Ribosome profiling provides a snapshot of ribosome positions and density across the transcriptome at a sub-codon resolution. By sequencing the entire set of ribosome-protected mRNA fragments, this powerful approach has been successfully used to measure ribosome dynamics and reveal the hidden coding potential of transcriptome. Since its conceptual inception, ribosome profiling has evolved into a versatile method with many innovative variations. It has been applied to study translation in diverse cell types, adapted to capture specific subsets of ribosomes and further improved to address translational regulation in multicellular organisms. The continuous development of ribosome profiling technologies over the coming decade promises a broad view of translational regulation of gene expression.

Introduction

Regulation of gene expression is crucial to implement genomic information and to determine properties of cells and organisms. The field of gene regulation has made great progress over the past decade, in large part stemming from technological development that enables increasingly high-throughput and cost-effective sequencing methods (Metzker, 2010). In addition to whole genome sequencing, a dizzying array of SEQs has been devised to quantify nucleic acid abundances, profile nucleotide modifications and map positions of trans-acting factors. Judging from the rate at which new ‘seq’ words are coined, high-throughput sequencing has taken the wheel in the post-genomic era. Complementing biochemical assays designed for individual molecules, the increasingly accumulating SEQ data sets promise holistic views of gene regulation at transcriptional and translational levels. For the latter, a particular exciting advance is the method of ribosome profiling, which provides a genome-wide snapshot of ribosome positions along mRNA. In this article, we begin with a brief description of general principles of ribosome profiling. We then discuss variations of ribosome profiling approach with a focus on the eukaryotic system. Notably, ribosome profiling has already led to many discoveries about translational control of gene expression. Given the broad scope of translational control mechanisms, this article does not intend to cover comprehensive aspects of translation processes. Instead, we primarily focus on technological advances of ribosome profiling and its broad applications.

Principles of Ribosome Profiling

The concept of ribosome profiling is based on the observation that each ribosome protects a footprint of about 30 nucleotides (nt) when bound to mRNA (Steitz, 1969). This initial observation was confirmed 20 years later by an in vitro ribosome footprinting assay (Wolin and Walter, 1988). It took another 20 years before the birth of ribosome profiling largely due to remarkable advances of sequencing technology (Ingolia et al., 2009). Taking advantage of the power of parallel sequencing, ribosome profiling resembles a global nuclease protection assay at the genome-wide scale. The sequence information derived from the entire set of ribosome-protected fragments (RPFs) permits identification of ribosome positions on mRNAs as well as the ribosome density at individual positions. Therefore, ribosome profiling is often referred to as genome-wide ribosome footprinting or Ribo-seq.

The method of ribosome profiling should not be confused with that of polysome profiling, a commonly used approach to separate ribosome subunits, monosome and polysome by sucrose sedimentation. Before ribosome profiling, polysome profiling was widely used to evaluate the overall translational status under different growth conditions. In coupling with microarray analysis, the changed ratio of mRNA abundance in different fractions reflects translational regulation (Arava et al., 2003). Although informative, the acquired data set suffers from limited resolution and does not provide the position of translating ribosomes. Sequencing-based ribosome profiling, in contrast, offers a quantitative snapshot of translation for individual mRNAs by revealing precise ribosome positions and the corresponding density (Ingolia et al., 2009).

As illustrated in Figure 1, the general procedure of ribosome profiling starts with immobilisation of ribosomes on mRNAs via
General procedures of ribosome profiling. Translating ribosomes are immobilised on mRNAs before or during cell lysis. Polysome is then converted into monosome by RNase digestion followed by size selection of ribosome-protected mRNA fragments. cDNA library is constructed using the purified mRNA fragments before sequencing.

