# Chapter 19

# Genome-Wide Profiling of Alternative Translation Initiation Sites

**Xiangwei Gao, Ji Wan, and Shu-Bing Qian**

## Abstract

Regulation of translation initiation is a central control point in protein synthesis. Variations of start codon selection contribute to protein diversity and complexity. Systemic mapping of start codon positions and precise measurement of the corresponding initiation rate would transform our understanding of translational control. Here we describe a ribosome profiling approach that enables identification of translation initiation sites on a genome-wide scale. By capturing initiating ribosomes using lactimidomycin, this approach permits qualitative and quantitative analysis of alternative translation initiation.

**Key words** Ribosome profiling, Translation, Initiation, Start codon, Genome-wide, Deep sequencing

## 1 Introduction

Translation initiation entails the ordered assembly of translation-competent ribosomes with initiator tRNA basepaired to the mRNA start codon at the ribosomal P-site [1]. In eukaryotes, the translation start codon is generally identified by the scanning mechanism [2]. It is believed that the first AUG codon encountered by the 48S preinitiation complex serves as the translation start codon. However, an increasing body of evidence suggests that eukaryotic ribosomes can recognize several alternative translation initiation sites (TISs) [3, 4]. It has been estimated that about 50 % of mammalian transcripts contain at least one upstream open reading frame (uORF) [5, 6]. To confound the matter further, many non-AUG triplets have been reported to act as alternative TISs for translation initiating [7]. Since there is no reliable way to predict non-AUG codons as potential initiators from *in silico* sequence analysis, there is a pressing need to develop experimental approaches for genome-wide TIS identification.

Ribosome profiling, based on deep sequencing of ribosome-protected mRNA fragments (RPF), has proven to be powerful in defining ribosome positions on the entire transcriptome [8].

Erik Dassi (ed.), *Post-Transcriptional Gene Regulation*, Methods in Molecular Biology, vol. 1358,
DOI 10.1007/978-1-4939-3067-8_19, 

However, the standard ribosome profiling captures all the ribosomes engaged on the mRNA. Therefore, it is not optimized for identifying TIS positions. Several strategies have been developed in order to achieve efficient capture of initiating ribosomes. Translation inhibitors like harringtonine act on the first round of peptide bond formation and potentially stall the initiating ribosomes [9]. Puromycin dissociates the ribosome into subunits but exhibits higher sensitivity towards elongating ribosomes than initiating ribosomes [10]. However, both compounds do not seem to arrest the ribosome at the start codon in a definitive manner. To improve the resolution of TIS mapping, we took advantage of a translation inhibitor lactimidomycin (LTM) that acts preferentially on the initiating ribosomes [11]. Two different approaches were developed to effectively capture the initiating ribosomes from mammalian cells (Fig. 1). The first approach, global translation initiation sequencing (GTI-seq), permits high-resolution mapping of TIS positions via LTM pretreatment [12]. The second approach, quantitative translation initiation sequencing (QTI-seq), enables

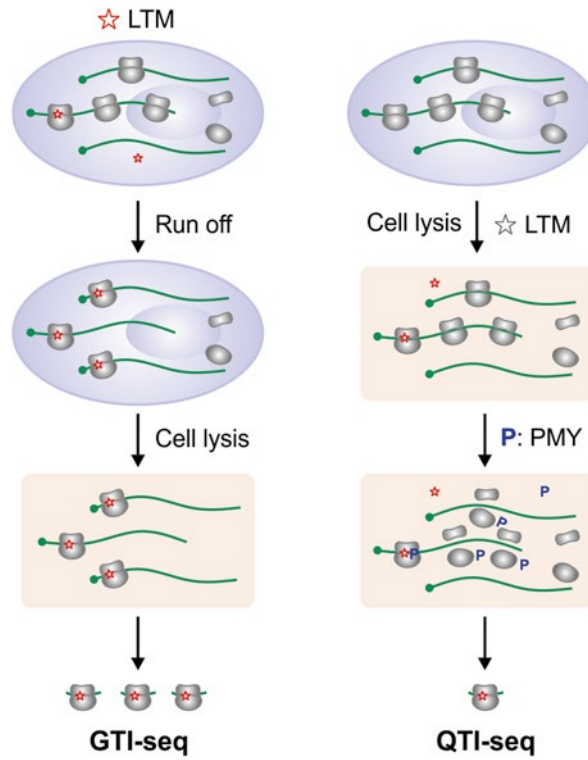


**Fig. 1** Schematic of GTI-seq (*left*) and QTI-seq (*right*). In GTI-seq, cells are pretreated with LTM to immobilize the initiating ribosomes followed by an incubation to dissociate the elongating ribosomes. In QTI-seq, LTM directly captures the initiating ribosomes from the lysates and relies on puromycin (PMY) to dissociate the elongating ribosomes

