default**		
n	Yes, 2	*bowtie*-n	No	Seed-based mapping approach	Mutually exclusive with v parameter
l	28	b2-L (only *Bowtie2*)	20	Length of the seed	Only usable with n; default – too long for RPF reads
v	No	N	Yes, 2	Mismatch-based mapping approach	Mutually exclusive with n parameter
m	Infinite	NA, must be filtered after mapping		Maximum number of multiple positions per read	Uniquely mapping; m=1; try to avoid large numbers
best	No	Always on		Reports alignment in best to worst order	Should be used always
strata	No	Different scoring concept		Must be combined with best; reports best positions only	Should be used always
k	1	g	20	Maximum number of reported alignment	Should be chosen carefully
a	No	NA		As k, but reports all alignment	Same as k=m
–	–	*bowtie*1	No	Use *Bowtie* instead of *Bowtie2*	*Bowtie* does not allow changing mapping parameters

contiguous sequences inferring them to be putative exons, then uses seed and extended alignment to match reads to possible splice sites (Trapnell et al., 2009). The pipeline of *TopHat* is more structured, with fewer possibilities for changing the mapping parameters (Table 1), whereas *Bowtie* allows flexible adjustment of the mapping parameters. A new version of *Bowtie*, *Bowtie*2, has been launched (Langmead and Salzberg, 2012) which however differs conceptually from *Bowtie* and can find gapped alignments of reads resulting from insertions or deletions or sequencing errors. Note that *Bowtie*2 is suitable for reads longer than 50 nt.

In general, mapping can be defined as a procedure to find the unique position of each read in the reference genome (Oshlack et al., 2010). The logical consequence of this is to discard all reads with more than one best position. As ribosomal footprints are very short, the proportion of reads mapping at more than one position increases with the size of the genome. In the ribosome profiling datasets a large fraction of the reads map at multiple positions to the genome (in the range of 30% and more), however there is no uniform strategy on how to handle them. Strategies range from considering only reads uniquely mapping to the genome (Guo et al., 2010; Baudin-Baillieu et al., 2014), to allowing more than 200 or unlimited number of positions (Ingolia et al., 2011). One strategy to estimate the true location of reads that map to many (multimappable) locations involves proportional assignment of reads based on the read density of the neighboring positions (Trapnell et al., 2010). We varied the number of the mapping positions when aligning reads from ribosome profiling data (Figure 3A) and observed a clear difference between unique mapping (Figure 3B,C) and allowing multiple positions (Figure 3D,E). The number of the mapped reads increases when multiple mapping positions are allowed (Figure 3A). Such scenarios are relevant mostly to genes with duplications or highly homologous isoforms, nevertheless the fraction of the discarded non-uniquely mappable reads can be in the range of 30%. Stringent criterion leads to sparse and incomplete coverage of each gene (Figure 3B–C), while allowing mapping to multiple positions in the genome improves significantly the coverage of a single gene (Figure 3D–E). The assignment of the reads mapping to multiple positions is also of crucial importance. Multimappable reads can be assigned randomly to one of the possible positions they map (Figure 3D) or to all possible positions (Figure 3E). However, choosing the mapping parameter in such a way that the first hit position is reported (Figure 3D) bears some caveats as the origin of the reads is unclear, i.e. whether they are from the same gene or originate from

