illumına®

Ribosome Profiling Enables Comprehensive Translation Analysis

Illumina's massively parallel sequencing advances analysis of genome-wide *in vivo* protein synthesis to individual nucleotide levels.
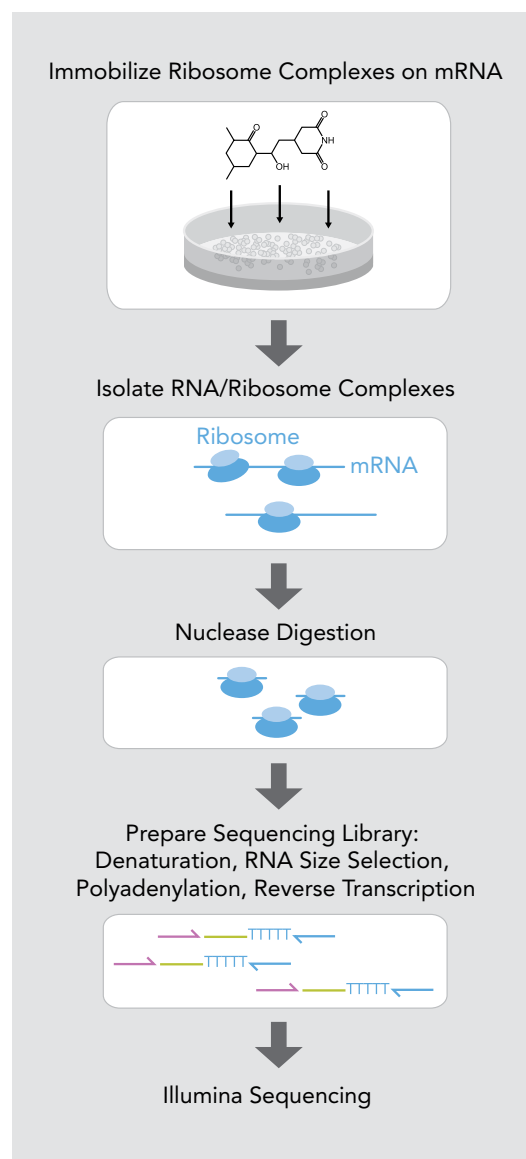
Introduction

Microarrays have the capacity to measure the expression levels of thousands of genes simultaneously[1]. Although microarrays have revolutionized mRNA analysis, they are only a proxy for the actual amount of protein synthesis occurring in cells. Multiple levels of transcriptional and post-transcriptional regulation contribute to the inherent challenges of proteomic analysis. A direct method for analyzing gene expression at the protein synthesis level is needed to investigate numerous regulatory events and other features that affect translational efficiency, such as mRNA splicing, ribosome pausing, and noncanonical initiation.

Ribosome footprinting is an established method for identifying the position of active ribosomes bound to target mRNA. Using nuclease digestion, the ribosome position and the translated message can be precisely determined by analyzing the protected ~30-nt area of the mRNA template[2]. After conversion of mRNA fragments to DNA, massively parallel sequencing provides millions of short reads, sampling each protein that is being actively translated. As shown in Figure 1, the ribosome profiling workflow begins with immobilization of ribosome complexes in the process of translating mRNA templates by cycloheximide treatment. RNA/ribosome complexes are harvested and digested with nuclease to trim away excess RNA. The ribosome-protected mRNA fragments are extracted and size-selected by gel electrophoresis. The resulting pool of 50–100 million purified RNAs represent all of the cellular ribosomes that were in the process of translation, and these 28-bp fragments are polyadenylated and reverse transcribed to make cDNA libraries for Illumina sequencing. Sequence identification determines the precise mRNA region being read by a ribosome, and the rate of protein synthesis is measured by examining the ribosome density of each RNA. The rate of protein synthesis can be defined under various conditions, providing a cellular snapshot of protein production.

In a recent study[3] Ingolia *et al.* provide proof-of-principle for ribosome profiling by examining protein synthesis in *Saccharomyces cerevisiae* that were grown under normal conditions and under the stress of amino acid starvation. The authors observed extensive translational control influencing protein abundance, differences in ribosome density between early and late peptide elongation, and an unexpected increase in initiation at favorable noncanonical initiation sites. Using Illumina sequencing to produce hundreds of millions of sequence reads per run marks a strategic advance towards more widespread use of high-throughput sequencing for transcriptomics and proteomics. As detailed in this report, ribosome profiling can be an essential tool for monitoring cell-wide protein production, with the potential to inform nearly every domain of biological and clinical research.