enrichment of initiating ribosomes from whole cell lysates [13]. Along with detailed methods, both the advantages and the disadvantages of these approaches are discussed in this chapter.

## 2 Materials

### 2.1 Preparation of Cell Lysates for GTI-seq

1. Growth medium: Dulbecco's modified Eagle medium (DMEM) supplemented with 10 % fetal bovine serum (FBS).
2. Phosphate-buffered saline (PBS).
3. 10 cm petri dishes.
4. Lactimidomycin: Dissolve in DMSO to the concentration of 50 mM and store it at −80 °C.
5. Cycloheximide: Dissolve in DMSO to the concentration of 100 mg/mL and store it at −80°C.
6. GTI-seq lysis buffer: 20 mM Hepes, pH 7.4, 100 mM KCl, 5 mM $MgCl_2$, 100 μg/mL cycloheximide, 2 % Triton X-100.

### 2.2 Preparation of Cell Lysates for QTI-seq

In addition to the materials for GTI-seq, other materials are listed below:

1. Lysing Matrix D.
2. QTI-seq buffer: 20 mM Hepes, pH 7.4, 100 mM KCl, 5 mM $MgCl_2$.
3. Puromycin: Dissolve in dd$H_2O$ to the concentration of 20 μg/mL and store it at −20 °C.
4. Creatine phosphate: Dissolve in nuclease-free $H_2O$ to the concentration of 1 M and store it at −20 °C.
5. Spermidine: Dissolve it in nuclease-free $H_2O$ to the concentration of 10 mM and store it at −20 °C.
6. Creatine phosphokinase: Dissolve in nuclease-free $H_2O$ to the concentration of 10 mg/mL and store it at −20 °C.
7. ATP solution (10 mM).

### 2.3 Sucrose Gradient Sedimentation

1. Polysome buffer: 10 mM Hepes, pH 7.4, 100 mM KCl, 5 mM $MgCl_2$.
2. Sucrose solution: First make the 60 % (wt/vol) sucrose solution using polysome buffer. Mix 1 volume of 60 % sucrose and 3 volumes of polysome buffer to get 15 % sucrose. Mix 3 volumes of 60 % sucrose and 1 volume of polysome buffer to get 45 % sucrose.
3. SW41 ultracentrifuge tubes.
4. Sucrose gradient maker.
5. Automated fractionation system.

### 2.4 RNase I Digestion and RPF Extraction

1. RNase I.
2. TRIzol LS reagent.
3. T4 Polynucleotide Kinase.
4. SUPERase_In.
5. 2× Novex TBE-urea sample buffer.
6. 10× Novex TBE-urea gel running buffer.
7. Novex 15 % denaturing polyacrylamide TBE-urea gel.
8. SYBR Gold: Dilute to 1:10,000 in TAE buffer and store it at −20 °C.
9. Spin-X column.
10. Glycogen (5 mg/mL).
11. 3 M NaOAc (pH 5.2).
12. RNA gel elution buffer: 300 mM NaOAc (pH 5.5), 1 mM EDTA, and 0.1 U/μL SUPERase_In. Store the buffer at −20 °C.

### 2.5 cDNA Library Construction

1. *E. coli* Poly(A) polymerase.
2. SuperScript III Reverse Transcriptase.
3. RNaseOUT Recombinant Ribonuclease Inhibitor:.
4. Reverse transcription primers: 5′-pGATCGTCGGACTGTA GAACTCTØCAAGC AGAAGACGGCATACGATTTTTTTT TTTTTTTTTTTTVN-3′. The initial p indicates 5′ phosphorylation, Ø indicates the abasic dSpacer furan, and V and N indicate degenerate nucleotides.
5. Novex 10 % denaturing polyacrylamide TBE-urea gel.
6. DNA gel elution buffer: 300 mM NaCl and 1 mM EDTA. Prepare in advance and store at 4 °C.