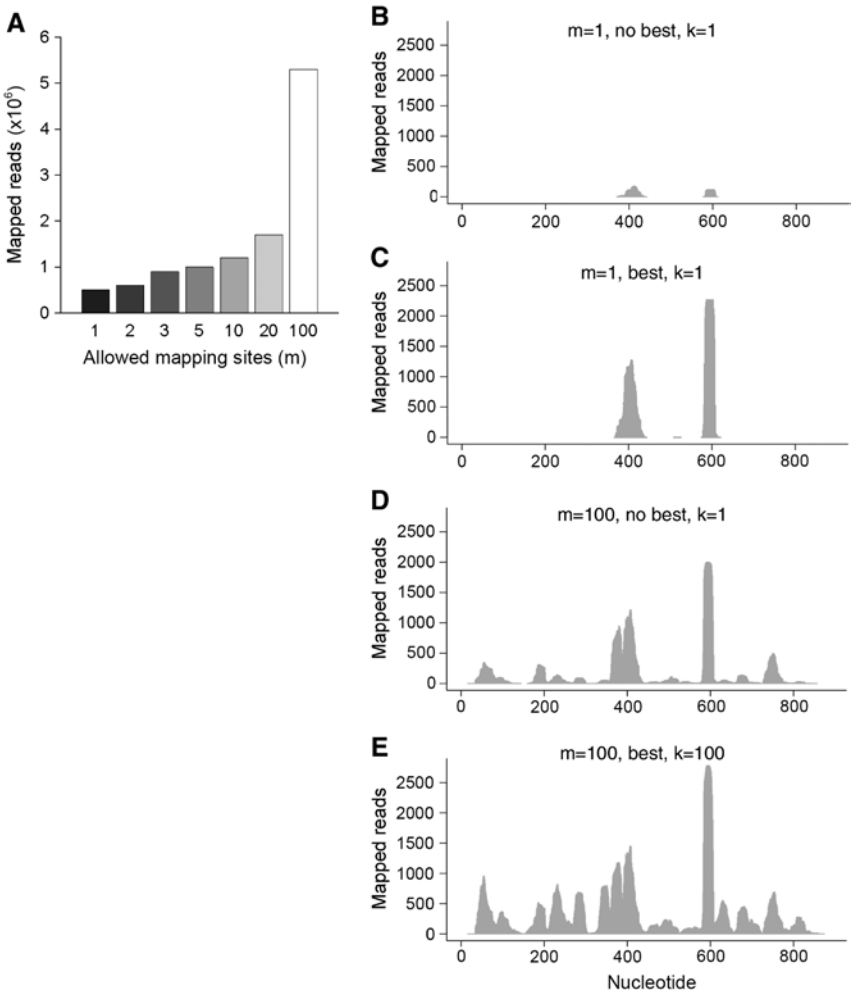

Figure 3: Effects of different mapping strategies on gene coverage.
(A) Total reads mapped to the genome allowing different number of maximal best mappable positions per read (at m=1, a read with one best position (uniquely mappable) is allowed, otherwise discarded; at m=100 a read can be mapped to the 100 best positions and all positions are recorded). Mapping of ribosome profiling data of mouse brain was performed with *Bowtie* using the mouse genome (assembly GRCm38) allowing two mismatches per read (-v2 –strata –best). (B–E) Different mapping strategies result in variations in the read coverage of a human *rplA* gene mapped from ribosome profiling data with *Bowtie* using the human genome (assembly GRCm38). Mapping with (B) a single hit (uniquely mappable) fulfilling the restrictions of maximum two mismatches (-v2 –m1), (C) with a single best hit (uniquely mappable) (-v2 –m1 –strata –best), (D) with default parameter and restrictions to maximum 100 positions with two mismatches, but with only one listed in the output (-v2 –m100), (E) with multimapping restricted to 100 multimappable best positions (i.e. lowest number of mismatches) and best positions listed in the output (-v2 –m100 –k100). Best, the parameters strata best are given to ensure that a multimappable position is counted as such only by the same minimum number of mismatches; no best, default mode with no strata best parameters chosen; m, maximum number of multiple positions per read; k, maximum number of reported alignment. Note settings as in B should be avoided. Parameters as in C show the most reliable data, although the coverage is incomplete. The loss of reads in D compared to E is most likely due to the reporting of only one of the valid alignment positions.

another position in the genome. Thus, it might artificially increase the total number of reads on a gene. In this context, mapping to multiple positions with equal weighting of all positions (Figure 3E) might be a better choice as it does not prefer between positions and maps uniformly to all best mappable positions. For some analysis, to avoid overinterpretation of the data (for example by differential analysis), the most conservative mapping with uniquely mappable reads (Figure 3C) might be the best choice.

The majority of the ribosome profiling datasets mapped with *Bowtie* do not set parameters to evaluate the quality of the alignments for a read (e.g. strata best), that compares, for example, whether a zero-mismatch mapping is better than an alignment with two mismatches. Usually, the first encountered alignment of a read is assigned to

it (Gerashchenko et al., 2012; Li et al., 2012; Subramaniam et al., 2014). Thus, when multimapping is allowed, a read with zero mismatches in a certain position may also be mapped to a different position with two mismatches. Consequently, it creates a bias because the best alignment would not be satisfied, but a read is randomly assigned to one of the two positions independent of the number of the mismatches. The choice of the parameters for the mapping are of crucial importance as they can result in significant variations in the mapping and gene coverage profiles (Figure 3B–E). For reproducibility of the results it is advisable to clearly state the mapping parameter in each publication.