Figure 1: Ribosome profiling workflow

Immobilize Ribosome Complexes on mRNA

Isolate RNA/Ribosome Complexes

Ribosome

mRNA

Nuclease Digestion

Prepare Sequencing Library:
Denaturation, RNA Size Selection,
Polyadenylation, Reverse Transcription

Illumina Sequencing

Ribosome-bound mRNAs are purifed from cells, size-selected, polyadenylated, and reverse transcribed to make a library ready for Illumina sequencing.

Results and Data Analysis

Determining Reading Frame Specificity

A total of 42 million fragments from budding yeast were generated and sequenced using the ribosome profiling assay. Of these, 16% (~7 million) aligned to known CDS regions, and the rest were derived from ribosomal RNA (rRNA). The authors have subsequently shown that the majority of rRNA contamination can be removed by subtractive hybridization. For any given mRNA, the first ribosome footprint sequences were entered on codons in the protein coding portion of the transcript, extending from the start codon to the stop codon. Footprint positions showed a strong preference for their nucleotide position within each codon, allowing statistical inference of the reading frame.

Measuring the Rate of Protein Synthesis

The rate of protein synthesis was quantified by examining the ribosome footprint density compared to the total abundance of mRNA to estimate protein expression levels. An approximate 100-fold range of translation efficiency was detected between various yeast genes, in addition to a class of genes that were translationally inactive. Translational efficiency contributed substantially to the dynamic range of gene expression, a biologically informative phenomenon that cannot be elucidated by mRNA abundance measurements alone.

Measurements of absolute protein abundance by mass spectrometry correlate better with calculated protein translation rate than with RNA abundance measurements. Thus, the rate of protein synthesis is a better predictive measure of overall protein abundance than measuring mRNA levels (Figure 2). Measuring the rate of protein synthesis by ribosome profiling in relation to overall protein levels measured by mass spectrometry can reveal new information about regulated protein degradation[4].

Measuring Ribosome Density

Previous reports[5] provide evidence that short mRNA transcripts have higher ribosome densities, but it was not clear what caused this trend or how it corresponded to actual translation. Ribosome profiling showed that ribosomes are preferentially located at the 5′ region of all transcripts, and in shorter genes this region makes up a greater percentage of their total length. In particular, there was a three-fold higher ribosome density in the first 30–40 codons that tapered to a uniform density through the length of the transcript until termination. This appears to be a general feature of translation that is independent of coding sequence length or translation level. Correcting estimates of protein synthesis rates for this effect substantially improves the correlation between ribosome density and protein abundance.

Discovering Novel uORFs

In the yeast experiments, a majority of unexplained ribosome footprints (< 2% of the total) were found in the 5′UTR. Translation occurring in the 5′UTR could be due to upstream open reading frames (uORFs), short translated sequences upstream of the CDS that can play a role in translation regulation. The authors discovered 153 translated uORFs, of which only 30 had been previously described.

Characterizing Noncanonical Initiation Sites

Another significant source of translation at the 5′UTR is initiation at noncanonical sites that are flanked by a strong initiation sequencing context[6]. Sequence information from known examples in the yeast genome allowed the identification of other candidate non-AUG initiation sites. A total of 143 non-AUG ORFs were found to contribute to the 5′UTR translation events, suggesting that initiation at certain favorable non-AUG sites is a regular occurrence.