**Figure 1** General procedures of ribosome profiling. Translating ribosomes are immobilised on mRNAs before or during cell lysis. Polysome is then converted into monosome by RNase digestion followed by size selection of ribosome-protected mRNA fragments. cDNA library is constructed using the purified mRNA fragments before sequencing.

elengation inhibitors. It is noteworthy that different inhibitors generate RPFs with slightly different sizes, presumably due to altered ribosome conformations in the presence of distinct compounds (Ingolia et al., 2011). Despite the advantage of stabilised ribosomes in revealing overall ribosome distribution on a given message, pre-treatment of elongation inhibitors may cause elevated ribosome density at the start codon. However, no drug treatment often leads to excess density of ribosomes at the stop codons. The true *in vivo* ribosome distribution may require rapid capture of translating ribosomes with minimal perturbation.

After cell lysis, nuclease digestion removes regions of mRNAs not protected by the ribosome, converting polysome into monosome carrying the RPF. The nuclease choice and the digestion condition need to be carefully considered to avoid incomplete digestion or overdigestion. In addition, samples from different species may have varied nuclease sensitivity. The commonly used RNase I works well for yeast and mammalian cells (Guo et al., 2010). However, it does not work for prokaryotic cells (Oh et al., 2011). In contrast, *Drosophila* cells are highly sensitive to RNase I (Dunn et al., 2013). As a result, micrococcal nuclease (MNase) has been chosen for both bacteria and *Drosophila* cells to achieve desirable results. Notably, MNase has a strong 3′ A/T bias that likely affects the mapping accuracy of RPFs. Other types of nucleases such as RNase A and T1 have also been successfully applied to mammalian cells as alternative nuclease choices. Nuclease digestion inevitably occurs on RNA segments of the ribosome, largely contributing to the components of non-RPFs. Importantly, nuclease overdigestion could disrupt ribosome integrity and alter the RPF size. Therefore, stringent optimisation steps are necessary during nuclease digestion to ensure high-quality RPF enrichment.

Following nuclease digestion, the RPFs are resolved on a gel and collected via a size selection. The typical RPF size is around 28 nt but varies depending on sample preparations. The RPF size corresponds to the relative ribosome conformation and different elongation inhibitors are known to stabilise distinct ribosome conformations and bias the RPF size distribution (Ingolia et al., 2011). For instance, cycloheximide leads to a more compact ribosome conformation than emetine. Interestingly, a recent study reported two radically distinct populations of RPFs corresponding to different stages of translating ribosomes (Lareau et al., 2014). The long RPF (28–30 nt) seems to be derived from the ribosome with non-rotated conformation that readily stabilised by cycloheximide. In contrast, the short RPF (20–22 nt) predominates in the presence of anisomycin that likely traps the ribosome in a rotated conformation. In addition, the RPF size also differs among different species. For example, ribosomes from *Arabidopsis* protect RPF of longer size (32 nt), presumably due to distinct conformation of plant ribosomes (Juntawong et al., 2014). The significance of the proper size selection is further substantiated by the finding that ribosomes protect multiple RPF sizes. For example, an RPF of 16 nt may represent a truncated mRNA, whereas disome-protected fragments could reach up to 65 nt (Guydosh and Green, 2014). Fortunately, the length heterogeneity mainly occurs at the 3′ end of RPFs, indicating that the 5′ end nucleotide is a robust marker for inferring ribosome positions.

The proper size selection of RPFs is followed by cDNA library construction with the entire procedure similar to RNA-seq. Readers are referred to recent papers describing detailed methods and protocols of ribosome profiling as well as RNA-seq (Ingolia et al., 2012). Commercial kits are also available for standardised library construction. Although computational analysis of Ribo-seq data is not the focus of this article, ribosome profiling is often conducted in parallel with RNA-seq because the translation efficiency of each message is estimated as the ratio of RPF density to mRNA abundance.

### Variations of Ribosome Profiling

The seminal experiment of ribosome profiling was first developed in budding yeast *Saccharomyces cerevisiae* (Ingolia et al., 2009). Since then, the approach has been rapidly applied to a broad range of species and modified for different purposes (Figure 2). We highlight some variations of ribosome profiling below.