As both nucleases (RNase I for eukaryotic and MNase for bacterial systems) that are used to produce ribosomal footprints, cleave also rRNA, the rRNA reads comprise a large fraction of the sequencing reads, despite their removal in the experimental procedure. rRNA mapping and subtraction of those reads can be done in an extra round before or after mapping to the genome. Thereby, the mapping of the rRNA reads should be strict, i.e. allowing only a single mismatch.

In summary, mapping defines the shape of dataset to be used for further analysis and hence is a crucial step for which the parameters should be chosen carefully. In studies aiming at reproducible and robust gene expression estimates, uniquely mappable reads aligned to a reference genome (i.e. m=1 strata best) should be selected as they bear the lowest bias. However, for some genes (e.g. isoforms, duplicates), using only uniquely mapped reads may result in a partial coverage of a gene (Figure 3B,C); an incomplete coverage cannot be used to extract specific positions on which the ribosomes may pause or enrichment of reads over specific codons. For such analysis a multiple alignment of the multimappable reads (i.e. m=10 a strata best) might be chosen to ensure a maximal gene coverage. This parameter set bears drawbacks in analyzing the coverage of simultaneous expression of genes sharing large sequence identity (i.e. isoforms and duplicates). Such genes should be carefully assessed and might be separately compared only with their uniquely mappable reads or their expression should be confirmed with alternative methods (e.g. qRT-PCR).

Normalization of the read counts

Following mapping, read counts, also called gene counts, are collected and assigned to each gene or non-coding RNAs. Overlapping genes can be an issue here. As the ribosome profiling protocol is strand-specific, overlapping genes on different strands are well resolved. For genes overlapping on the same strand, as commonly observed in *E. coli* in which the coding sequence of the one gene falls into the end of the coding sequence of another, some read counting tools correct for this by randomly distributing the reads to the two overlapping genes (Anders et al., 2015), while other tools do not recognize overlapping features (Quinlan, 2014).

A commonly applied approach for normalization of the read counts is reads per kilobase of exon per million mapped reads, rpkM, (Mortazavi et al., 2008), which accounts for the differences in the sequencing depth (i.e. total number of the mapped reads) between sequencing libraries and for the length variation of each gene (i.e. per kilobase). Note that for short genes this normalization can give quite high rpkM values despite the presence of only few raw counts. Thus the detection limit should be set up using the raw counts (Ingolia et al., 2011). Other normalization approaches frequently applied in the RNA sequencing (RNA-Seq) might be applied too (Anders and Huber, 2010; Robinson et al., 2010; Dillies et al., 2013) but the statistical behavior of ribosome profiling data with those normalization procedures has not yet been tested (Olshen et al., 2013).

Further downstream analysis and post-processing

In the RNA-Seq datasets, several tools are used to identify differentially expressed (DE) genes (Guo et al., 2013), some of which (e.g. *DESeq* tool) have been applied in a few ribosome profiling studies (Baudin-Baillieu et al., 2014; Sidrauski et al., 2015). Still, they require a test that the ribosome profiling read counts follow the underlying distributions required by many tools designed for DE analysis of RNA-Seq, for example *DESeq* (Anders and Huber, 2010), *EdgeR* (Robinson et al., 2010), and *baySeq* (Hardcastle and Kelly, 2010). In all cases, a careful and conservative interpretation of the data is needed because, unlike RNA-Seq (Dillies et al., 2013), no uniform pipeline exists for ribosome profiling data. So far, only one tool has been developed specifically for ribosome profiling data (Olshen et al., 2013). Instead of performing DE analysis, a simple fold-change analysis can be carried out (Dunn et al., 2013) with the assumption that most of the genes are unchanged.

