Methods in Molecular Biology 2252

Springer Protocols

Vyacheslav M. Labunskyy *Editor*

Ribosome Profiling

Methods and Protocols

Humana Press

METHODS IN MOLECULAR BIOLOGY

Series Editor
John M. Walker
School of Life and Medical Sciences
University of Hertfordshire
Hatfield, Hertfordshire, UK

For further volumes:
http://www.springer.com/series/7651

For over 35 years, biological scientists have come to rely on the research protocols and methodologies in the critically acclaimed *Methods in Molecular Biology* series. The series was the first to introduce the step-by-step protocols approach that has become the standard in all biomedical protocol publishing. Each protocol is provided in readily-reproducible step-by-step fashion, opening with an introductory overview, a list of the materials and reagents needed to complete the experiment, and followed by a detailed procedure that is supported with a helpful notes section offering tips and tricks of the trade as well as troubleshooting advice. These hallmark features were introduced by series editor Dr. John Walker and constitute the key ingredient in each and every volume of the *Methods in Molecular Biology* series. Tested and trusted, comprehensive and reliable, all protocols from the series are indexed in PubMed.

Ribosome Profiling

Methods and Protocols

Edited by

Vyacheslav M. Labunskyy

Department of Dermatology, Boston University School of Medicine, Boston, MA, USA

Humana Press

Editor
Vyacheslav M. Labunskyy
Department of Dermatology
Boston University School of Medicine
Boston, MA, USA

ISSN 1064-3745 ISSN 1940-6029 (electronic)
Methods in Molecular Biology
ISBN 978-1-0716-1149-4 ISBN 978-1-0716-1150-0 (eBook)
https://doi.org/10.1007/978-1-0716-1150-0

Cover Illustration Caption: Figure of ribosomes, with the binding site of the translational control protein studied in Chapter 5.

This Humana imprint is published by the registered company Springer Science+Business Media, LLC, part of Springer Nature.
The registered company address is: 1 New York Plaza, New York, NY 10004, U.S.A.

Preface

While many experimental techniques have been developed to assess gene expression, it remains challenging to quantify translation of proteins at the genome-wide level. Ribosome profiling has recently emerged as a powerful tool to monitor protein translation with high resolution in vivo, but methodological challenges and lack of standardized protocols are the major barriers to wide use of ribosome profiling-based approaches. This volume provides a single-source reference on the current state of the ribosome profiling method describing step-by-step experimental protocols for quantitative analysis of translation in a variety of model organisms. It also presents an overview of the existing software tools and includes detailed description of methods for statistical analysis, data processing, and visualization of ribosome profiling data.

Boston, MA, USA *Vyacheslav M. Labunskyy*

Contents

Contributors

Aleksandra S. Anisimova • *Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia; Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia; Max Perutz Labs, Medical University of Vienna, University of Vienna, Vienna, Austria*

Hanna Barlit • *Department of Dermatology, Boston University School of Medicine, Boston, MA, USA*

Alexander Bartholomäus • *GFZ German Research Centre for Geosciences, Section Geomicrobiology, Potsdam, Germany*

Carine Beaupere • *Department of Dermatology, Boston University School of Medicine, Boston, MA, USA*

Etienne Boileau • *Section of Bioinformatics and Systems Cardiology, Klaus Tschira Institute for Integrative Computational Cardiology, Heidelberg, Germany; Department of Internal Medicine III (Cardiology, Angiology, and Pneumology), University Hospital Heidelberg, Heidelberg, Germany; DZHK (German Centre for Cardiovascular Research) Partner Site Heidelberg/Mannheim, Heidelberg, Germany*

Gloria Ann Brar • *Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, CA, USA*

Lorenzo Calviello • *Max Delbruck Centre for Molecular Medicine in the Helmholtz Association, Berlin Institute for Medical Systems Biology, Berlin, Germany*

Florence Caro • *Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA, USA*

Xun Chen • *Broad Institute of MIT and Harvard, Cambridge, MA, USA*

Massimiliano Clamer • *IMMAGINA Biotechnology S.r.l., Povo, Italy*

Joseph L. DeRisi • *Department of Biochemistry and Biophysics, University of California, San Francisco, CA, USA*

Dion Dickman • *Department of Neurobiology, University of Southern California, Los Angeles, CA, USA*

Christoph Dieterich • *Section of Bioinformatics and Systems Cardiology, Klaus Tschira Institute for Integrative Computational Cardiology, Heidelberg, Germany; Department of Internal Medicine III (Cardiology, Angiology, and Pneumology), University Hospital Heidelberg, Heidelberg, Germany; DZHK (German Centre for Cardiovascular Research) Partner Site Heidelberg/Mannheim, Heidelberg, Germany*

Sergey E. Dmitriev • *Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia; Sirius University of Science and Technology, Sochi, Russia; Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia; Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia*

Artyom A. Egorov • *Phystech School of Biological and Medical Physics, Moscow Institute of Physics and Technology (State University), Dolgoprudny, Russia; Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia; Sirius University of Science and Technology, Sochi, Russia*

Irina A. Eliseeva • *Institute of Protein Research, Russian Academy of Sciences, Pushchino, Russia*

Maxim V. Gerashchenko • *Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA*
Wendy V. Gilbert • *Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA*
Vadim N. Gladyshev • *Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA*
Dermot Harnett • *IRI Life Sciences, Institute of Biology, Humboldt-Universität zu Berlin, Berlin, Germany*
Zoya Ignatova • *Department of Chemistry, Institute for Biochemistry and Molecular Biology, University of Hamburg, Hamburg, Germany*
Dorota Klepacki • *Center for Biomolecular Sciences, University of Illinois at Chicago, Chicago, IL, USA*
Ivan V. Kulakovskiy • *Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia; Institute of Protein Research, Russian Academy of Sciences, Pushchino, Russia; Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia; Sirius University of Science and Technology, Sochi, Russia; Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia*
Vyacheslav M. Labunskyy • *Department of Dermatology, Boston University School of Medicine, Boston, MA, USA*
Fabio Lauria • *Institute of Biophysics, CNR Unit at Trento, Povo, Italy*
Vsevolod J. Makeev • *Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia; Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia; Phystech School of Biological and Medical Physics, Moscow Institute of Physics and Technology (State University), Dolgoprudny, Russia*
Alexander S. Mankin • *Center for Biomolecular Sciences, University of Illinois at Chicago, Chicago, IL, USA*
Joel McManus • *Carnegie Mellon University, Pittsburgh, PA, USA*
Eelco Meerdink • *IRI Life Sciences, Institute of Biology, Humboldt-Universität zu Berlin, Berlin, Germany*
Sezen Meydan • *National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA*
Armaghan Naik • *Scientific & Digital Innovation, Sanofi Pasteur, Cambridge, MA, USA*
Eugene Oh • *Center for Cancer Research, Massachusetts General Hospital, Charlestown, MA, USA; Department of Medicine, Harvard Medical School, Boston, MA, USA*
Uwe Ohler • *IRI Life Sciences, Institute of Biology, Humboldt-Universität zu Berlin, Berlin, Germany; Max Delbruck Centre for Molecular Medicine in the Helmholtz Association, Berlin Institute for Medical Systems Biology, Berlin, Germany*
Emily Nicole Powers • *Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, CA, USA*
Shu-Bing Qian • *Division of Nutritional Sciences, Cornell University, Ithaca, NY, USA*
Jingfan Qiu • *Department of Pathogen Biology, Nanjing Medical University, Nanjing, China*
Manish K. Rai • *Department of Dermatology, Boston University School of Medicine, Boston, MA, USA*
Sara I. Shoushtari • *Department of Dermatology, Boston University School of Medicine, Boston, MA, USA*
Pieter Spealman • *Center for Genomics and Systems Biology, New York University, New York, NY, USA*

Dominique Sydow • *IRI Life Sciences, Institute of Biology, Humboldt-Universität zu Berlin, Berlin, Germany*
Toma Tebaldi • *Internal Medicine & Yale Cancer Center, Yale University School of Medicine, New Haven, CT, USA*
Nora Vázquez-Laslop • *Center for Biomolecular Sciences, University of Illinois at Chicago, Chicago, IL, USA*
Gabriella Viero • *Institute of Biophysics, CNR Unit at Trento, Povo, Italy*
Ilya E. Vorontsov • *Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia; Institute of Protein Research, Russian Academy of Sciences, Pushchino, Russia*
Yinuo J. Wang • *Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA; Arrakis Therapeutics, Waltham, MA, USA*

Part I

Experimental Protocols

Chapter 1

Monitoring Bacterial Translation Rates Genome-Wide

Eugene Oh

Abstract

Modern DNA sequencing technologies have allowed for the sequencing of tens of thousands of bacterial genomes. While this explosion of information has brought about new insights into the diversity of the prokaryotic world, much less is known of the identity of proteins encoded within these genomes, as well as their rates of production. The advent of ribosome profiling, or the deep sequencing of ribosome-protected footprints, has recently enabled the systematic evaluation of every protein-coding region in a given experimental condition, the rates of protein production for each gene, and the variability in translation rates across each message. Here, I provide an update to the bacterial ribosome profiling approach, with a particular emphasis on a simplified strategy to reduce cloning time.

Key words Ribosome profiling, Ribo-seq, Bacterial ribosome profiling, Bacterial translation

1 Introduction

Capturing translation rates can reveal which genes are being made, how much, and when they are needed [1, 2]. Recent advances in sequencing technologies have now made it possible to systematically assess translation rates under any experimental condition. Ribosome profiling, or the deep sequencing of ribosome-protected mRNA footprints, allows for the genome-wide monitoring of ongoing protein synthesis and complements other global approaches (such as RNA-seq and mass spectrometry-based proteomics) that enable exploration and discovery of diverse facets of gene regulation [3]. Indeed, ribosome profiling of various bacterial species has been used (i) to define unannotated open reading frames, especially those that encode for small polypeptides [4, 5]; (ii) to find translation initiation sites [6–9]; (iii) to investigate specific aspects of translation, including initiation [10] and frameshifting [11, 12]; (iv) to measure translation efficiencies and examine causes of their variation [13]; (v) to estimate stoichiometries of protein complexes [14]; (vi) to ascertain sites of ribosome pausing and their causes [15–17]; (vii) to determine codon usage patterns

Vyacheslav M. Labunskyy (ed.), *Ribosome Profiling: Methods and Protocols*, Methods in Molecular Biology, vol. 2252, https://doi.org/10.1007/978-1-0716-1150-0_1,

[18]; (viii) to characterize the mechanism of antibiotic action [19, 20]; and (ix) to identify stress-induced regulons [21–23].

While ribosome profiling was originally developed for the budding yeast [24], fundamental differences in bacterial translation demanded a substantial overhaul of the eukaryotic protocol [25, 26]. Firstly, bacteria can sense the immediate loss of nutrients by rapidly shutting down translation initiation, causing ribosomes to run off, which, in turn, leads to a localized paucity of ribosome occupancy at the 5′ end of messages. To minimize this effect, a rapid filtration strategy was developed for cell harvesting in lieu of traditional methods based on centrifugation. Secondly, the nuclease used to generate ribosome-protected footprints for eukaryotes, RNase I, had long been known to be inhibited by the 30S subunit of bacterial ribosomes [27]. This necessitated a new digestion strategy. While micrococcal nuclease was chosen for its ability to be quenched, use of other ribonucleases might prove beneficial [28]. Additionally, recent protocols have provided advancements that should be taken into consideration [29, 30]. However, it is worth noting that the cloning of mRNA footprints into a sequenceable library is time-consuming, taking up to 6 days of continuous work. Much of this time is spent on purifying each enzymatic step through lengthy gel extractions. Here, I describe a protocol that eliminates two gel purification steps, by substituting them with SPRI (solid phase reverse immobilization)-based purifications.

2 Materials

2.1 Cell Harvesting by Rapid Filtration

1. Growth media: LB broth (Lennox), LB broth (Miller), or MOPS minimal medium containing 0.2% glucose.
2. 15 mL culture tube.
3. 1 L baffled flask.
4. 90 mm filtration apparatus.
5. Vacuum line.
6. Nitrocellulose filter discs, 90 mm diameter, 0.2 μm pore size.
7. Scoopula spatula.
8. Liquid nitrogen.
9. 50 mL conical and 18 gauge needle.

2.2 Cell Pulverization

1. Styrofoam box.
2. Liquid nitrogen.
3. Mixer mill, 10 mL grinding jar, and 12 mm grinding ball (*see* **Note 1**).
4. 1 M tris(hydroxymethyl)aminomethane (Tris), pH 8 stock solution in nuclease-free water.

5. 1 M NH_4Cl stock solution in nuclease-free water.
6. 1 M $MgCl_2$ stock solution in nuclease-free water.
7. 10% Triton X-100 stock solution in nuclease-free water.
8. 10% Nonidet P-40 stock solution in nuclease-free water.
9. 100 mM chloramphenicol stock solution in 100% ethanol (store at −20 °C).
10. DNase I, RNase-free (*see* **Note 2**).
11. Lysis buffer: 20 mM Tris, pH 8, 100 mM NH_4Cl, 10 mM $MgCl_2$, 0.4% Triton X-100, 0.1% Nonidet P-40, 1 mM chloramphenicol, and 100 unit/mL DNase I, RNase-free in nuclease-free water.
12. 50 mL conical and 18 gauge needle.
13. Rounded-edge spatula.

2.3 mRNA Enrichment and Fragmentation

1. 1 M ethylenediaminetetraacetic acid (EDTA), pH 8 stock solution in nuclease-free water.
2. 3 M sodium acetate (NaOAc), pH 5.5 stock solution in nuclease-free water.
3. Resuspension buffer: 10 mM EDTA and 60 mM NaOAc in nuclease-free water.
4. MICROBExpress bacterial mRNA enrichment kit (*see* **Note 3**).
5. 10 mM Tris, pH 7 in nuclease-free water (prepare from a 1 M stock solution).
6. MEGAclear clean-up kit (*see* **Note 3**).
7. 2× alkaline hydrolysis buffer: 1 μL of 0.5 M EDTA, pH 8, 30 μL of 0.1 M Na_2CO_3, and 220 μL of 0.1 M $NaHCO_3$ (prepare fresh each time).
8. GlycoBlue coprecipitant (15 mg/mL stock) or equivalent carrier.
9. Isopropanol.
10. Microcentrifuge.
11. 80% ethanol in nuclease-free water (prepare from a 200 proof stock solution).