#### Ribosome profiling across diverse species

The translation machinery differs significantly between prokaryotes and eukaryotes. It is thus not surprising that the ribosome profiling methods used for different species are not identical. For instance, ribosome profiling of bacteria rely on streptomycin to freeze the ribosome and require MNase to digest mRNA (Li
Ribosome Profiling: Principles and Variations

Species-tailored ribosome profiling  
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Selective ribosome profiling  
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Initiating ribosome profiling  
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tissue-specific ribosome profiling  
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2 Variations of ribosome profiling. Modified ribosome profiling protocol has been developed for different species ranging from bacteria to yeast to mammalian cells. Selective ribosome profiling is designed to capture ribosomes synthesising nascent chains interacting with chaperones or targeting different locations such as ER. Initiating ribosome profiling captures ribosomes stalled at the start codon. Tissue-specific ribosome profiling permits cell-type-specific ribosome profiling in multicellular organisms.

et al., 2012). Nevertheless, the basic principles remain the same. Although the size of RPF varies, the ribosome footprints obtained from different species bear an equally strong 3-nt periodicity. This salient feature allows for measurement of average decoding rates of the 61 sense codons in vivo (Gardin et al., 2014). However, owing to the presence of many confounding factors influencing ribosome dynamics, full interpretation of the decoding process could be harder than was originally hoped. Another common feature revealed by ribosome profiling is that ribosomes tend to pause during initiation and termination (Han et al., 2014). Knowing these characteristics of ribosome behaviour will facilitate the detection of unique translational response upon stress conditions. For instance, oxidative stress, heat shock and proteotoxic stress all lead to elevated ribosome density in the beginning of the coding region (~50 codons) (Gerashchenko et al., 2012; Liu et al., 2013; Shalgi et al., 2013). The exact molecular mechanism underlying the early ribosome pausing remains to be elucidated.

The relatively small genome of bacteria permits quantifying absolute protein synthesis rates by coupling ribosome profiling data with proteomic results (Li et al., 2014). In addition, careful measurement of ribosome occupancy upon starvation for single amino acids permits remodelling of translation elongation in Escherichia coli (Subramaniam et al., 2014). Compared to prokaryotic cells, eukaryotes have received more attention partly because of their relatively bigger genome and the separate process of translation. Applying ribosome profiling to sporulating yeast cells revealed extensive translational control in meiosis (Brar et al., 2012). With ribosome profiling, mammalian cells offer multiple aspects of translational control, such as nutrient signalling (Hsieh et al., 2012; Thoreen et al., 2012), microRNA regulation (Guo et al., 2010) and stress response (Liu et al., 2013; Shalgi et al., 2013). Translational control plays a crucial role in development, at least in certain stages when transcription is silent. For example, during the Drosophila oocyte-to-embryo transition, ribosome profiling reveals widespread changes in the post-transcriptional landscape (Kronja et al., 2014). Extensive translational regulation, including putative long non-coding RNAs (lncRNAs), has also been uncovered by ribosome profiling during development of zebrafish (Chew et al., 2013). The application of ribosome profiling in seedlings of Arabidopsis further substantiate the role of translational control in stress response of plant cells to hypoxia (Juntawong et al., 2014).

Protozoan parasites rely on the host translation machinery for protein synthesis. The method of ribosome profiling designed for the host cell should be feasible to study translational regulation of parasites. The translational control has long been appreciated in parasites such as Leishmania major, because their protein-coding genes are organised into long polycistronic transcription units (Kolev et al., 2010). Despite the lack of transcriptional control, many parasites are capable of tightly regulating gene expression throughout their complex life cycle. Comparative ribosome profiling of Trypanosoma brucei not only demonstrated a translational landscape between the bloodstream and procyclic parasite forms but also revealed translation of previously un-annotated coding regions (Jensen et al., 2014; Vasquez et al., 2014). Most recently, the gene expression of the malaria-causing parasite, Plasmodium falciparum, was investigated using ribosome
profiling during five intra-erythrocytic developmental cycles (Caro et al., 2014). Notably, the expression of only a few genes was controlled more stringently via translation rather than by transcription.