Still, a fascinating issue of ribosome profiling is the ability to record the position of ribosomes with single nucleotide resolution (Ingolia et al., 2009; Woolstenhulme et al., 2015), which enables detecting ribosomal pausing

(i.e. specific positions at which ribosomes pause) or encoding events (e.g. readthrough or frameshifting) (Li et al., 2012; Michel et al., 2012; O'Connor et al., 2013). The alignment of the ribosomal reads to the open-reading frame is called calibration in which the start codon is assigned to the ribosomal P-site (Ingolia et al., 2009) or the stop codon is assigned to the A-site (Woolstenhulme et al., 2015). If the ribosomes are not completely halted during the isolation procedure, it will compromise the calibration and would not allow for codon resolution (Ingolia et al., 2009). While ribosomal footprints of eukaryotic ribosomes can be calibrated using both stop and start codons, i.e. both 5′ and 3′ of the reads, reads from bacterial systems give only codon resolution when calibrated using their 3′ ends, most likely because of the sharp cleavage of the MNase at the 3′ of the reads but not at the 5′ ends (Woolstenhulme et al., 2015). Another approach to gain positional information of the translating ribosomes is center-weighted or center-assigned approach (Li et al., 2012). A defined number of nucleotides are excluded from both 5′ and 3′ sides of a read and the remaining centrally positioned nucleotides are weighted equally. This approach delivers less sharp resolution and defines the position of the ribosomal A- or P-sites with a subcodon resolution. Thus, it has limited applications and cannot be used for determining the reading frame (Woolstenhulme et al., 2015). Both, calibrated and center-weighted ribosomal reads can be used to assess ribosomal enrichment over specific codons (Li et al., 2012; Ishimura et al., 2014) or to determine sequences over which ribosomes transiently pause (Li et al., 2012; Woolstenhulme et al., 2015).

In the library preparation, RNA fragments over a typical length range of 25–35 nt, tightly distributed around a peak of ~28 nt, are selected from the gel upon ribonucleolytic digestion (Ingolia et al., 2009; Guydosh and Green, 2014). It should be noted that reads outside this range may also bear some biological information and, dependent on the specific question, might also be included in the library preparation. Reads shorter than the average length of ~28 nt represent different conformational states of the elongating ribosome (Lareau et al., 2014) or report on ribosomes stalled over 3′ truncated mRNAs (Guydosh and Green, 2014). In turn, longer reads may be informative on frameshifting events (O'Connor et al., 2013). When comparing expression level on a gene basis in the DE analysis, all reads independent of their length might be considered under the assumption that each ribosome read produces one protein. For more specific analysis, including ribosomal stalling at specific positions, the reads should be separated by their length and each length group should be treated separately.

Computational demand and infrastructure

Raw data from one sequencing lane of Illuminas HiSeq machine (Illumina, San Diego, USA) can reach a size of more than 20 GB (uncompressed). Preprocessing and mapping of these raw files easily exceeds another 20 GB; discarding the intermediate preprocessing file and keeping only compressed raw files requires hard disk space for one lane of about 20 GB. The demand of RAM varies dependent on the type of analysis and programming languages. For example, using a simple Perl hash index build on each of the ~4 million nucleotides of the relatively small *E. coli* genome requires more than 4 GB of RAM. Mapping with BWT-based algorithms demands relatively low memory (Langmead et al., 2009; Li and Durbin, 2009). For example, the human genome can be mapped with <8 GB of RAM (Langmead et al., 2009; Li and Durbin, 2009). The mapping programs offer an option to use more than one CPU in parallel to increase speed (Langmead et al., 2009). Many of pre- and post-processing steps are not implemented as full programs but as a collection of scripts or even in-house scripts (Anders et al., 2015).

Conclusions

Ribosome profiling is a powerful technology to study translation *in vivo* on a cell-wide scale. While introducing this approach we are beginning to appreciate the variety of mechanisms that control translation and gene expression. However, non-standardized sample preparation and ambiguous processing of the data has produced some inconsistencies and has challenged direct comparisons between different studies. Experimentally, ribosome profiling is a multistep procedure that is in constant development and improvement of the single experimental steps. The task would be to understand the intrinsic bias of each step in order to carefully design the experimental protocol and interpret the data.

The analysis of data is complex, in part because of the short read lengths. Particularly crucial is the mapping procedure and normalization that defines the data set for further downstream analysis. The goal in the data analysis is to develop a uniform protocol, at least for mapping and normalization, as the broadness of the downstream analysis does not allow full standardization of this part of the pipeline. With the development of more standardized ribosome profiling technology and optimized sample preparation, we will move to a higher reproducibility of the data and a more accurate quantitative understanding of the mechanisms of translational control.

Acknowledgments: This work was supported by the Deutsche Forschungsgemeinschaft (FOR 1805) and European Union (grant NICHE ITN) to ZI.