### 2.6 DNA Circularization and Re-linearization

1. Circularization kit.
2. Ape I.

### 2.7 DNA Amplification and Deep Sequencing

1. Phusion High-Fidelity DNA Polymerase.
2. Oligo primers for PCR: sense 5′-CAAGCAGAAGACGGC ATA-3′; antisense: 5′-AATGATACGGCGACCACCGACAG GTTCAGAGTTCTACAGTCCGACG-3′ to generate DNA species containing Illumina cluster generation sequences on each end and a sequencing primer-binding site.
3. Novex 8 % denaturing polyacrylamide TBE-urea gel.
4. 5× Novex TBE Running Buffer.

## 3 Methods

### 3.1 Preparation of Cell Lysates for GTI-seq

In this approach, human embryo kidney (HEK) 293 cells are pretreated with LTM for 30 min to allow the elongating ribosomes to run off. Different cell lines have varied LTM sensitivity and may require pilot experiments to determine both the LTM concentration and the incubation period (*see* **Note 1**). The following experimental procedure is based on HEK293 cells.

1. Grow HEK293 cells in growth medium in a 5 % $CO_2$ cell culture incubator with 95 % humidity. Split HEK293 cells the day before the treatment to at least four 10 cm petri dishes. Adjusting the splitting ratio so that the cells are at approximately 80 % confluence on the day of experiment.
2. Replace the medium with 5 mL fresh, pre-warmed medium containing 50 μM LTM followed by incubation for 30 min (*see* **Note 2**).
3. Aspirate medium from each dish and immediately cool the dishes on ice. Gently wash the cells with 5 mL of ice-cold PBS containing 100 μg/mL of cycloheximide.
4. Aspirate the PBS thoroughly from all the petri dishes and place them on ice.
5. Add 400 μL of ice-cold lysis buffer to one petri dish (*see* **Note 3**).
6. Detach the cells by scraping the entire petri dish and pipetting up and down several times to lyse the cells.
7. Transfer the lysates from the first petri dish to the second dish and lyse the cells. Repeat **step 6** until all the petri dishes are done (*see* **Note 4**).
8. Transfer the cell lysates from the last petri dish to a 1.6 mL Eppendorf tube and place the tube on ice.
9. Remove the debris from the cell lysates by centrifugation for 10 min at 13,000 × *g* at 4 °C. Transfer the soluble supernatant to a new 1.6 mL Eppendorf tube and store the tube on ice.

### 3.2 Preparation of Cell Lysates for QTI-seq

In this approach, human embryo kidney (HEK) 293 cells are used to enrich initiating ribosomes from the cell lysates. Sequential treatment of LTM and puromycin is conducted on the cell lysates. The detailed experimental procedure is described below.

1. Grow HEK293 cells in growth medium in a 5 % $CO_2$ cell culture incubator with 95 % humidity. Split HEK293 cells the day before the treatment to at least four 10 cm petri dishes. Adjusting the splitting ratio so that the cells are at approximately 80 % confluence on the day of experiment.
2. Aspirate medium from each dish and immediately cool the dishes on ice. Gently wash the cells with 5 mL of ice-cold PBS.

3. Aspirate the PBS thoroughly from all the petri dishes and place them on ice.
4. Add 400 μL of ice-cold QTI-seq buffer containing 5 μM LTM to the first petri dish (*see* **Note 4**).
5. Detach the cells by scraping the entire petri dish and pipetting up and down several times to suspend all the cells.
6. Transfer the buffer from the first petri dish to the second dish to detach the cells. Repeat **step 5** until all the petri dishes are done.
7. Transfer the collected cells from all the petri dishes to a 2 mL Eppendorf tube containing Lysing Matrix-D and place the tube on ice.
8. Lyse cells by vortexing 20 s for six times with a 40-s interval on ice.
9. Remove the debris from the cell lysates by centrifugation for 10 min at 13,000 × *g* at 4 °C. Transfer the soluble supernatant to a new 1.6 mL Eppendorf tube and store the tube on ice.
10. Supplement the following reagents into the cell lysates: 10 mM creatine phosphate, 0.1 mM spermidine, 40 μg/mL creatine phosphokinase, 0.8 mM ATP, and 25 μM of puromycin. Incubate the mixture at 35 °C for 15 min.

