Beginners guide to ribosome profiling

Luke Power (Ribomaps, Cork, Ireland)

Introduction

To synthesize proteins, cells must first transcribe an mRNA which specifies the sequence of amino acids, the building blocks of proteins. The next step involves translation of this mRNA into protein using the cell's protein-synthesizing machinery called ribosomes. Many gene expression studies rely solely on RNA-seq, which provides information of relative abundances of mRNAs in a cell; however, RNA-seq data ignore gene regulation at the translational level. Developed by Nicholas Ingolia and Jonathan Weissman, ribosome profiling (Ribo-seq) is a technique that provides a genome-wide view of *in vivo* translation. Ribo-seq is based on the principle that a translating ribosome protects a short stretch of mRNA within its structure. Once ribosomes are 'frozen' in the act of translation using translation elongation inhibitors, RNA-digesting enzymes known as RNases can be added to destroy any mRNA that is unprotected by the arrested ribosomes. After RNase digestion, ribosomes are enriched and the ribosome-protected mRNA is then isolated and converted into Illumina-compatible cDNA libraries. These ribosome-protected mRNA fragments are commonly called RPFs or ribosome footprints. Mapping these sequenced RPFs to the transcriptome provides a 'snapshot' of translation that reveals the positions and densities of ribosomes on individual mRNAs transcriptome-wide. This snapshot can help determine which proteins were being synthesized in the cell at the time of the experiment. Ribo-seq enables the identification of alternative mRNA translation start sites, the confirmation of annotated open reading frames (ORFs) as well as upstream (uORFs) that may be involved in the regulation of translation, the distribution of ribosomes on an mRNA and the rate at which ribosomes decode codons.

The steps involved in performing ribosome profiling

Harvesting and lysate preparation

Ribosome profiling protocols have been developed for budding yeast, mammalian cell lines, tissue samples, a range of bacterial species, plant and archaea. Each protocol follows a series of steps that are outlined in Figure 1. The first step is the lysis of cultured cells or collected tissue samples. These samples could be pre-treated with drugs or subjected to external stress conditions to investigate how these external factors impact the cell or tissue translationally. Lysis preparation is done in two parts, harvesting of cells/tissue and subsequent mechanical and chemical breakdown.

Harvesting is an important consideration due to the speed of translation (e.g., yeast ribosomes decode 9.6 codons a second), therefore the ribosomes must be harvested and flash-frozen immediately to preserve the position of the ribosome on mRNA. Harvesting can be done in many ways. One approach is vacuum filtration, in which a liquid culture is poured into a filter attached to a vacuum pump. As the cells remain on the filter they can then be scraped into a tube containing liquid nitrogen to flash-freeze the samples.

Samples are resuspended in an enriched polysome lysis buffer containing the following components:

- Salts such as sodium chloride, magnesium chloride or potassium chloride to stabilize the ribosomes,
- Some form of detergent (such as NP-40 or Triton X-100) to puncture the cell membrane, which is also important to isolate ribosomes bound to the endoplasmic reticulum,
- A deoxyribonuclease (such as Turbo DNAse) that degrades nuclear and mitochondrial/chloroplast genomic DNA,
- A translation-inhibiting drug (like cycloheximide or harringtonine) to stall the ribosome and entrap them over a region of mRNA,
- DTT, a reducing agent used to stop oxidative compounds from interfering with the RNA.

Organisms that contain a cell wall need to be first cryo-mechanically broken down using a cryogenic mixer mill or bead beater. Samples are clarified through centrifugation and the supernatant is recovered.