References

Anders, S. and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq – a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Andreev, D.E., O'Connor, P.B., Fahey, C., Kenny, E.M., Terenin, I.M., Dmitriev, S.E., Cormican, P., Morris, D.W., Shatsky, I.N., and Baranov, P.V. (2015). Translation of 5′ leaders is pervasive in genes resistant to eIF2 repression. eLife *4*, e03971.

Artieri, C.G. and Fraser, H.B. (2014). Accounting for biases in riboprofiling data indicates a major role for proline in stalling translation. Genome Res. *24*, 2011–2021.

Aspden, J.L., Eyre-Walker, Y.C., Phillips, R.J., Amin, U., Mumtaz, M.A., Brocard, M., and Couso, J.P. (2014). Extensive translation of small Open Reading Frames revealed by Poly-Ribo-Seq. eLife *3*, e03528.

Baudin-Baillieu, A., Legendre, R., Kuchly, C., Hatin, I., Demais, S., Mestdagh, C., Gautheret, D., and Namy, O. (2014). Genome-wide translational changes induced by the prion [PSI+]. Cell Rep. *8*, 439–448.

Bazzini, A.A., Johnstone, T.G., Christiano, R., Mackowiak, S.D., Obermayer, B., Fleming, E.S., Vejnar, C.E., Lee, M.T., Rajewsky, N., Walther, T.C., et al. (2014). Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. EMBO J. *33*, 981–993.

Becker, A.H., Oh, E., Weissman, J.S., Kramer, G., and 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. Protocols *8*, 2212–2239.

Blanchard, S.C., Kim, H.D., Gonzalez, R.L., Jr., Puglisi, J.D., and Chu, S. (2004). tRNA dynamics on the ribosome during translation. Proc. Natl. Acad. Sci. USA *101*, 12893–12898.

Blobel, G., and Sabatini, D. (1971). Dissociation of mammalian polyribosomes into subunits by puromycin. Proc. Natl. Acad. Sci. USA *68*, 390–394.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Chen, C., Zhang, H., Broitman, S.L., Reiche, M., Farrell, I., Cooperman, B.S., and Goldman, Y.E. (2013). Dynamics of translation by single ribosomes through mRNA secondary structures. Nat. Struct. Mol. Biol. *20*, 582–588.

Chew, G.L., Pauli, A., Rinn, J.L., Regev, A., Schier, A.F., and Valen, E. (2013). Ribosome profiling reveals resemblance between long non-coding RNAs and 5′ leaders of coding RNAs. Development *140*, 2828–2834.

Cunningham, F., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., et al. (2015). Ensembl 2015. Nucleic Acids Res. *43*, D662–669.

Datta, A.K. and Burma, D.P. (1972). Association of ribonuclease I with ribosomes and their subunits. J. Biol. Chem. *247*, 6795–6801.

Dillies, M.A., Rau, A., Aubert, J., Hennequet-Antier, C., Jeanmougin, M., Servant, N., Keime, C., Marot, G., Castel, D., Estelle, J., et al. (2013). A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. Briefings Bioinformatics *14*, 671–683.

Dingwall, C., Lomonossoff, G.P., and Laskey, R.A. (1981). High sequence specificity of micrococcal nuclease. Nucleic Acid Res. *9*, 2659–2673.

Dohm, J.C., Lottaz, C., Borodina, T., and Himmelbauer, H. (2008). Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic Acids Res. *36*, e105.

Dresden, M. and Hoagland, M.B. (1965). Polyribosomes from *Escherichia coli*: Enzymatic method for isolation. Science *149*, 647–649.

Dunn, J.G., Foo, C.K., Belletier, N.G., Gavis, E.R., and Weissman, J.S. (2013). Ribosome profiling reveals pervasive and regulated stop codon readthrough in Drosophila melanogaster. eLife *2*, e01179.

Freddolino, P.L., Amini, S., and Tavazoie, S. (2012). Newly identified genetic variations in common *Escherichia coli* MG1655 stock cultures. J. Bact. *194*, 303–306.

Fritsch, C., Herrmann, A., Nothnagel, M., Szafranski, K., Huse, K., Schumann, F., Schreiber, S., Platzer, M., Krawczak, M., Hampe, J., et al. (2012). Genome-wide search for novel human uORFs and N-terminal protein extensions using ribosomal footprinting. Genome Res. *22*, 2208–2218.