Viruses are obligate intracellular parasites and capable of subverting the host translation machinery via diverse strategies (Walsh and Mohr, 2011). Ribosome profiling of a human cell line infected with human cytomegalovirus (HCMV) identified several hundred new open reading frames (ORFs) (Stern-Ginossar et al., 2012). Given the high medical relevance of virus infection, it is anticipated that more studies will be undertaken to uncover the hidden coding potential of different viral genomes. Ribosome profiling of virus-infected cells could also help unravel the complex network of host response and dissect the diverse strategies that viruses used to escape innate immunity.

Selective ribosome profiling

Ribosome profiling typically uses bulky samples containing all the ribosomes to achieve a global pattern of translation across the transcriptome. It remains a formidable task to isolate ribosomes translating subsets of mRNAs or in different subcellular locations. One strategy is to target the newly synthesised polypeptides attached to the ribosome. Selective ribosome profiling was first developed in *E. coli* by cross-linking the chaperone trigger factor (TF) with the newly synthesised polypeptide (Oh et al., 2011). The ribosome complex associated with TF was then enriched by immunoprecipitation followed by RPF purification and sequencing. Given that the ribosome position corresponds to the length of the nascent chain, both the TF substrates and the point of action can be inferred from the TF-associated RPF sequences. More specific ribosome enrichment can be achieved by antibodies recognising nascent chains. Using conformation-specific antibodies, the co-translational folding status of the newly synthesised polypeptides was successfully probed by selective ribosome profiling (Han et al., 2012).

The eukaryotic cell is made up of many compartments and organelles. Proteins synthesised by cytosolic ribosomes are directed to particular locations in the cell. For instance, nearly one-quarter of the proteome is imported into the endoplasmic reticulum (ER). To investigate the translation of ER-associated ribosomes, Jan et al. (2014) developed an elegant approach called proximity-specific ribosome profiling. This is accomplished by co-expression of a spatially restricted biotin ligase (BirA) fusion protein together with ribosomes containing an AviTag. The *in vivo* biotinylation enables the recovery of ribosomes from defined locations such as ER. Remarkably, ER-specific ribosome profiling revealed distinct routes of co-translational translocation (Jan et al., 2014). The similar approach was also successfully applied to enrich ribosomes associated with mitochondria outer surface (Williams et al., 2014).

Mitochondria are unique among intracellular organelles in that they contain their own genome and the translation machinery (Wallace, 2012). The mitochondrial ribosomes are structurally more similar to bacterial ribosomes than the cytosolic ribosomes. Notably, The RNA components of the translation machinery (mt-rRNAs and mt-tRNAs) are supplied by mitochondria, whereas most protein components, including ribosomal proteins, translational factors and aminoacyl-tRNA synthetases, are encoded in the nucleus, synthesised in the cytoplasm and transported into the mitochondria. The human mitochondrial genome contains 37 genes of which only 13 are protein coding (Christian and Spremulli, 2012). Despite its small size, our understanding of its gene regulation and protein expression is very crude. Using an optimised protocol, Rooijers et al. (2013) enriched mitochondrial ribosomes and found that the mt-RPFs exhibit two populations. The short (24–29 nt) and the long (31–36 nt) RPFs may represent two conformations of mitochondrial ribosomes during translation. Unexpectedly, these footprints showed poor 3-nt periodicity, likely because of the relatively relaxed structure of mitochondrial ribosomes.