### 3.3 Sucrose Gradient Sedimentation

For both GTI-seq and QTI-seq, it is important to examine whether the dissociation of elongating ribosomes is complete. Sucrose gradient-based polysome profiling is commonly used for this purpose. As shown in Fig. 2, a complete dissociation of elongating ribosomes results in disassembly of polysome fractions with a corresponding increase of monosome.

1. Freshly make 12 mL sucrose density gradients in each SW41 ultracentrifuge tube. Use the BioComp Gradient Master (BioComp Inc.) to generate a 15–45 % (wt/vol) gradient. In brief, first fill the tube with 6 mL of 15 % sucrose solution

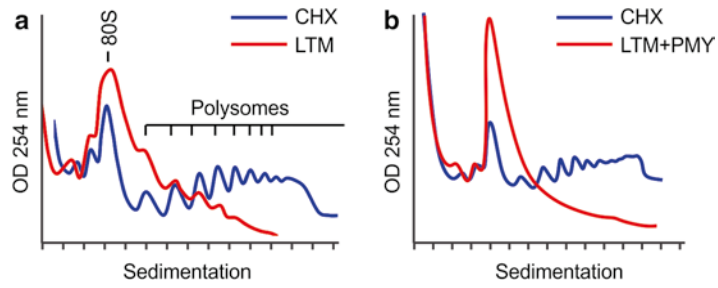


**Fig. 2** Polysome profiling to examine the dissociation of elongating ribosomes during GTI-seq (**a**) and QTI-seq (**b**)

and then add 6 mL of 45 % sucrose solution to the bottom of the tube using needle and syringe. Follow the manufacturer's instruction to mix the two layers of sucrose solution.

2. Gently load 600 μL of cell lysate onto the top surface of the sucrose gradients, followed by centrifugation for 100 min at 38,000 × *g*, 4 °C, in an SW41 rotor (*see* **Note 5**).
3. Fractionate the separated samples using an automated fractionation system (Isco) with continuous monitoring of $OD_{254}$ values. Fractions are collected into 1.6 mL Eppendorf tubes at 1.5 mL/min with 0.5 min per tube. The collected fractions can be either stored at −80 °C or immediately used for RNase I digestion.

### 3.4 RNase I Digestion and RPF Extraction

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 over-digestion. Following nuclease digestion, the RPFs are resolved on a gel and collected via a size selection. The typical RPF size in mammalian cells is around 28 nt (Fig. 3).

1. Take 10 μL from each fraction starting from the monosome and mix thoroughly to get a total of 300 μL ribosome fractions in a 1.6 mL Eppendorf tube.
2. Add *E. coli* RNase I into the mixed ribosome fractions (750 U per 100 $A_{260}$ units). Incubate the mixture at 4 °C for 1 h to convert the polysome into monosome (*see* **Note 6**).
3. Add 3 volume of TRIzol LS reagent to the digested ribosome samples. Extract total RNA according to the manufacturer's instructions. Dissolve the total RNA in 11 μL of RNase-free $H_2O$.
4. Dephosphorylate the purified RNA samples in a 15-μL reaction mixture by adding 1.5 μL 10× T4 polynucleotide kinase buffer, 0.5 μL RNase inhibitor SUPERase_In (20 U/μL), and 2 μL T4 polynucleotide kinase (10 U/μL). Incubate the reaction mixture at 37 °C for 2 h, followed by heat-inactivation at 65 °C for 20 min.
5. Mix the dephosphorylated sample with 15 μL 2× Novex TBE-Urea sample buffer (Invitrogen). Denature the sample in a heat block at 70 °C for 3 min. Load the samples on a Novex 15 % denaturing polyacrylamide TBE-urea gel (Invitrogen) and run at 160 V for 1.5 h.
6. Stain the gel with SYBR Gold to visualize the RNA fragments. Excise the gel bands corresponding to 28 nt RNA species (Fig. 3).