2.4 Extract Preparation, MNase Footprinting, and Monosome Isolation

1. Tabletop centrifuge compatible for spinning 50 mL conicals.
2. 1.5 mL tubes.
3. Microcentrifuge.
4. RNA/DNA spectrophotometer or Nanodrop.
5. Micrococcal nuclease (MNase) (*see* **Note 4**). Prepare a stock solution containing 250 units of MNase/μL in 10 mM Tris, pH 8 (store aliquots at −80 °C, avoid freeze-thaw cycles).

6. Superase•In RNase inhibitor (*see* **Note 5**).
7. 100 mM $CaCl_2$ stock solution in nuclease-free water.
8. 0.5 M ethylene glycol-bis(2-aminoethylether)-*N*,*N*,*N*,'*N*'--tetraacetic acid (EGTA), pH 8 stock solution in nuclease-free water.
9. SW 41 Ti swinging buckets, rotor, and ultracentrifuge.
10. Sucrose.
11. 1 M Tris, pH 8 stock solution in nuclease-free water.
12. 1 M NH_4Cl stock solution in nuclease-free water.
13. 1 M $MgCl_2$ stock solution in nuclease-free water.
14. 100 mM chloramphenicol stock solution in 100% ethanol (store at −20 °C).
15. 10% sucrose solution: 10% sucrose (w/v), 20 mM Tris, pH 8, 100 mM NH_4Cl, 10 mM $MgCl_2$, and 1 mM chloramphenicol in nuclease-free water.
16. 50% sucrose solution: 50% sucrose (w/v), 20 mM Tris, pH 8, 100 mM NH_4Cl, 10 mM $MgCl_2$, and 1 mM chloramphenicol in nuclease-free water.
17. SETON open-top polyclear tubes (*see* **Note 6**).
18. BIOCOMP Gradient Master and Piston Gradient Fractionator (*see* **Note 6**).
19. 2 mL screw cap tube.
20. Liquid nitrogen.

2.5 RNA Extraction

1. 20% sodium dodecyl sulfate (SDS) in nuclease-free water.
2. Acid phenol, pH 4.5.
3. Microcentrifuge.
4. 1.5 mL tubes.
5. Chloroform.
6. 3 M sodium acetate (NaOAc), pH 5.5.
7. Isopropanol.
8. 80% ethanol in nuclease-free water (prepare from a 200 proof stock solution).
9. 10 mM Tris, pH 7 in nuclease-free water (prepare from a 1 M stock solution).

2.6 Footprint Size Selection

1. RNA/DNA spectrophotometer or Nanodrop.
2. 10 mM Tris, pH 7 in nuclease-free water (prepare from a 1 M stock solution).
3. Agencourt AMPure XP beads (*see* **Note 7**).
4. Magnetic stand compatible for 1.5 mL tubes.

5. 1.5 mL tubes.
6. Isopropanol.
7. 80% ethanol in nuclease-free water (prepare from a 200 proof stock solution).

2.7 Dephosphorylation

1. NEB T4 polynucleotide kinase (PNK) (*see* **Note 7**).
2. 10× T4 PNK buffer (supplied with enzyme).
3. Superase•In RNase inhibitor.
4. 10 mM Tris, pH 7 in nuclease-free water (prepare from a 1 M stock solution).
5. 3 M sodium acetate (NaOAc), pH 5.5.
6. GlycoBlue coprecipitant (15 mg/mL stock).
7. Isopropanol.
8. Microcentrifuge.
9. 80% ethanol in nuclease-free water (prepare from a 200 proof stock solution).

2.8 Linker Ligation

1. NEB T4 RNA ligase 2, truncated K227Q (*see* **Note 7**).
2. 50% (w/v) PEG 8000 (supplied with enzyme).
3. 10× T4 RNA ligase reaction buffer (supplied with enzyme).
4. Superase•In RNase inhibitor.
5. 20 μM solution of oCJC88 oligonucleotide /5rApp/GATCG GAAGAGCACACGT/3ddC/ (*see* **Note 8**) dissolved in 10 mM Tris, pH 7.
6. Agencourt AMPure XP beads.
7. Isopropanol.
8. Magnetic stand compatible for 1.5 mL tubes.
9. 10 mM Tris, pH 7 in nuclease-free water (prepare from a 1 M stock solution).

2.9 Reverse Transcription

1. 10 mM dNTP stock solution.
2. 20 μM solution of oCJC160 oligonucleotide /5Phos/AGAT CGGAAGAGCGTCGTGTAGGGAAAGAGTGT/iSp18/GT GACTGGAGTTCAGACGTGTGCTCTTCCGATC (*see* **Note 8**) dissolved in 10 mM Tris, pH 7.
3. SuperScript III reverse transcriptase (*see* **Note 7**).
4. 5× First strand buffer (FSB) (supplied with enzyme).
5. 0.1 M DTT (supplied with enzyme).
6. Superase•In RNase inhibitor.
7. 1 N NaOH.
8. 2× Novex TBE-urea sample buffer (*see* **Note 7**).

9. 10% Novex TBE-urea gel, 12 wells (*see* **Note 7**).
10. 1× Tris-borate-EDTA (TBE) running buffer. Contains 0.089 M Tris-borate and 0.002 M EDTA buffered to pH 8.3.
11. SYBR gold nucleic acid gel stain (*see* **Note 7**).
12. 18 gauge needle, 0.5 mL tubes, and 1.5 mL tubes.
13. Microcentrifuge.
14. 10 mM Tris, pH 8.
15. Costar Spin-X centrifuge filters (*see* **Note 7**).
16. 5 M NaCl.
17. 0.5 M EDTA, pH 8.
18. GlycoBlue coprecipitant (15 mg/mL stock).
19. Isopropanol.
20. 80% ethanol in nuclease-free water (prepare from a 200 proof stock solution).

2.10 Circularization

1. CircLigase ssDNA ligase (*see* **Note 7**).
2. 10× CircLigase reaction buffer (supplied with enzyme).
3. 1 mM ATP (supplied with enzyme).
4. 50 mM $MnCl_2$ (supplied with enzyme).

2.11 rRNA Subtraction

1. 100 μM solution of hybridization oligonucleotides dissolved in 10 mM Tris, pH 8 (*see* **Note 8**):
   mix 77 μL of 100 μM oEO1055, /5Biosg/TCATCTCCGG GGGTAGAGCACTGTTTCG;
   4 μL of 100 μM oEO1056, /5Biosg/GGCTAAACCATG CACCGAAGCTGCGGCAG;
   17 μL of 100 μM oEO1057, /5Biosg/AAGGCTGAGGC GTGATGACGAGGCACT;
   and 2 μL of 100 μM oEO1058, /5Biosg/CGGTGCTGA AGCAACAAATGCCCTGCTT.
2. 20× saline-sodium citrate (SSC). Contains 3 M NaCl and 0.4 M sodium citrate buffered to pH 7 with HCl.
3. PCR thermocycler.
4. MyOne streptavidin C1 dynabeads (*see* **Note 7**).
5. Magnetic stand compatible for 1.5 mL tubes.
6. 1× binding and wash (B&W) buffer: 5 mM Tris, pH 7.5, 0.5 mM EDTA, 1 M NaCl, and 0.01% Tween.
7. 2× B&W buffer: 10 mM Tris, pH 7.5, 1 mM EDTA, 2 M NaCl, and 0.01% Tween.
8. 10 mM Tris, pH 8.
9. 5 M NaCl.

10. 0.5 M EDTA, pH 8.
11. GlycoBlue coprecipitant (15 mg/mL stock).
12. Isopropanol.
13. Microcentrifuge.
14. 80% ethanol in nuclease-free water (prepare from a 200 proof stock solution).

2.12 Library Amplification by PCR

1. High-Fidelity (HF) Phusion DNA polymerase (*see* **Note 7**).
2. 5× HF buffer (supplied with enzyme).
3. 10 mM dNTPs.
4. 10 μM solution of oCJC161 oligonucleotide resuspended in 10 mM Tris, pH 8:
   AATGATACGGCGACCACCGAGATCTACACTCTTTCCC TACACGACGCTCTTCCGATCT.
5. 10 μM solution of oCJC60–71 indexing oligonucleotides resuspended in 10 mM Tris, pH 8:
   oCJC60: CAAGCAGAAGACGGCATACGAGATCGAGTAA TGTGACTGGAGTTCAGACG, index: ATTACTCG.
   oCJC61: CAAGCAGAAGACGGCATACGAGATTCTCCG GAGTGACTGGAGTTCAGACG, index: TCCGGAGA.
   oCJC62: CAAGCAGAAGACGGCATACGAGATAATGAGC GGTGACTGGAGTTCAGACG, index: CGCTCATT.
   oCJC63: CAAGCAGAAGACGGCATACGAGATGGAATCT CGTGACTGGAGTTCAGACG, index: GAGATTCC.
   oCJC64: CAAGCAGAAGACGGCATACGAGATTTCTGAA TGTGACTGGAGTTCAGACG, index: ATTCAGAA.
   oCJC65: CAAGCAGAAGACGGCATACGAGATACGAATT CGTGACTGGAGTTCAGACG, index: GAATTCGT.
   oCJC66: CAAGCAGAAGACGGCATACGAGATAGCTT CAGGTGACTGGAGTTCAGACG, index: CTGAAGCT.
   oCJC67: CAAGCAGAAGACGGCATACGAGATGCGCAT TAGTGACTGGAGTTCAGACG, index: TAATGCGC.
   oCJC68: CAAGCAGAAGACGGCATACGAGATCATAGCC GGTGACTGGAGTTCAGACG, index: CGGCTATG.
   oCJC69: CAAGCAGAAGACGGCATACGAGATTTCGCG GAGTGACTGGAGTTCAGACG, index: TCCGCGAA.
   oCJC70: CAAGCAGAAGACGGCATACGAGATGCGCGA GAGTGACTGGAGTTCAGACG, index: TCTCGCGC.
   oCJC71: CAAGCAGAAGACGGCATACGAGATCTATCGC TGTGACTGGAGTTCAGACG, index: AGCGATAG).
6. 6× DNA gel loading dye.

7. Novex 8% TBE-polyacrylamide gel, 12 wells.
8. 1× TBE running buffer.
9. SYBR gold nucleic acid gel stain.
10. 18 gauge needle, 0.5 mL tubes, and 1.5 mL tubes.
11. Microcentrifuge.
12. 10 mM Tris, pH 8.
13. Costar Spin-X centrifuge filters.
14. 5 M NaCl.
15. 0.5 M EDTA, pH 8.
16. GlycoBlue coprecipitant (15 mg/mL).
17. Isopropanol.
18. 80% ethanol in nuclease-free water (prepare from a 200 proof stock solution).

2.13 Quantify, Sequence, and Analyze

1. Fragment analyzer or equivalent.
2. Read1 oligonucleotide sequence: ACACTCTTTCCCTACAC GACGCTCTTCCGATCT.
3. Indexing oligonucleotide sequence: GATCGGAAGAGCA CACGTCTGAACTCCAGTCAC.

3 Methods

3.1 Cell Harvesting by Rapid Filtration (~10–15 min for Harvesting)

1. Inoculate a single bacterial colony in 5 mL of growth medium. Use the appropriate growth medium for your strain of interest and/or experimental condition. Use of non-rich media has been shown to deplete specific amino acids more readily and will affect ribosome occupancy measurements. Culture cells overnight at 37 °C (or at a growth temperature required for your strain of interest) in a 15 mL culture tube.
2. Dilute overnight culture in 250 mL of growth medium (pre-warmed to 37 °C) in a 1 L baffled flask. Make sure cells go through more than five doublings. Starting OD_{600} should be less than 0.005.
3. Grow cells at 37 °C until culture reaches log-phase growth (or an OD_{600} of ~0.3–0.4). Culture conditions for log-phase growth should be optimized for each strain.
4. Connect filtration apparatus to a vacuum line in a 37 °C warm room. Make sure to use a 0.2 μm filter disc made of nitrocellulose. Filter discs made of different materials and pore sizes have been tested and do not filter as well. Cultures should be filtered at their growth temperature, if possible. Use a portable vacuum pump if a vacuum line is not set up in your warm room.

5. Filter liquid culture (*see* **Note 9**). Do not wait for liquid culture to filter fully. Immediately disassemble filtration apparatus once ~90–95% of culture has passed through the filter. This will minimize ribosome runoff, which occurs as soon as cells are deprived of growth medium.
6. Firmly scrape cells off filter disc with a Scoopula spatula. These spatulas are preferred because they have a long edge. Avoid scraping cells more than once, as this will help minimize ribosome runoff.
7. Quickly plunge Scoopula spatula into a 50 mL conical filled with liquid nitrogen. A second spatula (prechilled in liquid nitrogen) may be needed to dislodge cells that are firmly adhered to the Scoopula.
8. Remove excess liquid nitrogen. Pierce a 50 mL conical cap five times with an 18 gauge needle. Cap conical and invert to remove excess liquid nitrogen. Make sure vents face away from you.
9. Store frozen cells at −80 °C [STOP POINT] or continue to next section.

3.2 Cell Pulverization (~1 h)

1. Fill a Styrofoam box with liquid nitrogen. You may need to refill as liquid nitrogen evaporates.
2. Place a 12 mm grinding ball in a 10 mL grinding jar and submerge in a liquid nitrogen bath until liquid nitrogen stops boiling.
3. Prepare 1 mL of lysis buffer.
4. Fill a 50 mL conical with liquid nitrogen and add 650 μL of lysis buffer dropwise.
5. Remove excess liquid nitrogen. Pierce a 50 mL conical cap five times with an 18 gauge needle. Cap conical and invert to remove excess liquid nitrogen. Make sure vents face away from you.
6. Remove grinding jar (grinding ball included) from liquid nitrogen bath. Place frozen cells (from Subheading 3.1, **step 9**) and frozen lysis buffer pellets in grinding jar. Make sure liquid nitrogen has fully evaporated prior to assembling grinding jar.
7. Submerge closed jar in a liquid nitrogen bath until liquid nitrogen stops boiling.
8. Pulverize cells in a mixer mill at 15 Hz for 3 min.
9. Submerge jar in a liquid nitrogen bath until liquid nitrogen stops boiling.
10. Repeat **steps 8** and **9** four more times.
11. Transfer pulverized cell powder into a 50 mL conical filled with liquid nitrogen. Fill conical with liquid nitrogen up to the

25 mL line. Do not overfill. Use a rounded-edge spatula (pre-chilled in liquid nitrogen) to scrape out the pulverized cell powder and gently deposit cell powder into liquid nitrogen.