Figure 2: Protein abundance correlates better with ribosome density than with mRNA abundance
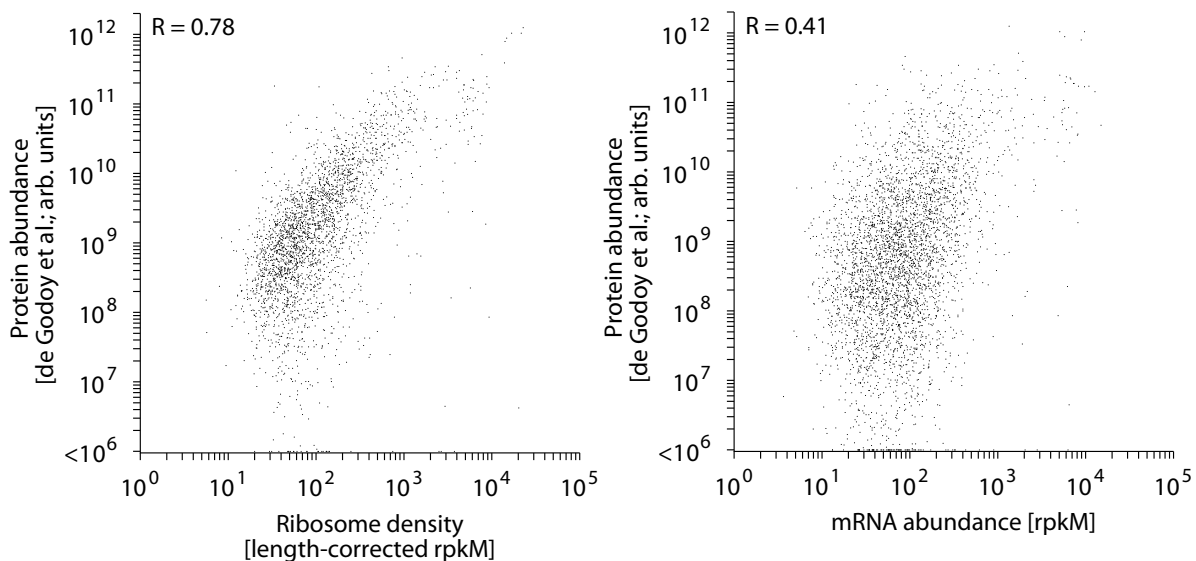
Correlation between protein abundance measured by mass spectrometry (summed ion intensity in haploid samples[4]) compared to ribosome density (left panel, R = 0.78) and mRNA abundance measurements (right panel, R = 0.41).

Figure 3: Changes in 5'UTR translation during starvation conditions in *S. cerevisiae*
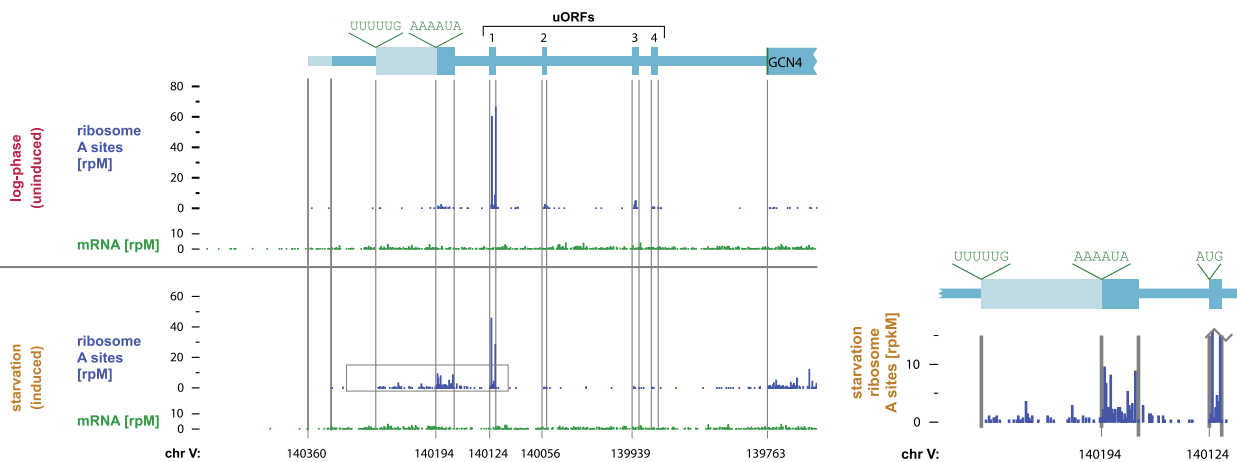
Left panel: Ribosome (blue) and mRNA (green) densities in the GCN4 5′UTR under repressive (log phase, upper) and inducive (starvation, lower) conditions. The four known uORFs are indicated. **Right panel:** Enlargement of boxed area in left panel, showing the ribosome density at proposed upstream translation initiation sites (UUUUUG and AAAAUA). From Ingolia *et al.* 2009 Science 324: 218–223. Reprinted with permission from AAAS.