Ribonuclease digestion

Ribonucleases are enzymes that break down and destroy RNA. Here, they are employed to cleave regions of mRNA that reside outside of the ribosome, ideally leaving only the fragments of mRNA that are stored within and protected by the ribosome. The choice of ribonuclease is important as some ribonucleases are not compatible with certain

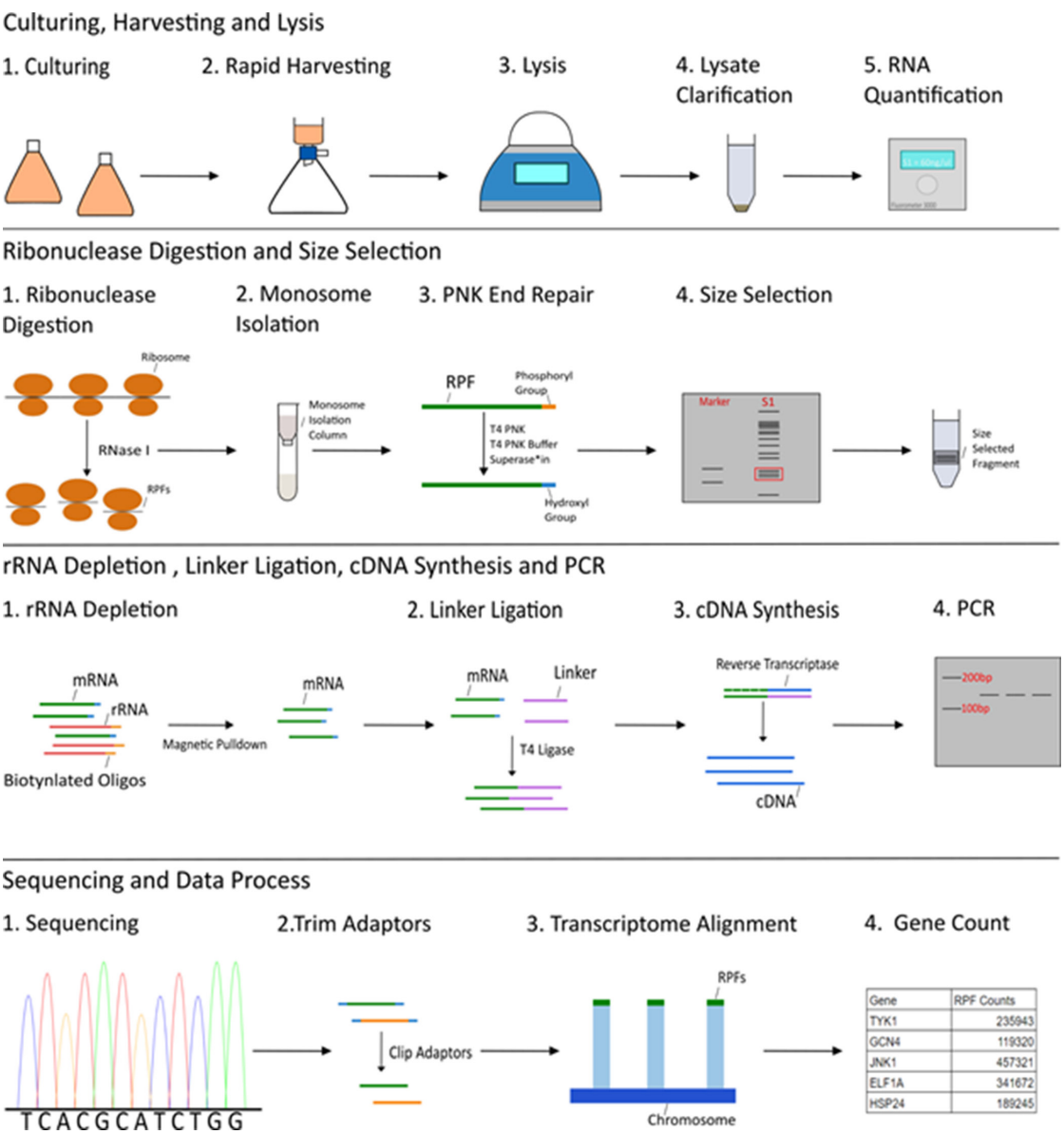

Figure 1. Schematic of the workflow involved with ribosome profiling, highlighting the most important steps of library generation.

species. For example, RNase I is a robust ribonuclease that is capable of providing good digestion in human cell types and yeast cell types but cannot be used for digestion in bacterial cell samples as it is capable of damaging the ribosome. Therefore, a different ribonuclease must be used (such as micrococcal nuclease, which offers a much more subtle digestion) to digest the mRNA outside of the ribosome. Each ribonuclease has its own strengths and weaknesses which should be taken into account when performing ribosome profiling.