12. Remove excess liquid nitrogen. Pierce a 50 mL conical cap five times with an 18 gauge needle. Cap and tap the conical until excess liquid nitrogen escapes through vents. Do not invert or you will lose your pulverized cell powder. Make sure vents face away from you.
13. Store pulverized cell powder at −80 °C [STOP POINT], or continue to next section. If storing cell powder, store with a vented cap. If genome-wide translation efficiency measurements are desired, save one-third of pulverized cell powder for Subheading 3.3 and two-thirds of cell powder for Subheading 3.4 (*see* **Note 10**). If only ribosome density measurements are needed, proceed with Subheading 3.4.

3.3 mRNA Enrichment (~4–5 h) and Fragmentation (~1.5 h)

1. For total RNA extraction, dissolve pulverized cell powder in 0.7 mL of resuspension buffer.
2. Proceed with Subheading 3.5, **steps 1–18** and return.
3. Remove 16S and 23S ribosomal RNAs by subtractive hybridization using the MICROBExpress bacterial mRNA enrichment kit (follow manufacturer's instructions). Begin with 20 μg of total RNA. Isopropanol precipitate rRNA-subtracted RNAs. Resuspend in 40 μL of 10 mM Tris, pH 7.
4. Remove small RNAs less than 100 nucleotides using the MEGAclear clean-up kit (follow manufacturer's instructions). Isopropanol precipitate enriched mRNAs. Resuspend in 25 μL of 10 mM Tris, pH 7.
5. Fragment enriched mRNA by alkaline hydrolysis. To 25 μL of enriched mRNA, add 25 μL of freshly prepared 2× alkaline hydrolysis buffer.
6. Incubate at 95 °C for 23 min and transfer to ice.
7. Precipitate fragmented RNA by adding 450 μL of 10 mM Tris, pH 7, 55 μL of 3 M NaOAc pH, 5.5, 2 μL of GlycoBlue coprecipitant, and 0.55 mL of 100% isopropanol. Vortex and incubate at −80 °C for 30 min or longer.
8. Pellet at 20,000 × *g* or top speed in a microcentrifuge for 30 min.
9. Wash pellet twice with 0.5 mL of ice-cold 80% ethanol.
10. Dry pellet for 5 min in a chemical fume hood.
11. Resuspend pellet in 50 μL of 10 mM Tris, pH 7.
12. Proceed with Subheading 3.6, **step 3**.

3.4 Extract Preparation (~30 min), MNase Footprinting (~1.5 h), and Monosome Isolation (~3.5 h)

1. Thaw pulverized cell powder in a room temperature water bath for 2 min (or until completely thawed).
2. Incubate on ice for 10 min.
3. Spin down 50 mL conical at 4000 × *g* in a tabletop centrifuge for 1 min at 4 °C. This is to collect the extract adhering to the conical wall.
4. Transfer thawed extract to a prechilled 1.5 mL tube.
5. Spin down insoluble debris at 20,000 × *g* or top speed in a microcentrifuge for 10 min at 4 °C.
6. Transfer clarified supernatant to a prechilled 1.5 mL tube. Do not disrupt pellet.
7. Measure RNA concentration by Nanodrop by diluting extract 1:100 in nuclease-free water. Measure A_{260} and calculate concentration. 1 A_{260} unit equals 40 μg/mL of total RNA.
8. Digest 0.5 mg of total RNA in a 0.2 mL reaction volume. To 0.5 mg of total RNA, add 750 units of MNase, 2.5 μL of Superase•In, and 10 μL of 100 mM $CaCl_2$. Bring up reaction volume to 0.2 mL with lysis buffer (*see* recipe in Subheading 2.2).
9. Incubate reaction at 25 °C for 1 h.
10. Quench reaction with 2 μL of 0.5 M EGTA and leave on ice.
11. Prechill SW 41 Ti swinging buckets and rotor at 4 °C.
12. Prepare 15 mL of 10% sucrose solution.
13. Prepare 15 mL of 50% sucrose solution.
14. Add 6 mL of 50% sucrose solution to an open-top polyclear tube. SETON brand tubes are less prone to cracking.
15. Layer 6 mL of 10% sucrose solution with a motorized pipette controller. Use the slowest setting available to avoid mixing the two layers.
16. Make a 10–50% gradient. Use the preset 10–50% (w/v) short program on a Gradient Master.
17. Carefully load gradients into prechilled buckets. Gradients should be stored at 4 °C until quenched reactions are ready to be loaded.
18. Load samples without disrupting the gradients. Balance with lysis buffer, if necessary (*see* recipe in Subheading 2.2).
19. Carefully attach the loaded buckets to the prechilled rotor. Set ultracentrifuge to spin an SW 41 Ti rotor and spin at 35,000 rpm (151,000 × *g*) for 2.5 h at 4 °C.
20. Fractionate using a Piston Gradient Fractionator. Set piston speed to 0.2 mm/s.

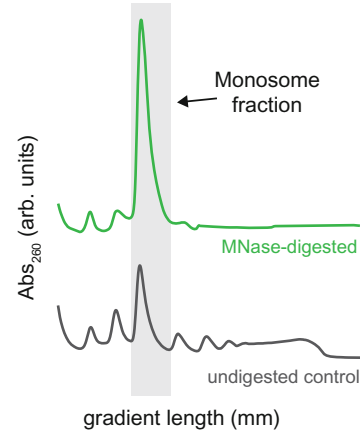

Fig. 1 Sucrose gradient profiles of mock versus MNase-digested polysomes. Polysomes were harvested from DH5α cells and run on a 10–50% sucrose gradient. Gradient lengths (mm) are plotted as a function of absorbance at 260 nm (in arbitrary units)

21. Manually collect monosome peak in a 2 mL screw cap tube (Fig. 1).
22. Flash freeze monosomes in liquid nitrogen and store at −80 °C [STOP POINT] or continue to next section.

3.5 RNA Extraction (~2 h)

1. To 0.7 mL of monosomes, add 40 μL of 20% SDS and 0.7 mL of acid phenol, pH 4.5 (prewarmed to 65 °C).
2. Vortex and incubate at 65 °C for 5 min.
3. Chill on ice for 5 min.
4. Spin at 20,000 × *g* or top speed in a microcentrifuge for 2 min.
5. Transfer aqueous layer to a new 1.5 mL tube.
6. Add 0.7 mL of room temperature acid phenol, pH 4.5.
7. Vortex and incubate at room temperature for 5 min.
8. Spin at 20,000 × *g* or top speed for 2 min.
9. Transfer aqueous layer to a new 1.5 mL tube.
10. Add 0.6 mL of chloroform.
11. Vortex and immediately spin at 20,000 × *g* or top speed for 1 min.
12. Transfer aqueous layer to new 1.5 mL tube.
13. Precipitate total RNA by adding 78 μL of 3 M NaOAc, pH 5.5, and 0.77 mL of 100% isopropanol. Vortex and incubate at −80 °C for 30 min or longer.
14. Pellet at 20,000 × *g* or top speed for 30 min.
15. Wash pellet twice with 0.5 mL of ice-cold 80% ethanol.

16. Dry pellet for 5 min in a chemical fume hood.
17. Resuspend pellet in 20 μL of 10 mM Tris, pH 7.
18. Store extracted RNA at −20 °C [STOP POINT] or continue to next section.

3.6 Footprint Size Selection (~1.5 h)

1. Quantify extracted RNA by Nanodrop.
2. Dilute 10 μg of extracted RNA in 50 μL of 10 mM Tris, pH 7.
3. Add 90 μL of AMPure XP beads (or SPRIselect beads). This method is based on a strategy for cloning microRNA libraries [31]. Make sure AMPure XP beads are fully resuspended prior to use. AMPure XP and SPRIselect beads are identical except SPRIselect beads are certified as RNAse-free. Nonetheless, AMPure XP beads have been used to purify RNA without observable degradation.
4. Mix by pipetting until beads are fully resuspended and incubate at room temperature for 5 min.
5. Place tube in a magnetic stand, and allow beads to settle until supernatant clears. RNA fragments greater than ~200 nucleotides will adhere to beads.
6. Carefully transfer supernatant to a new tube. This fraction contains RNA fragments that range in length from ~18 to 200 nucleotides.
7. Add 70 μL of AMPure XP beads and 90 μL of 100% isopropanol. Final concentrations: 10.6% PEG, 30% isopropanol.
8. Mix by pipetting until beads are fully resuspended and incubate at room temperature for 5 min.
9. Place tube in a magnetic stand, and allow beads to settle until supernatant clears.
10. Carefully transfer supernatant to a new tube.
11. Add 90 μL of AMPure XP beads and 270 μL of 100% isopropanol. Final concentrations: 7.5% PEG, 58% isopropanol.
12. Mix by pipetting until beads are fully resuspended, and incubate at room temperature for 5 min. RNA fragments ranging from ~18 to 100 nucleotides will adhere to beads (Fig. 2). RNA fragments outside the desired range of ~18–45 nucleotides can be removed by gel extraction following the reverse transcription reaction (*see* Subheading 3.9).
13. Place tube in a magnetic stand, and allow beads to settle until supernatant clears.
14. Discard supernatant.
15. Wash twice with 0.5 mL of room temperature 80% ethanol. Do not disrupt beads.
16. Dry beads for 5 min in a chemical fume hood.

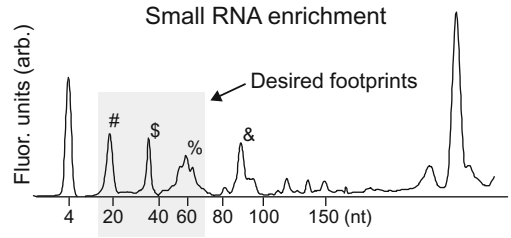

Fig. 2 RNA bioanalyzer trace of SPRI-selected small RNA footprints. RNA was quantified in fluorescence units (*y*-axis). #, $, %, and & signs reflect discrete, contaminating rRNA bands (at indicated sizes), which are used as internal sizing standards. The desired footprint lengths are enclosed within the gray box

17. Resuspend beads with 30 μL of 10 mM Tris, pH 7.
18. Mix by pipetting until beads are fully resuspended and incubate at room temperature for 2 min.
19. Place tube in a magnetic stand and allow beads to settle until supernatant clears.
20. Transfer eluate into a new 1.5 mL tube and store at −20 °C [STOP POINT] or continue to next section.

3.7 Dephosphorylation (~2.5 h)

1. To 15 μL of eluate, add 2 μL of 10× T4 PNK buffer (supplied with enzyme) and 1 μL of Superase•In. Make a master mix if performing many reactions. To monitor cloning efficiency, you can perform the same reaction with the control oligo oNTI199 in parallel (*see* **Note 11**).
2. Add 2 μL of T4 PNK.
3. Mix by pipetting.
4. Incubate reaction at 37 °C for 1 h.
5. Heat kill enzyme at 75 °C for 10 min.
6. Precipitate by adding 480 μL of 10 mM Tris, pH 7, 55 μL of 3 M NaOAc, pH 5.5, 2 μL of GlycoBlue coprecipitant, and 0.55 mL of 100% isopropanol. Vortex and incubate at −80 °C for 30 min or longer.
7. Pellet at 20,000 × *g* or at top speed in a microcentrifuge for 30 min at 4 °C.
8. Wash pellet twice with 0.5 mL of ice-cold 80% ethanol.
9. Dry pellet for 5 min in a chemical fume hood.
10. Resuspend pellet in 7 μL of 10 mM Tris, pH 7.
11. Store at −20 °C [STOP POINT] or continue to next section.

3.8 Linker Ligation (~3 h)

1. Denature 7 μL of dephosphorylated RNA at 80 °C for 2 min and return to ice.
2. Add 8 μL of 50% (w/v) PEG 8000 (supplied with enzyme), 2 μL of 10× T4 RNA ligase reaction buffer (supplied with

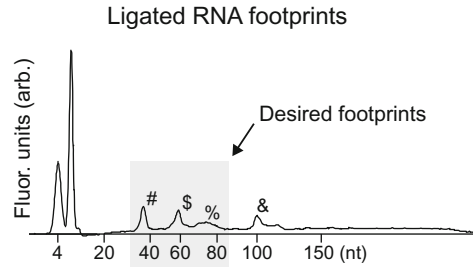

Fig. 3 Small RNA bioanalyzer trace of RNA footprints ligated with linker. RNA was quantified in fluorescence units (*y*-axis). #, $, %, and & signs reflect discrete, contaminating rRNA bands (at indicated sizes), which are used as internal sizing standards. RNA footprints will increase by ~18 nt, which is the length of the linker. The desired footprint lengths are enclosed within the gray box

enzyme), 1 μL of Superase•In, and 1 μL of 20 μM oCJC88 oligonucleotide. Make a master mix if performing many reactions.

3. Add 1 μL of T4 RNA ligase 2, truncated K227Q.
4. Mix by pipetting.
5. Incubate at 25 °C for 2 h.
6. To 20 μL of ligation reaction, add 29 μL of nuclease-free water, 75 μL of AMPure XP beads, and 129 μL of 100% isopropanol. Final concentrations: 7.5% PEG, 51% isopropanol.
7. Mix until beads are fully resuspended by pipetting and incubate at room temperature for 5 min. RNA fragments greater than ~35 nucleotides will adhere to beads (Fig. 3).
8. Discard supernatant.
9. Resuspend beads in 11.5 μL of 10 mM Tris, pH 7.
10. Place tube in a magnetic stand and allow beads to settle until supernatant clears.
11. Transfer eluate into a fresh 1.5 mL tube and store at −20 °C [STOP POINT] or continue to next section.

3.9 Reverse Transcription (~3.5 h)

1. To 11.5 μL of ligated RNA, add 1 μL of 10 mM dNTP and 1 μL of 20 μM oCJC160 oligonucleotide.
2. Denature at 65 °C for 5 min and return to ice.
3. Add 4 μL of 5× FSB buffer (supplied with enzyme), 1 μL of 0.1 M DTT (supplied with enzyme), and 1 μL of Superase•In. Make a master mix if performing many reactions.
4. Add 1 μL of SuperScript III reverse transcriptase.
5. Mix by pipetting.
6. Incubate at 50 °C for 30 min.
7. Add 2.3 μL of 1 N NaOH.