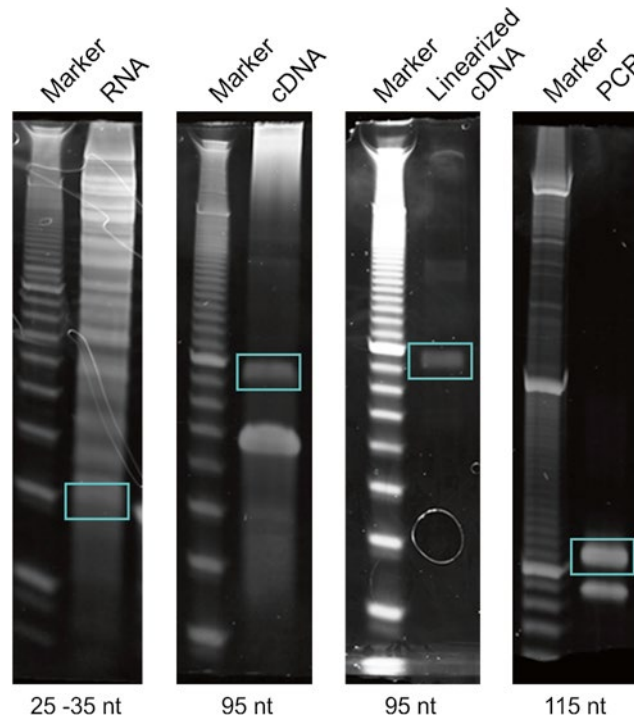


**Fig. 3** Size selection during cDNA library construction. The ribosome-protected mRNA fragments, single stranded DNA, linearized DNA, and PCR products are *highlight in box*

7. Physically disrupt the gel by using centrifugation through the bottom holes of the tube. Extract the RNA fragments by soaking the gel slices overnight in RNA gel elution buffer.
8. Remove the gel debris using a Spin-X column by centrifugation at 12,000 × *g* for 2 min at room temperature.
9. Mix the elution with 4 μL glycogen, 40 μL NaOAc (3 M, pH 5.2), and 900 μL ice-cold ethanol. Place the sample at −20 °C for 30 min.
10. Centrifuge the mixture at 14,000 × *g* for 15 min at 4 °C. Wash the pellet once with 70 % ethanol and centrifuge at 14,000 × *g* for 5 min at 4 °C.
11. Aspirate the liquid and air-dry the pellet. Dissolve the RNA pellet in 5.2 μL of nuclease-free $H_2O$.

### 3.5 cDNA Library Construction

The entire procedure of cDNA library construction from purified RPF is similar to regular ribosome profiling [14], which includes poly(A) tailing and reverse transcription. The custom-designed primers may include barcodes for multiplex analysis (*see* **Note 7**).

1. Perform poly-(A) tailing reaction in a 8 μL reaction mixture by adding 0.8 μL poly-(A) polymerase buffer (10×), 1 μL ATP

(10 mM), 0.4 μL SUPERase_In, and 0.6 μL *E. coli* poly-(A) polymerase (5 U/μL). Incubate the reaction mixture at 37 °C for 45 min.

2. Mix the poly(A)-tailed RNA samples with 1 μL dNTP (10 mM) and 1 μL synthesized primer (25 mM). Incubate the reaction mixture at 65 °C for 5 min followed by ice incubation for 5 min.
3. For reverse transcription, set up a 20 μL reaction mixture by adding 2 μL 10× buffer, 4 μL $MgCl_2$ (25 mM), 2 μL DTT (100 mM), 1 μL RNaseOUT (40 U/μL), and 1 μL SuperScript III (200 U/μL). Incubate the reaction mixture at 50 °C for 50 min.
4. Mix the reverse transcription sample with 20 μL Novex TBE-Urea sample buffer (2×). Denature the sample in a heat block at 70 °C for 3 min. Load the samples on a Novex 10 % denaturing polyacrylamide TBE-urea gel and run at 160 V for 1.5 h.
5. Stain the gel with SYBR Gold to visualize the single-strand DNA. Excise the gel bands corresponding to the size of 95 nt (Fig. 3).
6. Physically disrupt the gel by using centrifugation through the bottom holes of the tube. Extract the single-strand DNA by soaking the gel slices overnight in DNA gel elution buffer.
7. Remove the gel debris using a Spin-X column by centrifugation at 12,000 × *g* for 2 min at room temperature.
8. Mix the elution with 4 μL glycogen, 40 μL NaOAc (3 M, pH 5.2), and 900 μL ice-cold ethanol. Place the sample at −20 °C for 30 min.
9. Centrifuge the mixture at 14,000 × *g* for 15 min at 4 °C. Wash the pellet once with 70 % ethanol and centrifuge at 14,000 × *g* for 5 min at 4 °C.
10. Aspirate the liquid and air-dry the pellet. Dissolve the DNA pellet in 11.5 μL nuclease-free $H_2O$.