Garber, M., Grabherr, M.G., Guttman, M., and Trapnell, C. (2011). Computational methods for transcriptome annotation and quantification using RNA-seq. Nat. Methods *8*, 469–477.

Gerashchenko, M.V. and Gladyshev, V.N. (2014). Translation inhibitors cause abnormalities in ribosome profiling experiments. Nucleic Acids Res. *42*, e134.

Gerashchenko, M.V., Lobanov, A.V., and Gladyshev, V.N. (2012). Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress. Proc. Natl. Acad. Sci. USA *109*, 17394–17399.

Giannoulatou, E., Park, S.H., Humphreys, D.T., and Ho, J.W. (2014). Verification and validation of bioinformatics software without a gold standard: a case study of BWA and Bowtie. BMC Bioinform. 15 (Suppl. 16), S15.

Gonzalez, C., Sims, J.S., Hornstein, N., Mela, A., Garcia, F., Lei, L., Gass, D.A., Amendolara, B., Bruce, J.N., Canoll, P., et al. (2014). Ribosome profiling reveals a cell-type-specific translational landscape in brain tumors. J. Neurosci. *34*, 10924–10936.

Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature *466*, 835–840.

Guo, Y., Li, C.I., Ye, F., and Shyr, Y. (2013). Evaluation of read count based RNAseq analysis methods. BMC Genom. 14 (Suppl. 8), S2.

Guttman, M., Russell, P., Ingolia, N.T., Weissman, J.S., and Lander, E.S. (2013). Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. Cell *154*, 240–251.

Guydosh, N.R. and Green, R. (2014). Dom34 rescues ribosomes in 3′ untranslated regions. Cell *156*, 950–962.

Hafner, M., Renwick, N., Brown, M., Mihailovic, A., Holoch, D., Lin, C., Pena, J.T., Nusbaum, J.D., Morozov, P., Ludwig, J., et al. (2011). RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. RNA *17*, 1697–1712.

Hardcastle, T.J. and Kelly, K.A. (2010). baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. BMC Bioinformatics *11*, 422.

Homer, N., Merriman, B., and Nelson, S.F. (2009). BFAST: an alignment tool for large scale genome resequencing. PloS One *4*, e7767.

Ingolia, N.T. (2010). Genome-wide translational profiling by ribosome footprinting. Methods Enzymol. *470*, 119–142.

Ingolia, N.T. (2014). Ribosome profiling: new views of translation, from single codons to genome scale. Nat. Rev. Genet. *15*, 205–213.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218–223.

Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell *147*, 789–802.

Ingolia, N.T., Brar, G.A., Rouskin, S., McGeachy, A.M., and Weissman, J.S. (2012). The ribosome profiling strategy for monitoring translation *in vivo* by deep sequencing of ribosome-protected mRNA fragments. Nat. Protocols *7*, 1534–1550.

Ishimura, R., Nagy, G., Dotu, I., Zhou, H., Yang, X.L., Schimmel, P., Senju, S., Nishimura, Y., Chuang, J.H., and Ackerman, S.L. (2014). RNA function. Ribosome stalling induced by mutation of a CNS-specific tRNA causes neurodegeneration. Science *345*, 455–459.

Jackson, R. and Standart, N. (2015). The awesome power of ribosome profiling. RNA *21*, 652–654.

Jackson, T.J., Spriggs, R.V., Burgoyne, N.J., Jones, C., and Willis, A.E. (2014). Evaluating bias-reducing protocols for RNA sequencing library preparation. BMC Genomics *15*, 569.

Jan, C.H., Williams, C.C., and Weissman, J.S. (2014). Principles of ER cotranslational translocation revealed by proximity-specific ribosome profiling. Science *346*, 1257521.

Jayaprakash, A.D., Jabado, O., Brown, B.D., and Sachidanandam, R. (2011). Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. Nucleic Acids Res. *39*, e141.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res. *12*, 996–1006.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.

Kramer, G., Boehringer, D., Ban, N., and Bukau, B. (2009). The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. Nat. Struct. Mol. Biol. *16*, 589–597.

Kuersten, S., Radek, A., Vogel, C., and Penalva, L.O. (2013). Translation regulation gets its 'omics' moment. Wiley Interdiscipl. Rev. RNA *4*, 617–630.

Lamm, A.T., Stadler, M.R., Zhang, H., Gent, J.I., and Fire, A.Z. (2011). Multimodal RNA-seq using single-strand, double-strand, and CircLigase-based capture yields a refined and extended description of the C. elegans transcriptome. Genome Res. *21*, 265–275.