The quality of the ribonucleic digestion can be visualized using plots such as a triplet periodicity plot or a metagene profile (such as the example shown in Figure 2) as they can show the triplet decoding of the ribosome. As codons are encoded in groups of three nucleotides, a strong digestion on a metagene profile would give a very clear repeating pattern where the number of RPFs in one frame would be much higher than in the other two frames.

Size selection

After nuclease digestion, the samples undergo T4 PNK end repair (which tailors the ends of RNA by removing the 3′-phosphoryl groups generated from RNase I cleavage and adds a hydroxyl group to prepare them for subsequent linker ligation), followed by fractionation on a 15% PAGE-urea gel. (Urea is added to denature the

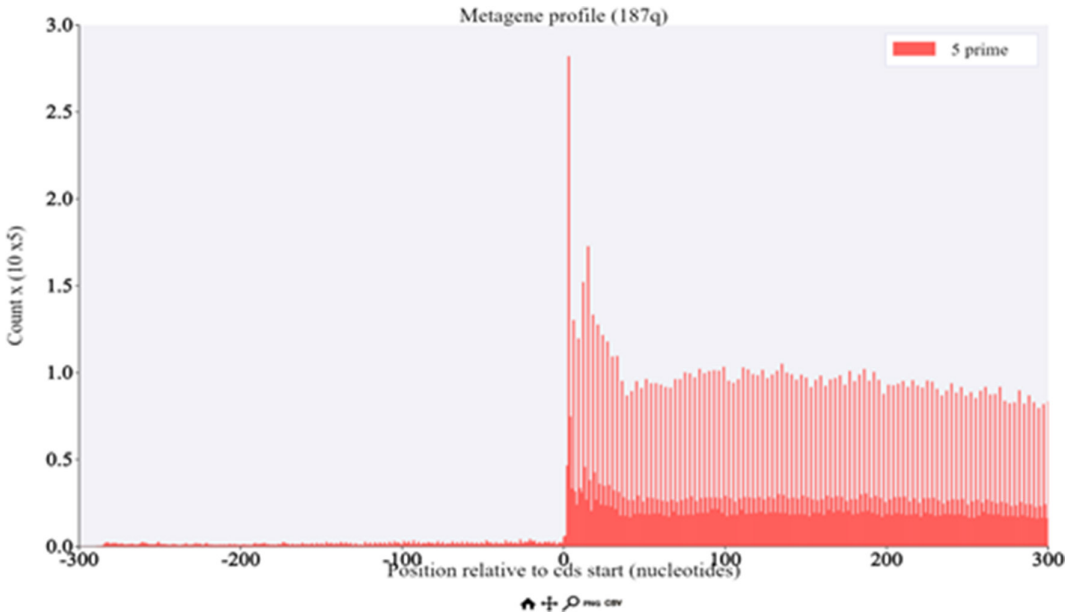

Figure 2. A metagene profile representing the number of RPF reads in a 300-nt region flanking the start of the CDS (*CoDing Sequence*). In this graph, there are multiple bars increasing vertically in groups of three. Each bar denotes one of the three frames an RPF can be decoded from. There is a clear triplet-like pattern observed in which the read counts follow a pattern of high–low–low, high–low–low, showing a large number of RPFs present in the first frame over subsequent frames, indicating strong RNAse digestion

RNA, thereby preventing any secondary structures from forming.) The use of a size selection marker allows for the determination of bands that are within a specific size. RPFs or footprints should be approximately 28 nucleotides in length (although this depends on the cell type). Therefore the size selection marker should contain bands that are just above and below the desired footprint size (e.g., 24 and 32 nucleotides in length). Figure 3 is an example of a gel photo which highlights the region where the ribosomal footprints are located. This use of bracketing the samples with the size selection markers allows for easy identification of the region of the gel to be excised out. Gel slices containing the size-selected RNA of interest are excised using a scalpel and the RNA is then extracted and purified.