Profiling of initiating ribosomes

Translation can be divided mechanistically into initiation, elongation and termination. Regular ribosome profiling captures all the ribosomes engaged on mRNAs. It has long been a challenge to separate ribosomes at different stages. Specific enrichment of initiating ribosomes enables experimental identification of translation initiation sites (TISs). The feature of elevated ribosome density near the beginning of the coding region, although helpful in defining the start codon, is not sufficient for unambiguous identification of TISs. This is particularly true when multiple TIS sites exist on a single mRNA. Indeed, many alternative TISs are associated with either upstream open reading frames (uORFs) or downstream open reading frames (dORFs). It has been estimated that about 50% of mammalian transcripts contain at least one uORF (Calvo et al., 2009). To confound the situation further, many non-AUG codons can serve as alternative initiators (Hann et al., 1988; Meiron et al., 2001; Vagner et al., 1996). As there is no reliable way to predict alternative TISs from *in silico* sequence analysis, there is an urgent need to develop experimental approaches for genome-wide TIS identification.

Several strategies have been developed in order to achieve efficient capture of initiating ribosomes. Translation inhibitors such as harringtonine act on the first round of peptide bond formation. A short incubation period allows the run-off of elongating ribosomes, thereby specifically halting ribosomes at all possible TIS codons. Indeed, ribosome profiling of harringtonine-treated cells uncovered an unexpected abundance of alternative TISs, including many non-AUG initiators in the 5′ UTR (Ingolia et al., 2011). However, it appears that harringtonine does not completely arrest the initiating ribosomes. As a result, computational tool such as support vector machine (SVM)-based peak calling is needed to identify potential TIS sites. As harringtonine binds to the free 60S subunit first before it blocks the peptidyltransferase centre (PTC), it is uncertain whether the presence of this compound affects physiological TIS selection.

To overcome these problems and improve the resolution of TIS profiling, Lee et al. (2012) utilised a different translation inhibitor lactimidomycin to specifically capture the initiating ribosomes. Unlike harringtonine, lactimidomycin uses a mechanism similar but not identical to that used by cycloheximide (Schneider-Poetsch et al., 2010). Cycloheximide prevents translocation by locking the ribosome E-site. With an extra macrocycle, the much larger lactimidomycin can only bind to the
empty E-site of the ribosome (Garreau de Loubresse et al., 2014). As only the E-site of initiating ribosome is empty (the elongating ribosomes have their E-site occupied by a deacylated tRNA), lactimidomycin acts preferentially on the initiating ribosomes. Indeed, profiling of RPFs marked by lactimidomycin revealed remarkable enrichment of ribosomes at the annotated start codon. The single-nucleotide resolution permits identification of many alternative TISs with high confidence, as evidenced by experimental validation (Lee et al., 2012). In coupling with profiling of elongating ribosomes using cycloheximide, high-resolution profiling of initiating ribosomes greatly facilitates identification of hidden coding potential of genomes.

It would be desirable if the lactimidomycin-associated ribosome density reflects the rate of initiation in vivo. However, this is not always the case. During the runoff of elongating ribosomes, new rounds of initiation may occur. To circumvent the issue of 5′ end RPF inflation, Gao et al. (2015) devised a new approach for achieving direct capture of initiating ribosomes from cell lysates. Instead of relying on runoff to dissociate the elongating ribosomes, cell lysates were treated with puromycin to deplete the non-initiating ribosomes. Without the bias-introducing runoff incubation, the much improved quantitative feature permits direct comparison of initiation rates between normal and starvation conditions. This quantitative translation initiation sequencing (QTI-seq) promises great potential in illuminating the dynamic nature of translational regulation. Given the fact that much of the translational control occurs at the initiation stage (Sonenberg and Hinnebusch, 2009), systemic mapping of TIS positions and precise measurement of the corresponding initiation rate would transform our understanding of translational control.