### 3.6 DNA Circularization and Re-linearization

1. Circularize the single-stranded DNA in 20 μL of reaction mixture by adding 2 μL CircLigase buffer (10×), 1 μL $MnCl_2$ (50 mM), 4 μL Betaine (5 M), and 1.5 μL CircLigase II (100 U/μL). Circularization is performed at 60 °C for 1.5 h followed by heat-inactivation at 80 °C for 10 min (*see* **Note 8**).
2. Purify the circularized DNA using ethanol precipitation mentioned above. Dissolve the DNA pellet in 8.25 μL of nuclear-free $H_2O$.
3. Re-linearize the circular single-stranded DNA in 10 μL of reaction mixture by adding 0.75 μL APE 1 (10 U/μL) and 1 μL Buffer 4 (NEB). The reaction is carried out at 37 °C for 1 h (*see* **Note 9**).

4. Mix the re-linearized DNA sample with 10 μL Novex TBE-Urea sample buffer (2×). Denature the sample in a heat block at 70 °C for 3 min. Load the samples on a Novex 10 % denaturing polyacrylamide TBE-urea gel (Invitrogen) and run at 160 V for 1.5 h.
5. Stain the gel with SYBR Gold to visualize the single-strand DNA. Excise the DNA gel band corresponding to the size of 95 nt (Fig. 3).
6. Physically disrupt the gel by using centrifugation through the bottom holes of the tube. Extract the single-strand DNA by soaking the gel slices overnight in DNA gel elution buffer.
7. Remove the gel debris using a Spin-X column by centrifugation at 12,000 × *g* for 2 min at room temperature.
8. Mix the elution with 4 μL glycogen, 40 μL NaOAc (3 M, pH 5.2), and 900 μL ice-cold ethanol. Place the sample at −20 °C for 30 min.
9. Centrifuge the mixture at 14,000 × *g* for 15 min at 4 °C. Wash the pellet once with 70 % ethanol and centrifuge at 14,000 × *g* for 5 min at 4 °C.
10. Aspirate the liquid and air-dry the pellet. Dissolve the DNA pellet in 10 μL nuclease-free $H_2O$.

### 3.7 DNA Amplification and Deep Sequencing

1. Amplify the single-stranded DNA template by setting up a 25 μL PCR reaction mixture containing 5 μL DNA template, 5 μL HF buffer (5×), 0.5 μL dNTP (10 mM), 1.25 μL each primer (10 mM), 0.5 μL Phusion polymerase (2 U/μL), and 11.5 μL distilled $H_2O$.
2. Perform PCR with an initial denaturation at 98 °C for 30 s, followed by 8–14 cycles of 10-s denaturation at 98 °C, 20-s annealing at 60 °C, and 10-s extension at 72 °C (*see* **Note 10**).
3. Separate PCR products on a nondenaturing 8 % polyacrylamide TBE gel by running at 160 V for 1 h.
4. Stain the gel with SYBR Gold to visualize the PCR product. Excise the DNA band corresponding to the size of 115 bp (Fig. 3)
5. Physically disrupt the gel by using centrifugation through the bottom holes of the tube. Extract the single strand DNA by soaking the gel slices overnight in DNA gel elution buffer.
6. Remove the gel debris using a Spin-X column by centrifugation at 12,000 × *g* for 2 min at room temperature.
7. Mix the elution with 4 μL glycogen, 40 μL NaOAc (3M, pH 5.2), and 900 μL ice-cold ethanol. Place the sample at −20 °C for 30 min.