rRNA depletion

After recovering the RNA from the gel slices, the next step is the removal of ribosomal RNA (rRNA). rRNA is the most predominant RNA (and can make up for about 80% of cellular RNA) in the cell. This is an issue, as an abundance of rRNA leads to fewer RPFs being sequenced, resulting in less mapping reads, essentially reducing the useful size of the library. Most contaminating rRNAs are generated from ribonuclease digestion of the ribosome, nicking off RPF-sized fragments. One way of depleting rRNA is using biotinylated oligos designed to hybridize to the predominant rRNAs. After mixing the samples of interest with a depletion mix containing the designed oligos, the biotinylated oligo-bound rRNA can be removed using streptavidin conjugated to magnetic beads. After this form of clean-up, the sample should be depleted of the

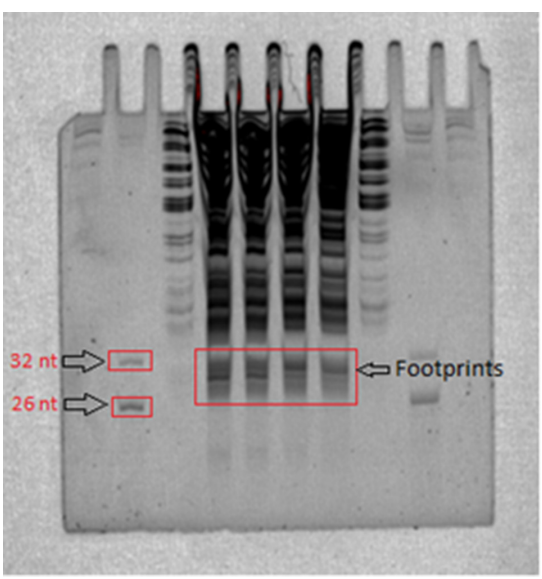

Figure 3. Size selection gel. In this gel photo; the footprints are highlighted within the red box. Samples were bracketed with size selection markers of 26 and 32 nucleotides in length

majority of its rRNA contents, thereby allowing for more RPFs to be sequenced later on.

Linker ligation and reverse transcription

In order to convert the ribosome footprints to DNA, they must be reverse transcribed. However, reverse transcriptase requires a primer to initiate polymerization. To overcome this, either the RPFs can be tailed with a poly A polymerase or else a single-stranded RNA linker of a known sequence can be ligated to the 3′-end of the RPF. There are several benefits of using 3′-linkers instead of polyadenylation tailing. One is the incorporation of random nucleotides at the 5′-end of the linker which can act as unique molecular identifiers (UMIs) to aid removal of PCR duplicates during subsequent analysis. Random nucleotides at the 5′-end also have the added benefit of reducing potential ligation biases. Another benefit of the addition of 3′-linkers is that they can be designed to contain unique barcodes for each linker, allowing for multiplexing (essentially combining different samples together into a single pool of samples for deep sequencing) prior to cDNA synthesis. Linker ligation can be done by using an enzyme known as T4 RNA ligase truncated K227Q in conjunction with adenylated linkers to join the ends of samples to the linker strands. The ligated product can then be isolated by either running the samples on a 15% PAGE-urea gel or via enzymatic linker digestion to cleave and remove any non-ligated linkers. Purified linker-ligated RPFs can then be converted into DNA by standard reverse transcriptase reaction.

Polymerase chain reaction

Following cDNA synthesis, libraries are amplified by polymerase chain reaction (PCR). PCR is a reaction that amplifies DNA exponentially, causing it to double in size every cycle. PCR is done in three stages (known as denaturation, annealing and extension) in which the DNA is subjected to rapid cycles of heating and cooling. These three stages are cycled through until a desired concentration of library is made. This library can then be sequenced to generate a bioinformatic library containing RPF sequenced reads.