### Tissue-specific ribosome profiling

Our understanding of the basic mechanisms of translation has reached a great deal of molecular detail, but much of this knowledge was obtained biochemically using cells in culture. Determining the precise biological role of such events in the intact organism remains to be a pressing and formidable challenge. The cellular heterogeneity of tissues and organs confounds our efforts to achieve cell-type-specific genomic and proteomic interrogation. Many techniques require some form of cell isolation, such as fluorescent activated cell sorting (FACS) or laser capture microdissection (LCM). In addition to the technical difficulties of these methods, the procedure of cell isolation inevitably introduces stresses and may cause ribosome re-distribution. Other methods for assaying translation in vivo are based on polysome mRNA pull-down, relying on tagged ribosomal proteins, such as HA-tagged Rpl22 (Sanz et al., 2009) or EGFP-fused Rpl10a (Heiman et al., 2008). Following genetic targeting to specific cell populations, affinity purification allows simple and rapid isolation of polysome-bound mRNAs from specific cell types amenable for profiling assays. Indeed, RiboTag mice have been successfully used to achieve liver cell-specific ribosome profiling (Gao et al., 2015) and brain tumour-specific translational landscapes (Gonzalez et al., 2014). It is conceivable that other techniques such as BacTRAP can achieve deep profiling of translatomes equally well in a cell-type-specific manner (Heiman et al., 2008).

The biggest challenge, however, is how to accomplish tissue-specific profiling of initiating ribosomes, which bear more reliable information in evaluating translational control. As QTI-seq directly captures the initiating ribosomes from cell lysates, the entire procedure should be applicable to tissue homogenates. This was indeed the case. Using liver-specific RiboTag mice, sequential treatment of liver homogenates with lactimidomycin and puromycin enriched initiating ribosomes (Gao et al., 2015). In comparison to mouse embryonic fibroblast cell, QTI-seq uncovered many genes whose translation is liver specific. The applicability of QTI-seq to solid tissues has the potential to greatly advance our understanding of the basic mechanisms of translation initiation from cellular to organismal levels.

### Bioinformatics of Ribosome Profiling

Like many high-throughput technology, one common theme associated with the advances of ribosome profiling is the increased reliance on computers for data processing and interpretation. Interpretation of high-throughput analysis requires both advanced programming skills and biological knowledge for insightful interpretation. It seems that the generation of data is no longer the rate-limiting step. Rather, the ability to effectively interpret large data sets is a significant roadblock. To facilitate this process, an online genome browser for viewing ribosome profiling data, GWIPS-viz, has been established (Michel et al., 2014). In addition, a searchable TIS database is also available for query of alternative TIS sites in mammalian cells (Wan and Qian, 2014). There is little doubt that integration of Ribo-seq data with other data sets, such as CHIP-seq, RNA-seq, miRNA profiling and proteomics, will present a fresh view of global post-transcriptional and translational gene regulation. The holistic pictures generated through ribosome profiling will surely allow us to see what we have not seen before and will reveal new paradigms for the coordinated regulation of the genome.

### Perspectives

Since the landmark paper describing the method of ribosome profiling (Ingolia et al., 2009), this powerful approach has been successfully applied to study translational control in diverse organisms ranging from bacteria to mammals, from total ribosomes to subpopulation of ribosomes and from cells in culture to solid tissues. The enormous biological breadth of translational control has led to an enhanced appreciation of proteome diversity and complexity. As we gain better insight into the mechanisms of translation, it is clear that the application of ribosome profiling will help paint a holistic picture of this paramount cellular process. Creative application of ribosome profiling in the past 5 years has immensely broadened the types of questions that can be posed and answered in the coming decade. For example, can the approach of ribosome profiling be adapted to monitor the scanning process of the 48S pre-initiation complex? Despite the strong
3-nucleotide periodicity revealed by ribosome profiling data, can the resolution be further improved to probe physiological frameshifting? Given the fact that the termination process is also regulated, is it possible to capture terminating ribosomes at the stop codon? Finally, there is much to be learned between ribosome binding versus protein production. How to isolate ribosomes engaged on mRNAs but with no protein synthesis? With the versatility of ribosome profiling on the rise, it will be exciting to watch the unveiling of answers to these questions and to see the inevitable surprises that will emerge.

References


initiation CUG codons located in its 5′UTR. *Biochemical and Biophysical Research Communications* **282**: 1053–1060.


**Further Reading**