8. Centrifuge the mixture at 14,000 × *g* for 15 min at 4 °C. Wash the pellet once with 70 % ethanol and centrifuge at 14,000 × *g* for 5 min at 4 °C.
9. Aspirate the liquid and air-dry the pellet. Dissolve the DNA pellet in 15 μL nuclease-free $H_2O$.
10. Quantify the PCR product by Agilent BioAnalyzer DNA 1000 assay.
11. Use approximately 3–5 pmol mixed DNA samples for cluster generation followed by sequencing using Illumina sequencing primer.

### 3.8 Sequencing Data Analysis

We conduct the sequencing data analysis using custom made R and Python scripts. The quality of TIS profiles can be evaluated by an aggregation plot showing all the RPF density across the transcriptome aligned at the annotated start codon. A high-quality TIS mapping is expected to see a significantly higher peak at the annotated start codon than other positions (Fig. 4a). A typical example of single gene TIS mapping is shown in Fig. 4b.

1. Categorize the raw reads of sequencing data into four groups according to barcode information (TG, AC, GA, and CT). After removing the 2-nt barcode, trim the reads by 10 nt from the 3′ end. The shortened reads are further trimmed by removing adenosine (A) stretch from the 3′ end with one mismatch allowed. Retain the processed reads for following analysis (*see* **Note 11**).
2. Map processed reads to transcriptome and hereafter unmapped reads to the corresponding genome using Tophat with parameters (--bowtie1 -p 10 --no-novel-juncs –G) [15]. Discard non-uniquely mapped reads to rule out ambiguity for the further analysis (*see* **Note 12**).

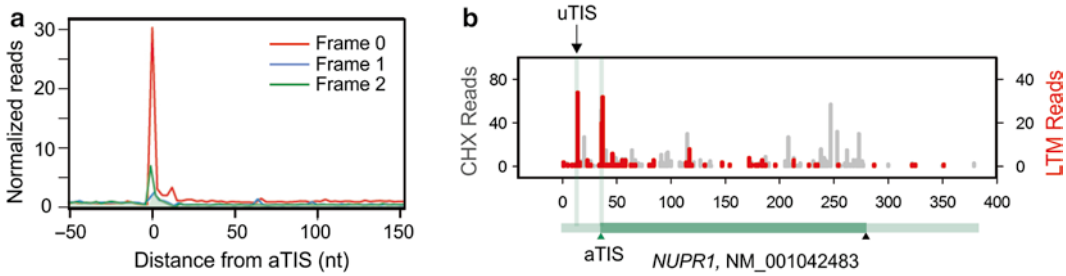


**Fig. 4** Sequencing data analysis. (**a**) Aggregation plot of LTM-associated ribosome density in HEK293 cells captured by QTI-seq. Normalized RPF reads are averaged across the entire transcriptome and aligned at the annotated start codon. Different reading frames are separated and color coded. (**b**) An example of gene (NUPR1) with an alternative TIS captured by QTI-seq. The corresponding gene structure is shown below the *X*-axis with *green triangle* as the annotated start codon, and *black triangle* as the stop codon

3. Define the 13th position (12 nt offset from the 5′ end) of the uniquely mapped read as the ribosome "P-site". Calculate the P-site density on each position of individual mRNA transcript according to the NCBI Refseq annotation.
4. Apply Zero-Truncated Negative (ZTNB) model for identifying statistically significant P-site peak on the mRNA. Fit a global ZTNB model over all the non-empty P-sites of the entire transcriptome; and for each individual transcript with more than 50 distinct P-site positions, fit a local ZTNB model of the non-zero P-sites (*see* **Note 13**).
5. Decide putative "start codon" as a peak by the following criteria: (a) *p*-value thresholds (0.05 for global ZTNB model and 0.01 for local ZTNB model); (b) the peak height should be a local optimal number in a fixed window (−15, +15) (*see* **Note 14**).
6. To make pairwise comparison for TIS efficiency between different conditions, apply upper quartile (UQ) normalization to each predicted TIS sites based on the population of total read count on each individual mRNA. Calculate the fold change of each TIS to estimate the differential translational initiation signals between two experimental conditions. Calculate the abundance of translation initiation signal using a window centering the predicted TIS codon (−1, +4) because read length variation in the QTI-Seq data may cause offset problem in defining P-site (*see* **Note 15**).