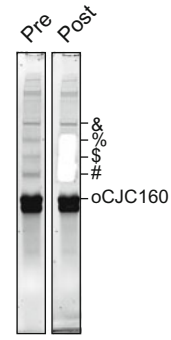

Fig. 4 Gel analysis of reverse transcription reaction. Reverse-transcribed RNA was resolved on a 15% TBE-urea gel (however, a 10% TBE-urea is recommended). Size selection of your desired fragment lengths should occur at this step. #, $, %, and & signs reflect discrete, contaminating rRNA bands, which are used as internal sizing standards, rather than DNA ladders. Left panel indicates gel before excision of bands (pre), while right panel indicates gel following excision of bands (post)

8. Hydrolyze RNA at 95 °C for 15 min.
9. Add 23 μL of 2× TBE-urea sample buffer.
10. Denature at 80 °C for 2 min and return to ice.
11. Set up a 10% TBE-urea gel in 1× TBE. Pre-run gel at 200 V for 1 h. Wash lanes prior to use.
12. Load samples and run at 200 V until loading dye runs out of gel. Instead of relying on loading controls, I now use the contaminating rRNA bands to approximate sizes (*see* **Note 12**).
13. Stain gel with a 1:10,000 dilution of SYBR gold reagent in 1× TBE for 2 min.
14. Size select desired bands. Cut below the lowest contaminating rRNA band (#) but well above the free oCJC160 oligonucleotide to below the highest contaminating rRNA band (&) (Fig. 4) (*see* **Note 13**). To maximize the recovery of all ribosome-protected mRNA footprints, cutting a larger band length is preferred. Use blue light source for detection, if possible. Avoid using UV as light source to prevent UV-induced crosslinking of library.
15. To recover DNA, pierce an 18 gauge needle through a 0.5 mL tube. Insert gel slice in pierced tube and nest inside a 1.5 mL tube.
16. Spin down at 20,000 × *g* or top speed in a microcentrifuge for 2 min to crush gel.
17. Add 0.5 mL of 10 mM Tris, pH 8. Incubate at −20 °C for 30 min.

18. Incubate at 70 °C for 10 min with shaking. Transfer gel slurry to a Spin-X cellulose acetate column using a wide-bore pipette (or cut a pipette tip with a razor blade).
19. Spin at 20,000 × *g* or top speed for 1 min. Transfer flow through to a new 1.5 mL tube.
20. Add 32 μL of 5 M NaCl, 1 μL of 0.5 M EDTA, 2 μL of GlycoBlue coprecipitant, and 0.55 mL of 100% isopropanol. Mix by vortexing and incubate at −80 °C for 30 min or longer.
21. Pellet at 20,000 × *g* or top speed for 30 min at 4 °C.
22. Wash pellet twice with 0.5 mL of ice-cold 80% ethanol.
23. Dry pellet for 5 min in a chemical fume hood.
24. Resuspend pellet in 15 μL of 10 mM Tris, pH 8.
25. Store at −20 °C [STOP POINT] or continue to next section.

3.10 Circularization (~2.5 h)

1. To 15 μL of reverse-transcribed ssDNA, add 2 μL of 10× CircLigase buffer (supplied with enzyme), 1 μL of 1 mM ATP (supplied with enzyme), and 1 μL of 50 mM $MnCl_2$ (supplied with enzyme). Make a master mix if performing many reactions.
2. Add 1 μL of CircLigase.
3. Mix by pipetting.
4. Incubate at 60 °C for 1 h. Use a heated lid, if possible.
5. Dope in 1 μL of CircLigase and incubate at 60 °C for 1 h.
6. Heat kill enzyme at 80 °C for 10 min.
7. Store reaction at −20 °C [STOP POINT] or continue to next section.

3.11 rRNA Subtraction (~3 h)

1. To 5 μL of circularized DNA, add 1 μL of 100 μM mix of hybridization oligonucleotides, 1 μL of 20× SSC, and 3 μL of water. Make a master mix if performing many reactions. Hybridization oligonucleotide sequences were designed for *E. coli* ribosomal RNAs and must be empirically determined for other bacterial species.
2. Incubate at 98 °C for 2 min.
3. Ramp down temperature to 37 °C over a span of 1 h in a PCR thermocycler.
4. Incubate at 37 °C for 20 min.
5. Prepare 25 μL of MyOne streptavidin C1 dynabeads using magnetic stand.
6. Wash dynabeads three times with 25 μL of 1× B&W buffer.
7. Resuspend in 10 μL of 2× B&W buffer.
8. Incubate dynabeads at 37 °C until needed.

9. Add 10 μL of hybridization reaction to pre-washed dynabeads.
10. Incubate at 37 °C for 15 min.
11. Place in a magnetic stand and recover supernatant.
12. Precipitate DNA by adding 0.48 mL of 10 mM Tris, pH 8, 32 μL of 5 M NaCl, 1 μL of 0.5 M EDTA, 2 μL of GlycoBlue coprecipitant, and 0.55 mL of 100% isopropanol. Mix by vortexing and incubate at −80 °C for 30 min or longer.
13. Pellet at 20,000 × *g* or top speed in a microcentrifuge for 30 min at 4 °C.
14. Wash pellet twice with 0.5 mL of ice-cold 80% ethanol.
15. Dry pellet for 5 min in a chemical fume hood.
16. Resuspend in 5 μL of 10 mM Tris, pH 8.
17. Store at −20 °C [STOP POINT] or continue to next section.

3.12 Library Amplification by PCR (~2.5 h)

1. Prepare a master mix containing 16.7 μL of 5× HF buffer (supplied with enzyme), 1.7 μL of 10 mM dNTPs, 4 μL of 10 μM oCJC161, 4 μL of 10 μM indexing oligonucleotide, 52 μL of water, and 0.8 μL of HF Phusion polymerase.
2. Add 79.2 μL of PCR master mix to 5 μL of rRNA-subtracted circularized DNA and vortex.
3. Aliquot 16 μL into five separate PCR tubes.
4. Perform PCR reaction. Set initial denaturation at 98 °C for 30 s. Cycle 14 times using following conditions: 98 °C for 10 s, 60 °C for 10 s, 72 °C for 5 s.
5. Remove a tube after cycle 6, 8, 10, 12 and 14.
6. Add 3.5 μL of 6× DNA gel loading dye.
7. Set up an 8% TBE-polyacrylamide gel in 1× TBE.
8. Load samples and run at 180 volts for 40 min.
9. Stain gel with a 1:10,000 dilution of SYBR gold reagent in 1× TBE for 2 min.
10. Excise desired bands (Fig. 5).
11. To precipitate library, pierce an 18 gauge needle through a 0.5 mL tube. Insert gel slice in pierced tube and nest inside a 1.5 mL tube.
12. Spin down at 20,000 × g or top speed in a microcentrifuge for 2 min to crush gel.
13. Add 0.5 mL of 10 mM Tris, pH 8. Incubate at −20 °C for 30 min.
14. Incubate at 70 °C for 10 min with shaking. Transfer gel slurry to a Spin-X cellulose acetate column using a wide-bore pipette.

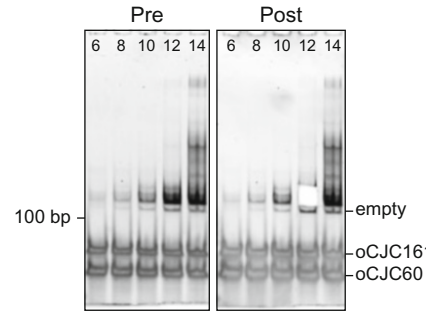

Fig. 5 Gel analysis of PCR amplification reaction. Circularized ssDNA was directly amplified without rRNA subtraction. PCR reactions were removed after the indicated cycles and resolved on an 8% TBE-polyacrylamide gel. Left panel indicates gel before excision of bands (pre), while right panel indicates gel following excision of bands (post)

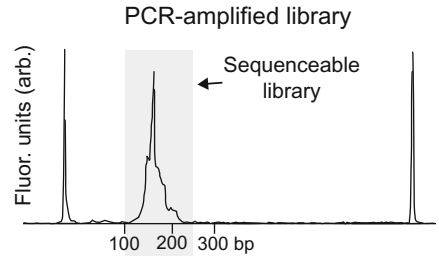

Fig. 6 High sensitivity small DNA fragment analyzer trace of amplified DNA library

15. Spin at 20,000 × *g* or top speed for 1 min. Transfer flow through to a fresh tube.
16. Add 32 μL of 5 M NaCl, 1 μL of 0.5 M EDTA, 2 μL of GlycoBlue coprecipitant, and 0.55 mL of 100% isopropanol. Mix by vortexing and incubate at −80 °C for 30 min or longer.
17. Pellet at 20,000 × *g* or top speed for 30 min at 4 °C.
18. Wash pellet twice with 0.5 mL of ice-cold 80% ethanol.
19. Dry pellet for 5 min in a chemical fume hood.
20. Resuspend pellet in 10 μL of 10 mM Tris, pH 8.
21. Store library at −20 °C [STOP POINT] or continue to next section.

3.13 Quantify, Sequence, and Alignment

1. Quantify library using a fragment analyzer (Fig. 6). The library should be between ~140 and 180 bp in length. Concentrations will vary based on cycling times and gel extraction efficiencies but should be greater than 2 nM. If more is needed, repeat PCR with additional cycling times or use more rRNA-subtracted circularized DNA as input.

2. Sequence using an Illumina sequencing platform.
3. Trim reads with cutadapt version 1.16.

```
cutadapt -m 15 -u 1 -a GATCGGAAGAGCACACGT -o $output_name
$input_fastq_file
```

4. Remove rRNA sequences and align with bowtie v.1.2.2. Use aligned reads for quantifying translation rates.

```
bowtie rRNA_sequences -v 2 -m 1 $input_fastq_file >
$output_aligned2rRNA -un %output_unaligned
```

```
bowtie reference_genome -v 2 -m 1 $output_unaligned >
$output_aligned2genome
```

5. Quantifying translation rates. Translation rates can be measured using various metrics [29, 30]. Studies on ribosome pausing (i.e., local translation) require a more refined analysis [16] when compared with quantifying bulk (or gene level) translation rates. Briefly, to measure gene level translation rates, the sum of aligned reads to an open reading frame (ORF) should be normalized to the length of the ORF and the total number of reads aligned to all ORFs. This gives the units reads per kilobase million (RPKM) for each gene. To account for variability in read length (~15–45 nucleotides), footprints can be trimmed from each side by a constant length (e.g., 7 nucleotides), with the remaining nucleotides given a score of 1/N (where N is the number of positions leftover after discarding each end).
6. Quality control metrics. To assess the extent of ribosome run-off, perform a metagene analysis. Generate a ribosome density profile for each ORF and scale each position by the mean density for that ORF. Exclude genes with less than 128 aligned reads. This gives differentially expressed genes equal weighting. Align each normalized ribosome density profile from the start codon and average across each position. If the average ribosome density is lower near the beginning of the message compared to the middle, then there is ribosome runoff. Nutrient deprivation and delays in cell harvesting are major causes of this. To overcome these issues, cells should be harvested during early-log to log-phase growth, and cell harvesting must be performed as quickly as possible.

4 Notes

1. Cell pulverization has been optimized using the Retsch MM301 mixer mill. If a different system is used, cell pulverization should be reoptimized such that cells are fully lysed while polyribosome complexes remain mostly intact.
2. This protocol uses DNase I, RNase-free (Roche, catalog no. 04716728001). It is essential to use RNase-free DNase I to preserve the integrity of polyribosome complexes. While RNase-free DNase I from other sources may be suitable for use, this may require further optimization, as each company uses non-standardized unit definitions for enzymatic activity.
3. For MICROBExpress bacterial mRNA enrichment and MEGAclear clean-up kits, follow manufacturer's instructions. Perform standard isopropanol precipitations to concentrate.
4. This protocol utilizes MNase from Sigma, catalog no. 10107921001. The source of MNase has been specifically optimized for this protocol. I have found that MNase from different sources show drastically differing activities (unpublished results).
5. Superase•In is a propriety RNase inhibitor that has been specifically validated to inhibit RNase I, RNase A, but not MNase (unpublished results).
6. While other gradient makers and fraction collectors can be employed, this protocol describes sucrose gradients made and collected using the BIOCOMP Gradient Master and Piston Gradient Fractionator. It is essential to use the open-top polyclear tubes from SETON when using the BIOCOMP Piston Gradient Fractionator, as the gaskets seal poorly with other brands.
7. Only the indicated reagents and enzymes have been used. While other sources might be compatible, these should be independently tested.
8. Special codes for oligonucleotides: /5rApp/, 5′ adenylation; /3ddC/, 3′ dideoxycytidinylation; /5Phos/, 5′ phosphorylation; /iSp18/, an 18-atom hexa-ethyleneglycol spacer; /5Biosg/, 5′ biotinylation, /3phos/, 3′ phosphorylation.
9. It is strongly advised to avoid centrifugation as a method of harvesting, as cells harvested by centrifugation versus rapid filtration show measurably different ribosome occupancy profiles [25]. Updated protocols even recommend directly freezing the liquid culture as a harvesting method, particularly for evaluating ribosome pause sites [16].

10. Purifying and sequencing mRNAs are needed for translation efficiency measurements, particularly when comparing genes encoded on distinct operons. For mRNA enrichment, save one-third of the pulverized cell powder, purify total RNA by acid phenol extraction, and prepare sequencing library.
11. The oligonucleotide oNTI199 AUGUACACGGAGUC GACCCGCAACGCGA/3phos/ (*see* **Note 8**) is often used as a control to gauge reaction efficiency.
12. 10 bp DNA ladder commonly used in most ribosome profiling protocols is now discontinued. As a replacement, other DNA ladders can be readily substituted, or specific oligonucleotides of defined lengths can be synthesized and used as a sizing standard.
13. All gel extraction steps can be replaced by an automated size selection approach (Pippin Prep, Sage Science). However, this requires a dedicated machine and may not be cost effective for some laboratories.

Acknowledgments

I thank Emily Powers and Gloria Brar for use of reagents and equipment.