Langmead, B. and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

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

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

Lee, S., Liu, B., Lee, S., Huang, S.X., Shen, B., and Qian, S.B. (2012). Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. Proc. Natl. Acad. Sci. USA *109*, E2424–2432.

Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, H. and Homer, N. (2010). A survey of sequence alignment algorithms for next-generation sequencing. Brief. Bioinform. *11*, 473–483.

Li, H., Ruan, J., and Durbin, R. (2008). Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. *18*, 1851–1858.

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

Lindgreen, S. (2012). AdapterRemoval: easy cleaning of next-generation sequencing reads. BMC Res. Notes *5*, 337.

Liu, B., Han, Y., and Qian, S.B. (2013). Cotranslational response to proteotoxic stress by elongation pausing of ribosomes. Mol. Cell *49*, 453–463.

Michel, A.M. and Baranov, P.V. (2013). Ribosome profiling: a Hi-Def monitor for protein synthesis at the genome-wide scale. Wiley Interdiscipl. Rev. RNA *4*, 473–490.

Michel, A.M., Choudhury, K.R., Firth, A.E., Ingolia, N.T., Atkins, J.F., and Baranov, P.V. (2012). Observation of dually decoded regions of the human genome using ribosome profiling data. Genome Res. *22*, 2219–2229.

Miettinen, T.P. and Bjorklund, M. (2015). Modified ribosome profiling reveals high abundance of ribosome protected mRNA fragments derived from 3′ untranslated regions. Nucl. Acids Res. *43*, 1019–1034.

Morris, D.R. (2009). Ribosomal footprints on a transcriptome landscape. Genome Biol. *10*, 215.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods *5*, 621–628.

Nedialkova, D.D. and Leidel, S.A. (2015). Optimization of codon translation rates via tRNA modifications maintains proteome integrity. Cell *161*, 1606–1618.

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

Oh, E., Becker, A.H., Sandikci, A., Huber, D., Chaba, R., Gloge, F., Nichols, R.J., Typas, A., Gross, C.A., Kramer, G., et al. (2011). Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor *in vivo*. Cell *147*, 1295–1308.

Olshen, A.B., Hsieh, A.C., Stumpf, C.R., Olshen, R.A., Ruggero, D., and Taylor, B.S. (2013). Assessing gene-level translational control from ribosome profiling. Bioinformatics *29*, 2995–3002.

Orelle, C., Carlson, S., Kaushal, B., Almutairi, M.M., Liu, H., Ochabowicz, A., Quan, S., Pham, V.C., Squires, C.L., Murphy, B.T., et al. (2013). Tools for characterizing bacterial protein synthesis inhibitors. Antimicrob. Agents Chemother. *57*, 5994–6004.

Oshlack, A., Robinson, M.D., and Young, M.D. (2010). From RNA-seq reads to differential expression results. Genome Biol. *11*, 220.

Pechmann, S., Chartron, J.W., and Frydman, J. (2014). Local slowdown of translation by nonoptimal codons promotes nascent-chain recognition by SRP *in vivo*. Nat. Struct. Mol. Biol. *21*, 1100–1105.

Pelechano, V., Wei, W., and Steinmetz, L.M. (2015). Widespread Co-translational RNA Decay Reveals Ribosome Dynamics. Cell *161*, 1400–1412.

Pestova, T.V, and Hellen, C.U. (2003). Translation elongation after assembly of ribosomes on the Cricket paralysis virus internal ribosomal entry site without initiation factors or initiator tRNA. Genes Dev. *17*, 181–186.

Plotkin, J.B. and Kudla, G. (2011). Synonymous but not the same: the causes and consequences of codon bias. Nat. Rev. Genet. *12*, 32–42.

Pruitt, K.D., Tatusova, T., and Maglott, D.R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. *35*, D61–65.

Quinlan, A.R. (2014). BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr. Prot. Bioinformatics *47*, 111211–111234.

Reid, D.W. and Nicchitta, C.V. (2012). Primary role for endoplasmic reticulum-bound ribosomes in cellular translation identified by ribosome profiling. J. Biol. Chem. *287*, 5518–5527.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Ron, E.Z., Kohler, R.E., and Davis, B.D. (1968). Magnesium ion dependence of free and polysomal ribosomes from *Escherichia coli*. J. Mol. Biol. *36*, 83–89.