## 4 Notes

1. The data quality of GTI-seq and QTI-seq appears to be dependent on cell lines. Some cells are more sensitive to LTM treatment than the others.
2. LTM is sensitive to light exposure. Shield the stock solution during the procedure.
3. To maximize the protein concentration of the lysates, it is recommended to use the minimum amount of the lysis buffer but sufficient to cover the petri dish.
4. Use the same lysis buffer for multiple dishes can minimize the total volume of the cell lysates.
5. Remove 600 μL of sucrose from the top of the gradient prior to the addition of the same amount of the sample. Do not disturb the gradient.
6. To avoid incomplete digestion or over-digestion, a pilot experiment is recommended to determine the digestion condition. A second sucrose gradient sedimentation of the sample is suggested to examine the status of digestion.

7. The sequence of bar code could be included in the custom designed primers for reverse transcription.
8. CircLigase I is also recommended for this step.
9. The step of re-linearization is optional. The circularization product can be used directly for PCR amplification.
10. It is recommended to minimize the cycles of PCR. A pilot PCR reaction with different cycles is suggested to determine the optimal cycles.
11. The size of RPF may affect the mapping accuracy. It is recommended to examine the size distribution of RPF as part of the quality control.
12. The non-uniquely mapped reads can be retained for certain analysis since they are true RPFs.
13. Global ZTNB model, which corresponds to a transcriptome-wide model trained on all the data, aims to suggest a peak height which is statically significant; while local ZTNB-model, which corresponds to a model only built on the data of individual gene, aims to compare peaks within the gene as well as control for the variation in gene expression levels.
14. It is suggested to compare different thresholds for optimal parameters to control false positive predictions.
15. Since the length range of RPFs is 25–35 nt, an offset of 1 or 2 nt is possible in TIS identification. A clustering step can be used to group neighboring putative TIS sites (within 1 or 2 nt) to avoid redundant peak calling.

## Acknowledgements

This work was supported by National Institutes of Health (DP2 OD006449, R01AG042400), Ellison Medical Foundation (AG-NS-0605-09), and US Department of Defense (W81XWH-14-1-0068).

### References

1. Hinnebusch AG (2014) The scanning mechanism of eukaryotic translation initiation. Annu Rev Biochem 83:779–812
2. Kozak M (2002) Pushing the limits of the scanning mechanism for initiation of translation. Gene 299:1–34
3. Iacono M, Mignone F, Pesole G (2005) uAUG and uORFs in human and rodent 5'untranslated mRNAs. Gene 349:97–105
4. Morris DR, Geballe AP (2000) Upstream open reading frames as regulators of mRNA translation. Mol Cell Biol 20:8635–8642
5. Calvo SE, Pagliarini DJ, Mootha VK (2009) Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. Proc Natl Acad Sci U S A 106:7507–7512
6. Resch AM, Ogurtsov AY, Rogozin IB, Shabalina SA, Koonin EV (2009) Evolution of alternative and constitutive regions of mammalian 5'UTRs. BMC Genomics 10:162
7. Touriol C, Bornes S, Bonnal S, Audigier S, Prats H, Prats AC, Vagner S (2003) Generation of protein isoform diversity by alternative initi-

ation of translation at non-AUG codons. Biol Cell 95:169–178

8. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324:218–223
9. Ingolia NT, Lareau LF, Weissman JS (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell 147:789–802
10. Fritsch C, Herrmann A, Nothnagel M, Szafranski K, Huse K, Schumann F, Schreiber S, Platzer M, Krawczak M, Hampe J, Brosch M (2012) Genome-wide search for novel human uORFs and N-terminal protein extensions using ribosomal footprinting. Genome Res 22:2208–2218
11. Schneider-Poetsch T, Ju J, Eyler DE, Dang Y, Bhat S, Merrick WC, Green R, Shen B, Liu JO (2010) Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin. Nat Chem Biol 6:209–217
12. Lee S, Liu B, Huang SX, Shen B, Qian SB (2012) Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. Proc Natl Acad Sci U S A 109:E2424–E2432
13. Gao X, Wan J, Liu B, Ma M, Shen B, Qian S (2014) Quantitative profiling of initiating ribosomes in vivo. Nat Methods 12(2):147–153
14. Ingolia NT, Brar GA, Rouskin S, McGeachy AM, Weissman JS (2012) The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. Nat Protoc 7:1534–1550
15. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105–1111