References

1. Brar GA, Weissman JS (2015) Ribosome profiling reveals the what, when, where and how of protein synthesis. Nat Rev Mol Cell Biol 16(11):651–664. https://doi.org/10.1038/nrm4069
2. Ingolia NT, Hussmann JA, Weissman JS (2019) Ribosome profiling: global views of translation. Cold Spring Harb Perspect Biol 11:5. https://doi.org/10.1101/cshperspect.a032698
3. Otto GM, Brar GA (2018) Seq-ing answers: uncovering the unexpected in global gene regulation. Curr Genet 64(6):1183–1188. https://doi.org/10.1007/s00294-018-0839-3
4. Ndah E, Jonckheere V, Giess A, Valen E, Menschaert G, Van Damme P (2017) REPARATION: ribosome profiling assisted (re-) annotation of bacterial genomes. Nucleic Acids Res 45(20):e168. https://doi.org/10.1093/nar/gkx758
5. Weaver J, Mohammad F, Buskirk AR, Storz G (2019) Identifying small proteins by ribosome profiling with stalled initiation complexes. MBio 10:2. https://doi.org/10.1128/mBio.02819-18
6. Giess A, Jonckheere V, Ndah E, Chyżyńska K, Van Damme P, Valen E (2017) Ribosome signatures aid bacterial translation initiation site identification. BMC Biol 15(1):76. https://doi.org/10.1186/s12915-017-0416-0
7. Meydan S, Marks J, Klepacki D, Sharma V, Baranov PV, Firth AE, Margus T, Kefi A, Vázquez-Laslop N, Mankin AS (2019) Retapamulin-assisted ribosome profiling reveals the alternative bacterial proteome. Mol Cell 74(3):481–493.e486. https://doi.org/10.1016/j.molcel.2019.02.017
8. Nakahigashi K, Takai Y, Kimura M, Abe N, Nakayashiki T, Shiwa Y, Yoshikawa H, Wanner BL, Ishihama Y, Mori H (2016) Comprehensive identification of translation start sites by tetracycline-inhibited ribosome profiling. DNA Res 23(3):193–201. https://doi.org/10.1093/dnares/dsw008
9. Schrader JM, Zhou B, Li G-W, Lasker K, Childers WS, Williams B, Long T, Crosson S, McAdams HH, Weissman JS, Shapiro L (2014)

The coding and noncoding architecture of the Caulobacter crescentus genome. PLoS Genet 10(7):e1004463. https://doi.org/10.1371/journal.pgen.1004463

10. Balakrishnan R, Oman K, Shoji S, Bundschuh R, Fredrick K (2014) The conserved GTPase LepA contributes mainly to translation initiation in Escherichia coli. Nucleic Acids Res 42(21):13370–13383. https://doi.org/10.1093/nar/gku1098
11. Hwang J-Y, Buskirk AR (2017) A ribosome profiling study of mRNA cleavage by the endonuclease RelE. Nucleic Acids Res 45 (1):327–336. https://doi.org/10.1093/nar/gkw944
12. Meydan S, Klepacki D, Karthikeyan S, Margus T, Thomas P, Jones JE, Khan Y, Briggs J, Dinman JD, Vázquez-Laslop N, Mankin AS (2017) Programmed ribosomal frameshifting generates a copper transporter and a copper chaperone from the same gene. Mol Cell 65(2):207–219. https://doi.org/10.1016/j.molcel.2016.12.008
13. Burkhardt DH, Rouskin S, Zhang Y, Li G-W, Weissman JS, Gross CA (2017) Operon mRNAs are organized into ORF-centric structures that predict translation efficiency. elife 6: e22037. https://doi.org/10.7554/eLife.22037
14. Li G-W, Burkhardt D, Gross C, Weissman JS (2014) Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. Cell 157(3):624–635. https://doi.org/10.1016/j.cell.2014.02.033
15. Li G-W, Oh E, Weissman JS (2012) The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. Nature 484(7395):538–541. https://doi.org/10.1038/nature10965
16. Mohammad F, Green R, Buskirk AR (2019) A systematically-revised ribosome profiling method for bacteria reveals pauses at single-codon resolution. elife 8:e42591. https://doi.org/10.7554/eLife.42591
17. Woolstenhulme CJ, Guydosh NR, Green R, Buskirk AR (2015) High-precision analysis of translational pausing by ribosome profiling in bacteria lacking EFP. Cell Rep 11(1):13–21. https://doi.org/10.1016/j.celrep.2015.03.014
18. Nakahigashi K, Takai Y, Shiwa Y, Wada M, Honma M, Yoshikawa H, Tomita M, Kanai A, Mori H (2014) Effect of codon adaptation on codon-level and gene-level translation efficiency in vivo. BMC Genomics 15(1):1115. https://doi.org/10.1186/1471-2164-15-1115
19. Kannan K, Kanabar P, Schryer D, Florin T, Oh E, Bahroos N, Tenson T, Weissman JS, Mankin AS (2014) The general mode of translation inhibition by macrolide antibiotics. Proc Natl Acad Sci U S A 111(45):15958–15963. https://doi.org/10.1073/pnas.1417334111
20. Marks J, Kannan K, Roncase EJ, Klepacki D, Kefi A, Orelle C, Vázquez-Laslop N, Mankin AS (2016) Context-specific inhibition of translation by ribosomal antibiotics targeting the peptidyl transferase center. Proc Natl Acad Sci U S A 113(43):12150–12155. https://doi.org/10.1073/pnas.1613055113
21. Haft RJF, Keating DH, Schwaegler T, Schwalbach MS, Vinokur J, Tremaine M, Peters JM, Kotlajich MV, Pohlmann EL, Ong IM, Grass JA, Kiley PJ, Landick R (2014) Correcting direct effects of ethanol on translation and transcription machinery confers ethanol tolerance in bacteria. Proc Natl Acad Sci U S A 111 (25):E2576–E2585. https://doi.org/10.1073/pnas.1401853111
22. Zhang Y, Burkhardt DH, Rouskin S, Li G-W, Weissman JS, Gross CA (2018) A stress response that monitors and regulates mRNA structure is central to cold shock adaptation. Mol Cell 70(2):274–286.e277. https://doi.org/10.1016/j.molcel.2018.02.035
23. Zhang Y, Xiao Z, Zou Q, Fang J, Wang Q, Yang X, Gao N (2017) Ribosome profiling reveals genome-wide cellular translational regulation upon heat stress in Escherichia coli. Genomics Proteomics Bioinformatics 15 (5):324–330. https://doi.org/10.1016/j.gpb.2017.04.005
24. Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324 (5924):218–223. https://doi.org/10.1126/science.1168978
25. Becker AH, Oh E, Weissman JS, Kramer G, 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 Protoc 8 (11):2212–2239. https://doi.org/10.1038/nprot.2013.133
26. Oh E, Becker AH, Sandikci A, Huber D, Chaba R, Gloge F, Nichols RJ, Typas A, Gross CA, Kramer G, Weissman JS, Bukau B (2011) Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. Cell 147(6):1295–1308. https://doi.org/10.1016/j.cell.2011.10.044

27. Datta AK, Burma DP (1972) Association of ribonuclease I with ribosomes and their subunits. J Biol Chem 247(21):6795–6801
28. Gerashchenko MV, Gladyshev VN (2017) Ribonuclease selection for ribosome profiling. Nucleic Acids Res 45(2):e6. https://doi.org/10.1093/nar/gkw822
29. Johnson GE, Li G-W (2018) Genome-wide quantitation of protein synthesis rates in bacteria. Meth Enzymol 612:225–249. https://doi.org/10.1016/bs.mie.2018.08.031
30. Mohammad F, Buskirk AR (2019) Protocol for ribosome profiling in bacteria. Bio Protoc 9 (24):e3468. https://doi.org/10.21769/BioProtoc.3468
31. Fishman A, Lamm AT (2019) QsRNA-seq: a protocol for generating libraries for high-throughput sequencing of small RNAs. Bio-protocol 9(5):e3179. https://doi.org/10.21769/BioProtoc.3179

Chapter 2

Identification of Translation Start Sites in Bacterial Genomes

Sezen Meydan, Dorota Klepacki, Alexander S. Mankin, and Nora Vázquez-Laslop

Abstract

The knowledge of translation start sites is crucial for annotation of genes in bacterial genomes. However, systematic mapping of start codons in bacterial genes has mainly relied on predictions based on protein conservation and mRNA sequence features which, although useful, are not always accurate. We recently found that the pleuromutilin antibiotic retapamulin (RET) is a specific inhibitor of translation initiation that traps ribosomes specifically at start codons, and we used it in combination with ribosome profiling to map start codons in the *Escherichia coli* genome. This genome-wide strategy, that was named Ribo-RET, not only verifies the position of start codons in already annotated genes but also enables identification of previously unannotated open reading frames and reveals the presence of internal start sites within genes. Here, we provide a detailed Ribo-RET protocol for *E. coli*. Ribo-RET can be adapted for mapping the start codons of the protein-coding sequences in a variety of bacterial species.

Key words Ribosome profiling, Ribo-Seq, Translation initiation, Start codons, Bacterial translation, Retapamulin, Pleuromutilin, Alternative proteome

1 Introduction

Translation initiation in bacteria ensues with the small (30S) ribosomal subunit recognizing a start codon of the protein coding region in mRNA [1]. The recognition of a translation initiation site (TIS) and recruitment of the initiator formylmethionyl-tRNA^{fMet} (fMet-tRNA^{fMet}) to the P site of the 30S subunit are assisted by the three initiation factors IF1, IF2, and IF3. Binding of the fMet-tRNA^{fMet} at the initiation codon defines the starting point of translation and sets up the reading frame. Upon dissociation of the IFs and binding of the large (50S) ribosomal subunit 70S initiation complex is formed at the start codon. Accommodation of the first elongator tRNA into the ribosomal A site, catalysis of the first peptide bond, and subsequent translocation of the ribosome to the second codon denotes the transition to the elongation step of translation [2, 3].

Vyacheslav M. Labunskyy (ed.), *Ribosome Profiling: Methods and Protocols*, Methods in Molecular Biology, vol. 2252, https://doi.org/10.1007/978-1-0716-1150-0_2,

AUG and GUG are the most commonly used start codons in the bacterial genes. However, some other triplets, such as UUG, CUG, AUU, and AUC, which can be decoded by the initiator fMet-tRNA^{fMet} can be also employed with varying efficiency for translation initiation [4]. In most of the bacterial genomes, the start codons of the ORFs are often preceded by a Shine-Dalgarno (SD) sequence (with the consensus GGAGG), which is fully or partially complementary to a stretch of nucleotides at the 3′ end of the 16S rRNA of the 30S subunit [5]. The presence of a SD sequence, however, is not a prerequisite for the start codon recognition or for efficient translation initiation [6–9], and it may play only a supportive role [10, 11]. Accessibility of the TIS to the ribosome, which depends on mRNA folding, is another important factor that affects initiation [6, 12–14]. Additional, yet still poorly understood, features help the ribosome to find and distinguish the start codon from similar ones within or outside of the ORFs [15, 16].

Accurate identification of TISs is crucial for correct annotation of the genomes, mapping the boundaries of the ORFs and gaining comprehensive information about the proteome. Several proteomics approaches based on identifying the N-terminal peptides of the proteins have been developed for experimental mapping of TISs in bacteria [17–19]. The best-suited proteomics technique for identifying the authentic N-terminal peptides relies on treating bacteria with actinonin, whose ability to inhibit peptide deformylase leads to retention of the formyl-methionine [17, 18]. Alternative approaches based on bottom-up mass spectrometry also hold potential for revealing TISs of unannotated ORFs [19]. These proteomics techniques, however, are limited by the size and abundance of the proteins and may not always report the authentic TIS of a gene or detect the presence of alternative start sites.

Computational approaches have also been widely employed for identifying boundaries of bacterial ORFs. Various algorithms that analyze codon usage, biases in nucleotide periodicity, the presence of known TIS signatures, or the conservation of the encoded proteins can relatively accurately predict the start codons of many genes [20–26]. However, these approaches often fail to distinguish between closely spaced putative start codons and are poorly applicable for identifying small ORFs [27, 28]. Furthermore, identification of alternative TISs that are utilized for expression of more than one protein product from a single coding sequence or detecting in-frame or out-of-frame ORFs within ORFs [29, 30] represents a formidable problem for the available computational techniques.

More sophisticated analysis of TISs has been carried out using ribosome profiling (Ribo-Seq), a genome-wide technique employing deep sequencing of ribosome-protected mRNA fragments. Ribo-Seq shows the distribution of ribosomes along the translated mRNAs [31]. Ribo-Seq, in combination with computational

algorithms and proteomics, has been utilized to map TISs, re-annotate prokaryotic genomes, and detect N-terminal protein extensions in various bacterial species [32–34]. Nevertheless, the direct use of Ribo-Seq for mapping ORF boundaries faces specific challenges. The mRNA coverage and resolution of the Ribo-Seq data near start codons vary between different genes. Furthermore, direct Ribo-Seq is poorly applicable for mapping internal TISs (iTISs) located within the coding regions because the footprints originated from elongating ribosome may obscure the footprints that come from the ribosomes engaged at an iTIS (Fig. 1). One solution to circumvent these limitations is to specifically capture ribosomes at start codons by arresting them at the translation