Characterizing Translation Under Log Phase and Starvation Conditions

Acute amino acid starvation in yeast results in global transcription and translation changes, including a marked decrease in overall protein synthesis coupled with increased expression of certain target genes[7,8]. Ribosome profiling allowed further characterization of the sevenfold translational induction of *GCN4*, a well-studied, translationally regulated gene. GCN4 translation regulation results from four 5′UTR uORFs. As shown in the left panel of Figure 3, translation of GCN4 uORF 1 occurs with some translation of uORFs 2–4. However, little translation of the main GCN4 CDS was observed during log-phase growth. This pattern of ribosome occupancy is consistent with the standard model of GCN4 5′UTR function, in which uORF 1 is constitutively translated but permissive for downstream reinitiation. In log phase growth, reinitiation occurs at uORFs 2–4 rather than at GCN4 itself. Upon starvation, a decrease in ribosome occupancy of the repressive uORFs was observed, and increased re-initiation occurred at the CDS, removing translational repression and leading to increased translation of the protein coding region (Figure 3, right panel). During log-phase growth, additional translation in the 5′UTR upstream of the characterized uORFs was observed, and this translation was greatly enhanced under starvation conditions (Figure 3). Translation was initiated from a noncanonical A*AAAUA* site, as well as from an upstream, in-frame noncanonical UUU*UUG* site. Sequences overlapping this region are important for control of uORF 1[9], suggesting a functional role for novel, non-AUG uORFs.

Conclusions and Perspective

During amino acid starvation conditions in yeast, a large increase in ribosome footprints was observed, specifically in 5′UTRs. Under these conditions, uORFs with non-AUG initiation sites showed a particularly

dramatic increase in ribosome density compared to those observed under non-starvation conditions. This phenomenon may be due to general changes in noncanonical initiation site stringency, occurring through the action of global translational regulators.

Ribosome profiling is a new technique that enables systematic monitoring of cellular translation processes and prediction of protein abundance. Highly precise positional information allows discovery of translational control elements such as uORFs and initiation at noncanonical start sites, as well as other events in protein synthesis, including programmed frameshifts, stop codon readthrough, regulated protein degradation, premature translation termination, and splice variant expression. High-resolution, gene-specific ribosome density profiles will promote analysis of translation rate, ribosomal pausing, protein synthesis, and folding.

Powered by Illumina's massively parallel, high-throughput sequencing, ribosome profiling allows detailed and accurate *in vivo* analysis of protein production. The ribosomal profiling technique is adaptable to many organisms, and might be modified to include tissue-specific translational profiling[10]. Prospective applications of ribosome profiling include global studies of translational control of gene expression. In particular, determining what regions of a message are being translated can help define the proteome of complex organisms, as well as advance the molecular characterization of normal and disease states[11].

References

1. Brown PO, Botstein D (1999) Exploring the new world of the genome with DNA microarrays. Nat Genet 21: 33–37.
2. Steitz JA (1969) Polypeptide chain initiation: nucleotide sequences of the three ribosomal binding sites in bacteriophage R17 RNA. Nature 224: 957–964.

3. Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman J (2009) Genome wide analysis *in vivo* of translation with nucleotide resolution using ribosomal profiling. Science 324: 218–223.

4. de Godoy LM, Olsen JV, Cox J, Nielsen ML, Hubner NC, et al. (2008) Comprehensive mass spectrometry-based proteome quantification of haploid versus diploid yeast. Nature 455: 1251–1254.

5. Arava Y, Wang Y, Storey JD, Brown PO, Herschlag D (2003) Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA 100: 3889–3894.

6. Meijer HA, Thomas AA (2002) Control of eukaryotic protein synthesis by upstream open reading frames in the 5' untranslated region of an mRNA. Biochem J 367: 1–11.

7. Hinnebusch AG (1988) Mechanisms of gene regulation in the general control of amino acid biosynthesis in *Saccharomyces cerevisiae*. Microbiol Rev 52: 248–273.

8. Hinnebusch AG (2005) Translational regulation of GCN4 and the general amino acid control of yeast. Annu Rev Microbiol 59: 407–450.

9. Grant CM, Miller PF, Hinnebusch AG (1995) Sequences 5' of the first upstream open reading frame in GCN4 mRNA are required for efficient translational reinitiation. Nucleic Acids Res 23: 3980–3988.

10. Heiman M, Schaefer A, Gong A, Peterson JD, Day M, et al. (2008) A translational profiling approach for the molecular characterization of CNS cell types. Cell 135: 738–748.

11. Holcik M, Sonenberg N (2005) Translational control in stress and apoptosis. Nat Rev Mol Cell Biol 6: 318–327.

FOR RESEARCH USE ONLY