Data analysis

The analysis of the sequencing data mainly depends on the aims of the experiment. A number of tools and pipelines are publicly available online that allow for a multitude of different data analyses, including uORF detection, differential gene expression, global translation rates, ribosome stalling, codon decoding rates, amongst others. A general flow of a ribosome profiling mapping pipeline would typically include the use of the software *FastQC* to determine basic quality metrics like read lengths and sequencing quality, utilizing another software called *Cutadapt* to demultiplex and cut away adapter sequences added during linker ligation and PCR, and *Bowtie* a short read aligner to bioinformatically remove remaining rRNA contaminants. The library can then be aligned to the organism's annotated genome/transcriptome, followed by the use of *Samtools* to convert the aligned reads into a sorted BAM file. From this, a gene count file can be generated using software such as *HT-seq* to count the number of reads aligned to each gene. A number of online browser-based platforms are available for visualizing ribosome profiles showing the mRNA positions of mapped RPFs allowing for further metadata analysis, each requiring different specific file types such as GTF files or Fasta files.

Conclusion and future prospects

One thing that is changing currently because of ribosomal profiling is genome annotation, as ribosome profiling pointed out that translation can occur outside of protein coding regions and that this translation is impactful, as it does something to the cell irrespective of whether it is productive or not. This advancement in genome annotation is exciting as it allows for a more advanced and in-depth look within the 'black box' that currently exists in genomes and to figure out why a particular change in their genome causes a particular phenotype or how a certain kind of chain of events occurs, and how this particular change affects gene expression and production of a particular protein, or maybe its own sequence. Ribosome profiling can accelerate our understanding of complex biological processes happening within the cell and, in turn, can be utilized to explore new ways for industrial exploitation. One such example is the possibility of accelerating and de-risking drug discovery through monitoring the side effects or toxicity of a drug on mRNA translation. ■

Further Reading

- Brar, G.A. and Weissman, J.S. (2015) Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nature Reviews Molecular Cell Biology*, 16, 651–664.
- McGlincy, N.J. and Ingolia, N.T. (2017) Transcriptome-wide measurement of translation by ribosome profiling. *Methods*, 126, 112–129.
- Kiniry, S.J., Judge, C.E., Michel, A.M. and Baranov, P.V. (2021) Trips-Viz: an environment for the analysis of public and user-generated ribosome profiling data. *Nucleic Acids Research*, 49, W662–W670.
- Yordanova, M.M., Loughran, G., Zhdanov, A.V., Mariotti, M., Kiniry, S.J., O'Connor, P.B.F., Andreev, D.E., Tzani, I., Saffert, P., Michel, A.M. et al. (2018) AMD1 mRNA employs ribosome stalling as a mechanism for molecular memory formation. *Nature*, 553, 356–360.
- Andreev, D.E., O'Connor, Patrick B.F., Loughran, G., Dmitriev, S.E., Baranov, P.V. and Shatsky, I.N. (2017) Insights into the mechanisms of eukaryotic translation gained with ribosome profiling. *Nucleic Acids Research*, 45, 513–526.
- van Heesch, S., Witte, F., Schneider-Lunitz, V., Schulz, J.F., Adami, E., Faber, A.B., Kirchner, M., Maatz, H., Blachut, S., Sandmann, C.L. et al. (2019) The Translational Landscape of the Human Heart. *Cell*, 178, 242–260.e229.
- Atger, F., Gobet, C., Marquis, J., Martin, E., Wang, J., Weger, B., Lefebvre, G., Descombes, P., Naef, F. and Gachon, F. (2015) Circadian and feeding rhythms differentially affect rhythmic mRNA transcription and translation in mouse liver. Proceedings of the National Academy of Sciences of the United States of America, 112, E6579–E6588.

Luke Power is a research scientist at Ribomaps Ltd. He received his Masters of Research (MRes) degree in translatomics from University College Cork. While researching in the Baranov lab in UCC, he successfully generated ribosome profiles of the oleaginous yeast Yarrowia lipolytica*, the first ribosome profiling data for this organism. His research interests include developing ribosome profiling to study translation in different species and optimizing the ribosome profiling protocol for both speed and data quality. Since joining Ribomaps, Luke has successfully performed ribosome profiling on cells, cell lines and tissues from many different species such as yeast, bacteria, plants, humans and other animals including rats and mice. His main duties include performing ribosome profiling on customer samples and the optimization and validation of protocols for different species. Email: Luke.Power@ribomaps.com*