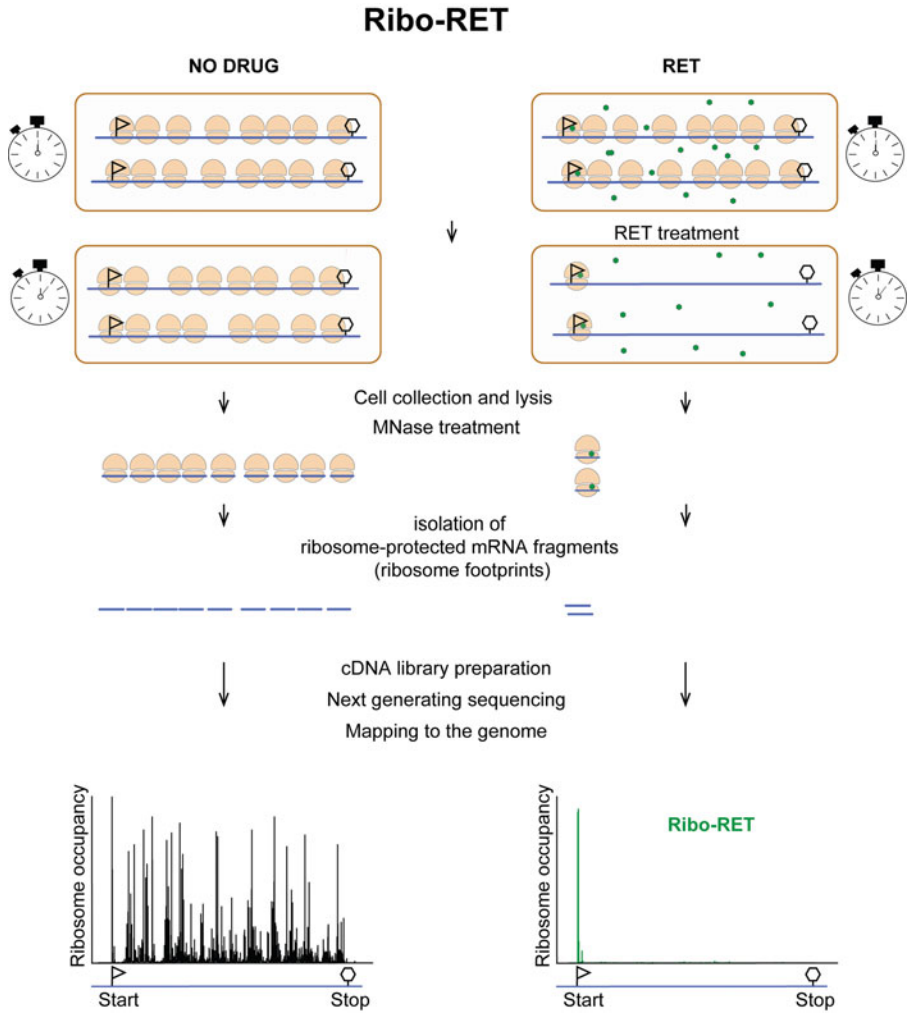

Fig. 1 The Ribo-RET experimental pipeline. Bacterial cells are first treated with RET at a concentration and incubation time that are pre-determined by metabolic labeling experiments. Upon cell collection and lysis, conventional Ribo-Seq experimental steps are performed. To assess the general translation status of the cells, a culture where the drug treatment is omitted (indicated as *no drug*) is prepared and processed in parallel

initiation step while clearing the mRNAs from the elongating ribosomes. Several such approaches have been carried out for eukaryotic systems [35–38]. When Ribo-Seq was performed with bacteria treated with the translation inhibitor tetracycline, it was noted that high peaks of ribosome density accumulated specifically at the start codons of the genes [39]. However, interpretation of the tetracycline-based Ribo-Seq data is not straightforward because this antibiotic can bind to elongating ribosomes. To be able to more reliably map translation start sites in bacteria, we used the bacterial translation initiation inhibitor retapamulin (RET) to develop RET-assisted Ribo-Seq or Ribo-RET [30]. RET binds to the ribosomal peptidyl transferase center, where it overlaps with both the A and P sites [40–42]. RET can readily bind to the initiating ribosome that carries fMet-tRNA^{fMet} in its P site, but its association with the elongating ribosome is barred due to the steric clash with the growing protein chain. Bound to the initiating ribosome, RET precludes placing of the aminoacyl-tRNA in the A site. As a result, it blocks the first peptide bond formation and arrests the ribosome at the start codon [30]. The use of Ribo-RET allowed for comprehensive mapping of TISs throughout the *E. coli* genome, revealing numerous unannotated genes outside of the coding regions [27] and identifying many unknown iTISs [30].

Ribo-RET consists of three main steps: (1) optimizing the RET treatment of bacterial cells in order to ensure high enrichment of translating ribosomes arrested at the start codons of the mRNAs (Fig. 1), (2) applying the Ribo-Seq protocol to prepare next-generation sequencing (NGS)-compatible cDNA libraries of the ribosomal footprints from untreated and RET-treated cells (Fig. 2), (3) and mapping the ribosomal footprints to the genome and identifying start sites.

Because Ribo-RET relies on a brief treatment of bacterial cells with RET, the minimum inhibitory concentration of RET (MIC_{RET}) should be first estimated, as it can greatly vary between different bacterial strains [43]. RET is highly active against many Gram-positive bacteria but shows relatively low activity against Gram-negative bacterial species. Genetic manipulation of the target strain, for example, inactivation of the drug efflux pumps (e.g., TolC in *E. coli*) or increasing permeability of the outer membrane, may be required to achieve sufficient level of translation inhibition by RET. The procedures described here were optimized for the RET-hypersusceptible *E. coli* strain BL21 Δ*tolC* for which MIC_{RET} is 0.06–0.12 μg/mL [30]. Newer antibiotics of the same class, e.g., lefamulin [44], are reported to have higher activity and potentially could be used directly against some bacteria that are naturally resistant to retapamulin. The subsequent Ribo-Seq steps have been largely adapted from published procedures [45, 46] and minimally modified using our own experience.

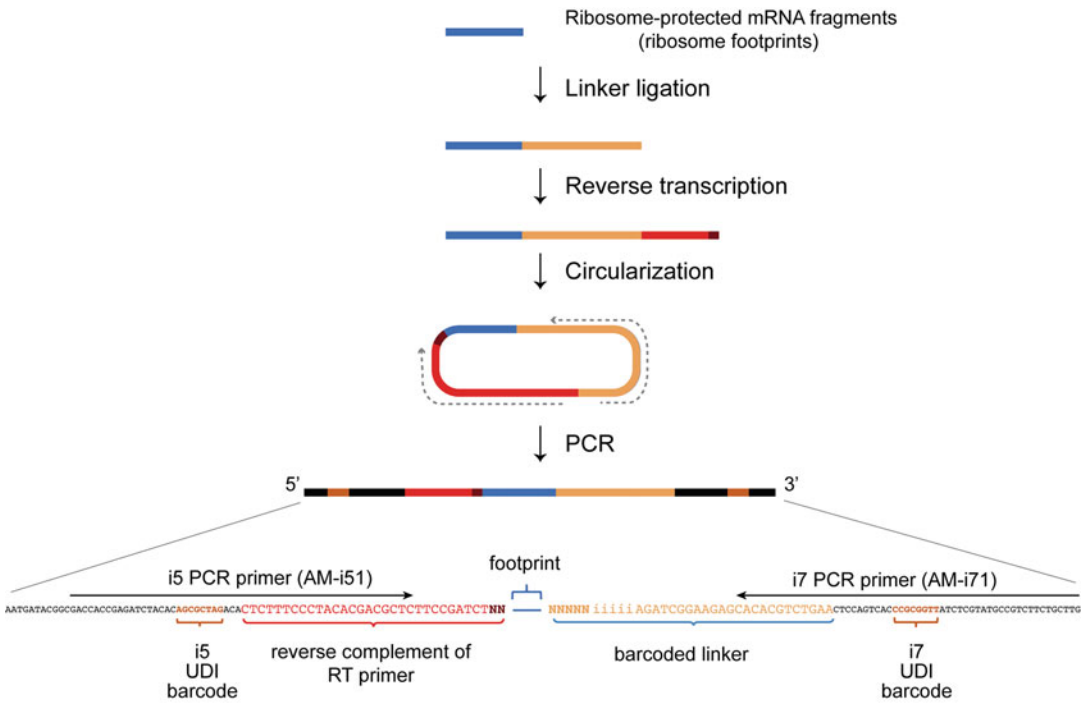

Fig. 2 The steps of the Ribo-Seq protocol for preparing samples amenable for next-generation sequencing. The general structure of the final product for Illumina sequencing is shown (*see* Subheading 3.12). Illumina unique dual indexes (UDI) (in this case derived from AM-i51 and AM-i71 PCR primers) (*see* Subheadings 2.12 and 3.12) are indicated. "N" indicates a random nucleotide; "iiiii" indicates a unique barcode sequence that resides within NI-810 through NI-817 linkers (*see* Subheadings 2.9 and 3.9)

2 Materials

2.1 Metabolic Labeling with RET

1. Growth medium M9AA-minus-Met: M9 minimal medium containing 19 amino acids (all natural amino acids except methionine), at a final concentration of 40 μg/mL each, supplemented with 3 μM thiamine (added right before use from the thiamine stock solution).
2. Thiamine stock solution: Dissolve thiamine in ddH_2O to the concentration of 3 mM and store at −20 °C.
3. Retapamulin: Dissolve in 100% ethanol to the concentration of 10 mg/mL and store at −20 °C.
4. [^{35}S]-L-Methionine with specific activity of ~1000 Ci/mmol at ~10 mCi/mL (*see* **Note 1**).
5. Trichloroacetic acid (TCA): Prepare 100% (weight/volume) TCA solution by adding 22.7 mL of ddH_2O to 50 g TCA, then dilute to a final concentration of 5% with ddH_2O, and store it at 4 °C.

6. Filter discs: Whatman Grade 3MM Chr Cellulose 0.34 mm thick, ∅ 2.5 cm circle discs.
7. Acetone: 99.9% ACS reagent grade.
8. Scintillation vials, scintillation cocktail, and scintillation counter.

2.2 Preparation of Cell Lysates for Ribo-RET

1. MOPS growth medium: prepare from MOPS EZ Rich Defined Medium Kit (*see* **Note 2**).
2. Retapamulin: prepare stock solution (*see* **item 3** of Subheading 2.1).
3. Glass filtration system capable of accommodating ∅ 90 mm filters, connected to a vacuum pump.
4. Filter discs: Millipore Express PLUS membrane hydrophilic polyethersulfone filters, ∅ 90 mm, 0.22 μm pore size.
5. Liquid nitrogen.
6. Stainless-steel Scoopula spatulas (scoopulas) ~15 cm long; container tubes compatible with holding liquid nitrogen and deep enough to encase 90% of the length of the scoopula; 50 mL conical tubes whose lids have been pierced 5–8 times with a 20 gauge needle.
7. Lysis buffer: 20 mM Tris–HCl, pH 8.0, 10 mM $MgCl_2$, 100 mM NH_4Cl, 5 mM $CaCl_2$, 0.4% Triton X-100, 0.1% NP-40. Store at −20 °C.
8. DNase I (RNase-free), 10 U/μL.
9. GMPPNP: dissolve in ddH_2O to the concentration of 100 mM and store at −20 °C.
10. SUPERase-In RNase inhibitor, 20 U/μL.

2.3 Pulverization of Cells

1. Mixer mill, 10 mL jar, 12-mm-diameter grinding ball.
2. Spatulas, 50 mL conical tubes with pierced lid (*see* **item 6** of Subheading 2.2).
3. Liquid nitrogen.

2.4 Preparation of Cell Lysate an of Ribosome-Protected mRNA Fragments

1. Tris solution: 10 mM Tris–HCl, pH 7.0.
2. SUPERase-In, 20 U/μL.
3. Micrococcal nuclease (MNase) solution: take MNase from the original vial and dilute it with 10 mM Tris pH 8.0 to 75 U/μL (*see* **Note 3**). Store at −80 °C.
4. EGTA solution: 0.5 mM EGTA at pH 8.0.
5. Nanodrop spectrophotometer.

2.5 Monosome Isolation by Sucrose Gradient Centrifugation

1. Gradient buffer: 20 mM Tris–HCl, pH 8.0, 10 mM $MgCl_2$, 100 mM NH_4Cl, prepared with DEPC-treated ddH_2O.
2. Sucrose solutions: 10% and 40% (weight/volume) solutions prepared in 20 mM Tris–HCl, pH 8.0, 10 mM $MgCl_2$, 100 mM NH_4Cl, prepared with DEPC-treated ddH_2O. Store at 4 °C.
3. Beckman SW41 rotor and open-top polyclear ultracentrifuge tubes.
4. Sucrose gradient maker.
5. Gradient fractionation system.

2.6 Phenol-Chloroform Extraction of Ribosome-Protected mRNA Fragments

1. SDS solution: 20% (weight/volume) SDS dissolved in RNase-free ddH_2O.
2. Acidic phenol solution: phenol/chloroform/isoamyl alcohol 125:24:1 at pH 4.5, molecular biology grade.
3. Chloroform, molecular biology grade.
4. 3 M NaOAc, pH 5.5 prepared in RNase-free ddH_2O.
5. Isopropanol, molecular biology grade.
6. 80% ethanol, ice-cold.
7. Tris solution: 10 mM Tris–HCl pH 7.0.

2.7 Size Selection of Ribosome-Protected mRNA Fragments

1. Tris solution: 10 mM Tris–HCl pH 7.0.
2. Novex 2× TBE-Urea Sample Buffer.
3. Control RNA oligonucleotides solutions: from 100 μM stock solutions, dilute each RNA oligonucleotide to 20 μM in DEPC-treated ddH_2O and store at −20 °C. The names and sequences of the control RNA oligos are as follows:

   o15, 5′_AUGUACACGGAGUCG_3′,

   o28, 5′_AUGUACACGGAGUCGACCCGCAACGCGA_3′,

   o45, 5′_AUGUACACGGAGUCGACCCGCAACGCGAUGUACACGGAGUCGAC_3′.
4. 15% TBE-Urea gel: 15% denaturing polyacrylamide TBE-Urea gel, 8 ×8 cm, 1 mm thick.
5. 10× TBE running buffer.
6. SYBR Gold nucleic acid gel stain.
7. RNA elution buffer: 300 mM NaOAc pH 5.5, 1 mM EDTA pH 8.0. Store at room temperature.
8. SUPERase-In, 20 U/μL.
9. Spin-X columns: Spin-X centrifuge tube filters with cellulose acetate membrane, pore size 0.22 μm.
10. Glycoblue.

2.8 Dephosphorylation

1. T4 polynucleotide kinase (PNK), 10 U/μL.
2. SUPERase-In, 20 U/μL.
3. Tris solution: 10 mM Tris–HCl pH 7.0.
4. 3 M NaOAc, pH 5.5.
5. Glycoblue.

2.9 Enzymatic Pre-adenylation of Linkers and Linker Ligation

1. 5′ DNA Adenylation Kit: Mth RNA Ligase.
2. Linker oligonucleotides [46] (linker-specific barcode is underlined):

   NI-810, 5′Phos/NNNNNATCGTAGATCGGAAGAGCA CACGTCTGAA/3′ddC.

   NI-811, 5′Phos/NNNNNAGCTAAGATCGGAAGAGCA CACGTCTGAA/3′ddC.

   NI-812, 5′Phos/NNNNNCGTAAAGATCGGAAGAGCA CACGTCTGAA/3′ddC.

   NI-813, 5′Phos/NNNNNCTAGAAGATCGGAAGAGCA CACGTCTGAA/3′ddC.

   NI-814, 5′Phos/NNNNNGATCAAGATCGGAAGAGCA CACGTCTGAA/3′ddC.