Salgado, H., Peralta-Gil, M., Gama-Castro, S., Santos-Zavaleta, A., Muniz-Rascado, L., Garcia-Sotelo, J.S., Weiss, V., Solano-Lira, H., Martinez-Flores, I., Medina-Rivera, A., et al. (2013). RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. Nucleic Acids Res. *41*, D203–213.

Schneider-Poetsch, T., Ju, J., Eyler, D.E., Dang, Y., Bhat, S., Merrick, W.C., Green, R., Shen, B., and Liu, J.O. (2010). Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin. Nat. Chem. Biol. *6*, 209–217.

Shalgi, R., Hurt, J.A., Krykbaeva, I., Taipale, M., Lindquist, S., and Burge, C.B. (2013). Widespread regulation of translation by elongation pausing in heat shock. Mol. Cell *49*, 439–452.

Sidrauski, C., McGeachy, A.M., Ingolia, N.T., and Walter, P. (2015). The small molecule ISRIB reverses the effects of eIF2a phosphorylation on translation and stress granule assembly. eLife 4.

Simpson, J.T. and Pop, M. (2015). The theory and practice of genome sequence assembly. Ann Rev Genomics Human Genet. *16*, 153–172.

Sorefan, K., Pais, H., Hall, A.E., Kozomara, A., Griffiths-Jones, S., Moulton, V., and Dalmay, T. (2012). Reducing ligation bias of small RNAs in libraries for next generation sequencing. Silence *3*, 4.

Steitz, J.A. (1969). Polypeptide chain initiation: nucleotide sequences of the three ribosomal binding sites in bacteriophage R17 RNA. Nature *224*, 957–964.

Subramaniam, A.R., Zid, B.M., and O'Shea, E.K. (2014). An integrated approach reveals regulatory controls on bacterial translation elongation. Cell *159*, 1200–1211.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511–515.

Tyagi, K. and Pedrioli, P.G. (2015). Protein degradation and dynamic tRNA thiolation fine-tune translation at elevated temperatures. Nucleic Acids Res. *43*, 4701–4712.

van Dijk, E.L., Jaszczyszyn, Y., and Thermes, C. (2014). Library preparation methods for next-generation sequencing: tone down the bias. Exp. Cell Res. *322*, 12–20.

Wen, J.D., Lancaster, L., Hodges, C., Zeri, A.C., Yoshimura, S.H., Noller, H.F., Bustamante, C., and Tinoco, I. (2008). Following translation by single ribosomes one codon at a time. Nature *452*, 598–603.

Wettstein, F.O., Staehelin, T., and Noll, H. (1963). Ribosomal aggregate engaged in protein synthesis: characterization of the ergosome. Nature *197*, 430–435.

Woolstenhulme, C.J., Guydosh, N.R., Green, R., and Buskirk, A.R. (2015). High-precision analysis of translational pausing by ribosome profiling in bacteria lacking EFP. Cell Rep. *11*, 13–21.

Wu, P.Y., Phan, J.H., and Wang, M.D. (2013). Assessing the impact of human genome annotation choice on RNA-seq expression estimates. BMC Bioinformatics 14 (Suppl. 11), S8.

Zhang, G. and Ignatova, Z. (2011). Folding at the birth of the nascent chain: coordinating translation with co-translational folding. Curr. Opin. Struct. Biol. *21*, 25–31.

Zhang, G., Fedyunin, I., Kirchner, S., Xiao, C., Valleriani, A., and Ignatova, Z. (2012). FANSe: an accurate algorithm for quantitative mapping of large scale sequencing reads. Nucleic Acids Res. *40*, e83.

Zhang, Z., Lee, J.E., Riemondy, K., Anderson, E.M., and Yi, R. (2013). High-efficiency RNA cloning enables accurate quantification of miRNA expression by deep sequencing. Genome Biol. *14*, R109.

Zhao, S. and Zhang, B. (2015). A comprehensive evaluation of ensembl, RefSeq, and UCSC annotations in the context of RNA-seq read mapping and gene quantification. BMC Genomics *16*, 97.

Zhuang, F., Fuchs, R.T., and Robb, G.B. (2012). Small RNA expression profiling by high-throughput sequencing: implications of enzymatic manipulation. J. Nucleic Acids Res. *40*, 360358.