   NI-815, 5′Phos/NNNNNGCATAAGATCGGAAGAGCA CACGTCTGAA/3′ddC.

   NI-816, 5′Phos/NNNNNTAGACAGATCGGAAGAGCA CACGTCTGAA/3′ddC.

   NI-817, 5′Phos/NNNNNTCTAGAGATCGGAAGAGCA CACGTCTGAA/3′ddC.
3. Oligo Clean & Concentrator kit.
4. T4 RNA ligase kit: T4 RNA ligase 2, truncated K227Q.
5. 5′ deadenylase, 50 U/μL.
6. $RecJ_f$ DNA exonuclease, 30 U/μL.

2.10 Reverse Transcription

1. dNTP mix: a mix of dATP, dCTP, dGTP, and dTTP each at 10 mM.
2. NI-802 DNA oligo: dilute with ddH_2O to 25 μM from a 100 μM stock solution of NI-802 DNA oligo [46] and store at −20 °C.

   NI-802, 5′Phos/NNAGATCGGAAGAGCGTCGTGTAGG GAAAGAG/iSp18/GTGACTGGAGTTCAGACGTGT GCTC.
3. Reverse transcriptase (RT) kit: SuperScript III Reverse Transcriptase, 200 U/μL.
4. SUPERase-In, 20 U/μL.
5. NaOH solution: 1 M NaOH.

6. Oligo Clean & Concentrator kit.
7. Tris solution: 10 mM Tris–HCl, pH 8.0.
8. 10% TBE-Urea gel: 10% denaturing polyacrylamide TBE-Urea gel, 8 × 8 cm, 1 mm thick.
9. 10× TBE running buffer.
10. SYBR Gold.
11. 20 gauge needles.
12. DNA elution buffer. To prepare 50 mL, mix the following:

5 M NaCl	3 mL
0.5 M EDTA pH 8.0	100 µL
1 M Tris–HCl pH 8.0	500 µL
DEPC-treated ddH_2O	46.4 mL

Store at room temperature.

13. Spin-X columns: Spin-X centrifuge tube filters with cellulose acetate membrane, pore size 0.22 µm.
14. Glycoblue.

2.11 Circularization

1. CircLigase ssDNA Ligase, 100 U/µL.

2.12 PCR Amplification

1. Phusion High-Fidelity DNA polymerase, 2000 U/mL.
2. 10 mM dNTP mix.
3. PCR primers. Prepare 100 µM stock solutions and store at −20 °C. The DNA primer pairs are as follows (Illumina unique dual indexes (UDI) are underlined):

AM-i51, 5′_AATGATACGGCGACCACCGAGATCTACAC<u>AGCGCTAG</u>ACACTCTTTCCCTACACGACGCTC_3′.

AM-i71, 5′_CAAGCAGAAGACGGCATACGAGAT<u>AACCGCGG</u>GTGACTGGATTCAGACGTGTG_3′.

AM-i52, 5′_AATGATACGGCGACCACCGAGATCTACAC<u>GATATCGA</u>ACACTCTTTCCCTACACGACGCTC_3′.

AM-i72, 5′_CAAGCAGAAGACGGCATACGAGAT<u>GGTTATAA</u>GTGACTGGAGTTCAGACGTGTG_3′.

AM-i53, 5′_AATGATACGGCGACCACCGAGATCTACAC<u>CGCAGACG</u>ACACTCTTTCCCTACACGACGCTC_3′.

AM-i73, 5′_CAAGCAGAAGACGGCATACGAGAT<u>CCAAGTCC</u>GTGACTGGAGTTCAGACGTGTG_3′.

AM-i54, 5′_AATGATACGGCGACCACCGAGATCTACAC<u>TATGAGTA</u>ACACTCTTTCCCTACACGACGCTC_3′.

AM-i74, 5′_CAAGCAGAAGACGGCATACGAGAT<u>TTGGACTT</u>GTGACTGGAGTTCAGACGTGTG_3′.

AM-i55, 5′_AATGATACGGCGACCACCGAGATCTACACAGGTGCGTACACTCTTTCCCTACACGACGCTC_3′.

AM-i75, 5′_CAAGCAGAAGACGGCATACGAGATCAGTGGATGTGACTGGAGTTCAGACGTGTG_3′.

AM-i56, 5′_AATGATACGGCGACCACCGAGATCTACACGAACATACACACTCTTTCCCTACACGACGCTC_3′.

AM-i76, 5′_CAAGCAGAAGACGGCATACGAGATTGACAAGCGTGACTGGAGTTCAGACGTGTG_3′.

4. 6× DNA loading dye.
5. 10 bp DNA ladder, ready-to-use 10–150 bp.
6. 8% TBE gel, 8 × 8 cm, 1 mm thick.
7. 1× TBE buffer.
8. SYBR Gold.
9. Tris solution: 10 mM Tris–HCl, pH 8.0.
10. DNA elution buffer, described in **item 12** of Subheading 2.10.

3 Methods

3.1 Optimization of RET Treatment

The goal of this experiment is to optimize the treatment of the bacterial cells with RET to achieve complete inhibition of protein synthesis (as estimated by incorporation of [^{35}S]-Met into polypeptides) in a short period of time (2–5 min) (Fig. 3) (*see* **Notes 4–5**).

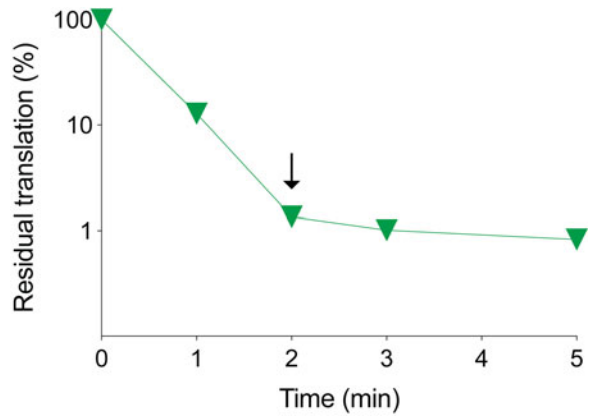

Fig. 3 Metabolic labeling experiment determining the residual protein synthesis in *E. coli* BL21 *ΔtolC* cells treated for the indicated times with RET. The arrow indicates the shortest incubation time where maximum protein synthesis inhibition was achieved

1. Grow an overnight cell culture at 37 °C in an incubator shaker in M9AA-minus-Met medium freshly supplemented with thiamine.

   Steps 2–9 are optimally carried out in a 37 °C room (warm room):

2. Dilute the overnight culture 1:200 into 5 mL of M9AA-minus-Met medium freshly supplemented with thiamine, and grow it with constant shaking until it reaches an A_{600} ~0.2. While the culture is growing, prepare the materials described in **steps 3–6** (*see* **Note 1**).
3. Dilute 1 μL (~10 μCi) of [^{35}S]-Met into 65 μL of M9AA-minus-Met freshly supplemented with thiamine. Prepare a set of 1.5 mL Eppendorf tubes labeled with the pre-chosen time points for the RET treatment of the cell culture (e.g., 0, 1, 2.5, 5, and 10 min). Pipette 2 μL of the diluted [^{35}S]-Met into each of the tubes.
4. Prepare a glass beaker with 500 mL of the 5% TCA (cover the beaker with aluminum foil to minimize exposure to the TCA fumes).
5. Label a set of filter discs with pencil (as the discs are going to be submerged in the TCA solution and solvent-based labels would be washed out), and place them in a shallow container, e.g., a Petri dish. Pre-soak each disc with ~25 μL of TCA solution and let them air-dry for 1–2 min.
6. Prepare a control sample for non-specific binding of [^{35}S]-Met to the filter: Add 28 μL of 5% TCA and to one of the tubes with 2 μL of the diluted [^{35}S]-Met (**step 4**). Pipette 25 μL of this mixture on one of the filter discs from **step 5**. Immediately place the disc in the beaker with 5% TCA.
7. Prepare the sample for 0-time RET treatment: Place a 28 μL aliquot of the culture at A_{600} ~0.2 (**step 2**) into one of the tubes with 2 μL of the diluted [^{35}S]-Met (**step 3**). Incubate for 1 min and then pipette 25 μL of this mixture onto one of the filter discs. Immediately place the disc in the beaker with 5% TCA.
8. Prepare the samples for the time course of RET treatment: Pipette 350 μL aliquot of the culture at A_{600} ~0.2 (**step 2**) into an Eppendorf tube. Add RET stock solution to reach a concentration 50–100× the MIC_{RET} (5 μg/mL for *E. coli* BL21 Δ*tolC*), vortex, and immediately start the timer.
9. At the required times, transfer 28 μL to the tubes with 2 μL of the diluted [^{35}S]-Met (**step 3**). Incubate for 1 min and then pipette 25 μL of this mixture on one of the filter discs (**step 4**). Immediately place the disc in the beaker with 5% TCA.
10. Once all the sample-containing filter discs are in the beaker, bring its content to boil under a fume hood, and keep boiling

for 5 min. Decant and discard the TCA, being careful to keep the filter discs inside the beaker. Add a fresh batch of 500 mL of 5% TCA. Boil for additional 5 min under a fume hood and discard the TCA solution. Add ~100 mL of acetone to the beaker; swirl it for ~3 min. Dry the filter discs for ~5 min under a fume hood.

11. Place the discs into scintillation vials; add scintillation liquid. Measure radioactivity in a scintillation counter.
12. Plot [^{35}S]-Met incorporation in the RET-treated samples relative to the 100% incorporation control sample (**step 7**) (Fig. 2). For the Ribo-RET procedure, choose the shortest RET treatment time where maximum protein synthesis inhibition was achieved (Fig. 3).

3.2 Collection of Cells Treated with RET

The aim of this procedure is to collect bacterial cells enriched with ribosomes stalled at initiation codons by the action of RET. Cells are exposed to RET under conditions optimized in Subheading 3.1. A control culture devoid of RET treatment is recommended to be processed in parallel.

1. Grow an overnight culture in MOPS medium (*see* **Note 6**) at 37 °C in an incubator shaker.

   Steps 2–4 are optimally carried out in a 37 °C temperature room. The scoopulas for scraping cells and the filtration apparatus connected to a vacuum line should be placed in the 37 °C room at least 30 min before cell harvesting takes place.
2. Dilute the overnight culture from **step 1** to A_{600} ~0.05 in MOPS media pre-warmed to 37 °C. Grow the diluted culture, shaking, until it reaches an A_{600} ~0.3 (*see* **Notes 6–8**).
3. Add RET to the cell culture according to the antibiotic concentration and time of exposure determined by the metabolic labeling experiments described in Subheading 3.1 (*see* **Note 4**), and continue shaking.
4. Filter the cells as rapidly as possible and immediately scrape them off the filter using a scoopula (*see* **Notes 6–8**). Quickly submerge the scoopula with the cell pellet in the container tube filled with liquid nitrogen.
5. Using another scoopula pre-chilled in liquid nitrogen, dislodge the frozen cells in a 50 mL conical tube (with pierced lid) containing ~10 mL of liquid nitrogen. Cell pellets can be stored at −80 °C (liquid nitrogen will evaporate through the lid holes). Otherwise, continue to Subheading 3.3.

3.3 Preparation of Cell Lysates

1. Freshly supplement 750 μL of lysis buffer with 7.5 μL DNase I (RNase-free), 22.5 μL GMPPNP, and 12 μL SUPERase-In.

2. Take the 50 mL tubes with the frozen cell pellets (*see* Subheading 3.2), and add ~10 mL of liquid nitrogen, re-filling to this volume with liquid nitrogen as needed. Slowly drip 650 μL of lysis buffer (**step 1**) in order to form small, discrete droplets.
3. Pre-chill the jar and grinding ball of the mixer mill in liquid nitrogen.
4. Take the frozen cell pellets and drops of frozen lysis buffer (care should be taken to keep them frozen at all times), and immediately transfer them into the pre-chilled jar with the grinding ball in it.
5. Lyse the frozen cells in the mixer mill by carrying five cycles of 3 min each at 15 Hz, re-chilling the jar in liquid nitrogen following each cycle.
6. Use a pre-chilled spatula to transfer the pulverized frozen cells into a 50 mL conical tube containing 10 mL of liquid nitrogen. Close with a pierced lid and either store at −80 °C or proceed to Subheading 3.4.

3.4 Preparation of Ribosome-Protected mRNA Fragments

Treatment of cell lysate with MNase results in degradation of the mRNA fragments not protected by stalled ribosomes and conversion of polysomes to monosomes (*see* **Note 9**).

1. Collect the lysates (*see* Subheading 3.3) to the bottom of the tubes by briefly spinning in a 4 °C centrifuge.
2. Thaw the lysates by placing the tubes for 2 min in a 30 °C water bath for 20 min, and transfer them to pre-chilled Eppendorf tubes.
3. Pellet insoluble debris at 20,000 × g for 10 min in a 4 °C microfuge.
4. Carefully transfer supernatants to pre-chilled Eppendorf tubes.
5. Blank the Nanodrop spectrophotometer with an aliquot of lysis buffer (from **step 1** of Subheading 3.3) diluted 1:100 with 10 mM Tris. Mix 2 μL of clarified lysate from **step 5** with 198 μL of 10 mM Tris and measure A_{260}.
6. Determine the concentration (A_{260} units/μL) of the lysates.
7. Dilute 22 A_{260} units of lysate into a final volume of 220 μL of lysis buffer (**step 1** of Subheading 3.3) (*see* **Note 10**).
8. Add 6 μL of SUPERase-In and 4.4 μL of the MNase solution (*see* **Note 3**). Incubate at 25 °C for 1 h with shaking at 1400 rpm in a thermomixer.
9. Quench the reactions by addition of 2 μL of EGTA solution and immediately place the tubes on ice.

1. Pre-chill the SW41 rotor and buckets to 4 °C.

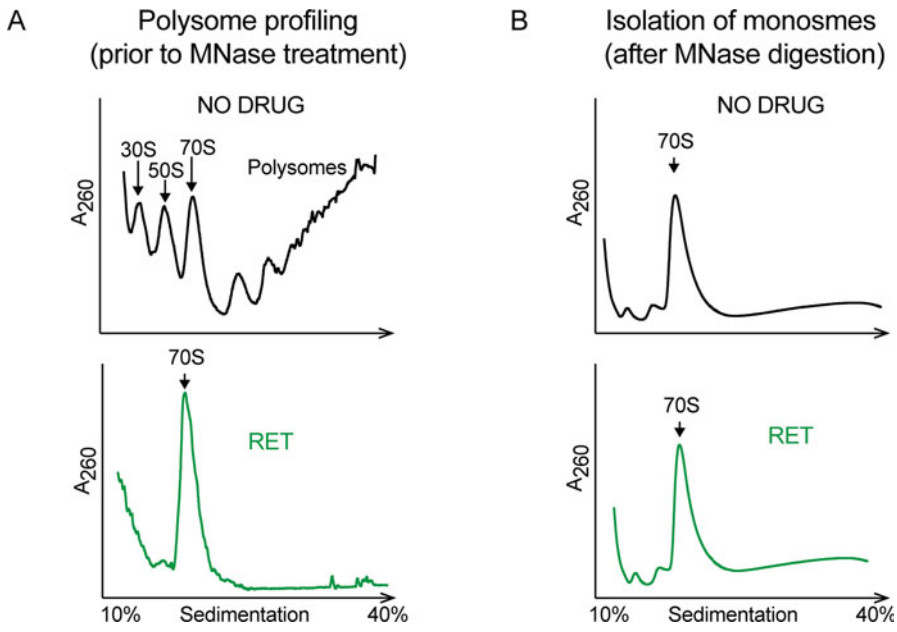

Fig. 4 (**a**) Sucrose gradient fractionation of the lysates prior to MNase digestion. Note that polysomes are collapsed into 70S ribosomes in the RET sample. (**b**) Sucrose gradient fractionation of the samples following digestion with MNase. The treatment with MNase should be optimized to collapse polysomes (in the control sample) but preserve the integrity of the 70S peak

3.5 Isolation of Monosomes by Sucrose Gradient Fractionation

2. Prepare a 10–40% (weight/vol) sucrose gradients in centrifuge tubes for the SW41 rotor using a gradient maker.
3. Weigh the tubes to ensure proper balance. If further balancing is required, remove or add small volumes of the 10% sucrose solution.
4. Carefully pipette the MNase-treated lysates (~230 μL) and, if desired, MNase-free control samples (Fig. 4) (*see* **Note 9**) (*see* Subheading 3.4) on top of the gradients.
5. Centrifugate at 190,000 × *g* (39,000 rpm) in SW41 rotor for 2 h at 4 °C.
6. Fractionate the gradients using an automated fractionation system with continuous monitoring of A_{254} optical density.
7. Collect the fractions corresponding to the 70S monosome peak, whose final volume typically ranges between 1.8 and 2 mL.
8. Flash-freeze the monosome fractions in liquid nitrogen. Fractions can be stored at −80 °C or can be immediately processed as described in Subheading 3.6.

1. To be able to perform the extraction in 1.5 mL Eppendorf tubes, divide the ~2 mL monosome fractions (*see* Subheading 3.5) in aliquots of ~700 μL. In the following steps, we describe

3.6 Phenol-Chloroform Extraction of Ribosome-Protected mRNA Fragments

the extraction for one of the aliquots, but all ~700 μL aliquots from the collected ~2 mL samples should be processed in parallel.

2. Add 40 μL of 20% SDS to the ~700 μL aliquot of monosome fraction.
3. Add 700 μL of acidic phenol pre-warmed to 65 °C. Mix by brief vortexing and incubate at 65 °C for 5 min with shaking at 1400 rpm in a thermomixer. Chill on ice for 5 min.
4. Spin at 20,000 × *g* for 2 min at room temperature in a microfuge, and transfer the top aqueous phase to a fresh tube. Add 700 μL of room-temperature acidic phenol. Mix by brief vortexing and incubate with shaking at room temperature for 5 min.
5. Spin at 20,000 × *g* for 2 min at room temperature, and transfer the top aqueous phase to a fresh tube. Add 600 μL chloroform and mix by vortexing.
6. Spin at 20,000 × *g* for 1 min at room temperature, and transfer the top aqueous phase to a fresh tube. Add 75 μL of 3 M NaOAc, pH 5.5, and mix. Add 800 μL of 100% isopropanol and briefly vortex. Chill at −80 °C for 30 min.
7. Pellet the extracted RNA at 20,000 × *g* for 1 h at 4 °C in a microfuge. Wash pellet with 800 μL of ice-cold 80% ethanol. Air-dry pellet for 5 min.
8. Use a total of 20 μL of Tris solution to resuspend all RNA pellets originated from the same monosome fraction. Snap-freeze and store at −80 °C or proceed to Subheading 3.7.

3.7 Size Selection of Ribosome-Protected mRNA Fragments

1. To quantify the concentration of RNA fragments, dilute 1 μL of RNA (see Subheading 3.6) with 9 μL of Tris solution, and estimate the concentration in a Nanodrop spectrophotometer. Optical density of 1 A_{260} equals ~40 μg/mL of RNA.
2. Prepare RNA for the sizing electrophoresis by placing 28 μg RNA into a final volume of 10 μL of Tris solution. Add 10 μL of 2× TBE-Urea Sample Buffer.
3. Prepare three individual samples of control RNA oligos for electrophoresis by combining 2 μL of the 20 μM oligo solutions (*see* Subheading 2.7) with 3 μL of Tris solution and 5 μL of 2× TBE-Urea Sample Buffer (*see* **Note 11**).
4. Pre-run the 15% TBE-Urea gel for 1 h at 200 V.
5. Denature samples from **steps 2** and **3** by incubating them at 80 °C for 2 min, and immediately place them on ice.

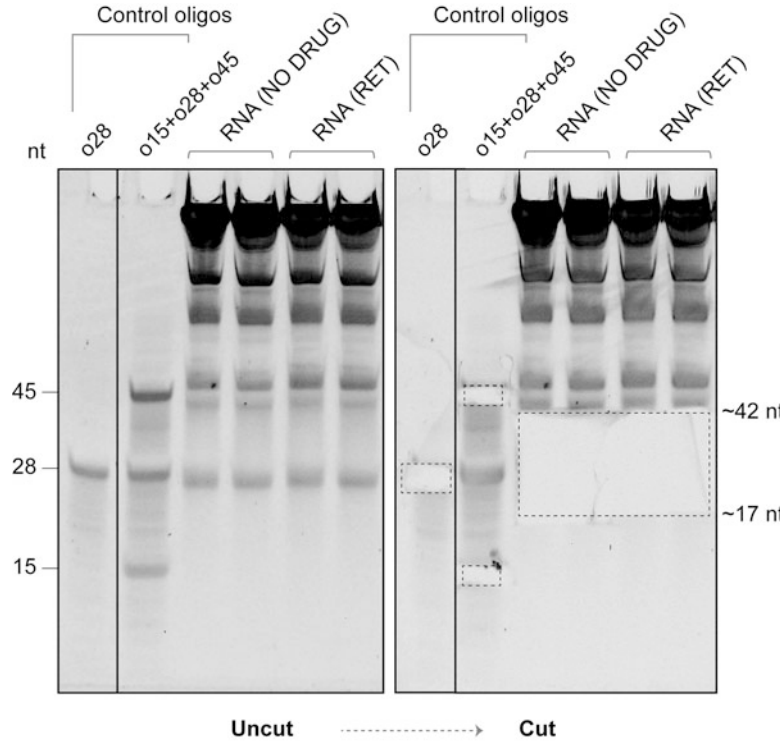

Fig. 5 Selection of mRNA fragments of the desired size ranges. The excised gel areas (right panel) are indicated by dashed rectangles

6. Wash the wells of the gel and load the RNA oligo controls (**step 3**) into individual lanes. To avoid overloading the lane, we recommend splitting the 20 μL RNA samples (**step 2**) into two 10 μL aliquots and loading them into two adjacent lanes. Run the gel at 200 V for 65 min. Stain the gel with 5 μL of SYBR Gold in 50 mL of 1× TBE for 5 min.
7. Excise desired bands whose sizes range between ~17–42 nt (Fig. 5). In parallel, excise the three control RNA oligos. Note that all the subsequent steps up to **step 13** of Subheading 3.12, described here for the ribosome-protected mRNA fragments, are carried out also with the control RNA oligos (*see* **Note 11**).
8. Recover the size-selected RNA fragments as follows: Place the gel slices in a 0.5 mL PCR tube whose bottom has been pierced with a 20 gauge needle. Nest the tube into a 2 mL collection tube. Spin at 20,000 × *g* for 2 min or until most of the gel has extruded into the collection tube.

9. Add to the crushed gel pieces 500 μL of RNA elution buffer supplemented with 2.5 μL SUPERase-In. Shake overnight in a thermomixer at 1400 rpm at 4 °C.
10. Spin samples for 10 sec in a microfuge, and transfer the gel suspension to a Spin-X column using a wide bore pipette tip. Spin at 20,000 × *g* for 3 min. Transfer eluate to a fresh tube.
11. Perform a second elution step by adding 200 μL of RNA elution buffer preheated to 70 °C to the Spin-X column. Shake in a thermomixer for 5 min at 1400 rpm at 70 °C. Spin the column for 3 min at 20,000 × *g* and combine the eluate with that obtained in **step 10**.
12. Add to the extracted RNA 2 μL of Glycoblue and briefly vortex. Add 750 μL of 100% isopropanol and vortex. Chill at −80 °C for 30 min.
13. Pellet RNA at 20,000 × *g* for 1 h at 4 °C in a microfuge. Aspirate supernatant. Wash the pellet by adding 800 μL of ice-cold 80% ethanol, spinning the tubes for 30 s at 20,000 × *g* in a microfuge, and aspirating the supernatants. Air-dry pellets for 5 min.
14. Resuspend the pellet in 15 μL of Tris solution.

3.8 Dephosphorylation of Ribosome-Protected mRNA Fragments

1. Prepare buffer master mix. For a single reaction:

10× T4 PNK buffer	2 μL
SUPERase-In	1 μL

2. Add 3 μL of buffer master mix to 15 μL of the RNA isolated in Subheading 3.7. Add 2 μL of T4 PNK to each tube. Incubate reaction at 37 °C for 1 h.
3. Heat inactive T4 PNK at 75 °C for 10 min.
4. Prepare precipitation buffer master mix. For a single reaction:

Tris solution	448 μL
3 M NaOAc, pH 5.5	50 μL
Glycoblue	2 μL

5. Add 500 μL of precipitation buffer master mix to each tube. Precipitate dephosphorylated RNA by adding 600 μL of 100% isopropanol and mix by vortexing. Chill at −80 °C for 30 min.
6. Pellet RNA at 20,000 × *g* for 1 h at 4 °C in a microfuge. Aspirate supernatant and wash pellet with 800 μL of ice-cold 80% ethanol as described in **steps 12** and **13** of Subheading 3.7. Air-dry pellet for 5 min.
7. Resuspend the RNA pellet in 5 μL of Tris solution.

3.9 Enzymatic Adenylation of the Linkers and Linker Ligation

Adenylation of the linkers (*see* **Note 12**):

1. For each linker (*see* Subheading 2.9), combine the following in a 0.2 mL PCR tube:

Linker (100 μM)	1.2 μL
5′ DNA adenylation 10× buffer	2 μL
ATP (1 mM)	2 μL
ddH_2O	12.8 μL
Mth RNA Ligase	2 μL

2. Incubate at 65 °C for 1 h.
3. Heat inactivate the Mth RNA Ligase at 85 °C for 5 min.
4. Add 30 μL of ddH_2O to the sample, and clean the sample using the Oligo Clean & Concentrator kit according to the manufacturer's instructions, except that the elution step is carried out with only 6 μL of ddH_2O. Store 5′-adenylated linkers at −20 °C or proceed immediately with the ligation steps.

Linker ligation:

5. Prepare the ligation reaction. For a single reaction:

50% PEG 8000	3.5 μL
T4 RNA ligase 10× buffer	1 μL
Pre-adenylated linker (20 μM) (**steps 1–4**)	0.5 μL
T4 RNA ligase	0.5 μL

6. Add 5 μL of dephosphorylated RNA (*see* Subheading 3.8) to the ligation reaction.
7. Incubate at 22 °C for 3 h.
8. Add 0.5 μL of 5'deadenylase and 0.5 μL of $RecJ_f$. Incubate at 30 °C for 45 min.
9. Clean the ligation reaction using Oligo Clean & Concentrator kit. Elute with 10 μL ddH_2O. Store at −80 °C or proceed to Subheading 3.10.

3.10 Reverse Transcription

1. Prepare the RT oligo master mix. For a single reaction:

10 mM dNTP	1 μL
25 μM NI-802	1 μL
ddH_2O	1.5 μL

2. Add 3.5 μL of RT oligo master mix to 10 μL of the ligated RNA (*see* Subheading 3.9). Denature at 65 °C for 5 min and chill it on ice for 5 min.
3. Prepare the RT buffer master mix. For a single reaction:

5× FSB buffer	4 μL
0.1 M DTT	1 μL
SUPERase-In	1 μL

4. Add 6 μL of the buffer master mix to the RNA from **step 2**. Add 1 μL of RT.
5. Incubate at 55 °C for 30 min.
6. Quench the reaction by hydrolyzing RNA templates: Add 2.3 μL of NaOH solution and incubate at 95 °C for 15 min. Note that this solution will turn pink in color.
7. Add 27.2 μL of ddH_2O to bring the reaction volume to 50 μL. Recover cDNA using Oligo & Clean Concentrator kit. Elute the purified cDNA in 8 μL of ddH_2O. Store at −80 °C or proceed with gel electrophoresis.
8. Pre-run the 10% TBE-Urea gel in 1× TBE buffer for 1 h at 200 V.
9. Prepare for electrophoresis the RT primer control:

NI-802 primer diluted to 1.25 μM	2 μL
Tris solution	6 μL
2× TBE-Urea Sample Buffer	8 μL

10. Prepare the cDNA samples from **step 7** for electrophoresis (as well as the cDNA samples generated using the processed control RNA oligos o15, o28, and o45; *see* **Note 11**) by adding 8 μL of 2× TBE-Urea Sample Buffer.
11. Denature the samples at 80 °C for 2 min and chill them on ice. Load samples onto the wells of the pre-run gel, and run electrophoresis at 200 V for 70 min (during this time the bromophenol blue dye may run out from the gel).
12. Stain the gel with 5 μL of SYBR Gold in 50 mL of 1× TBE for 5 min, and excise the desired bands (Fig. 6). In parallel, from

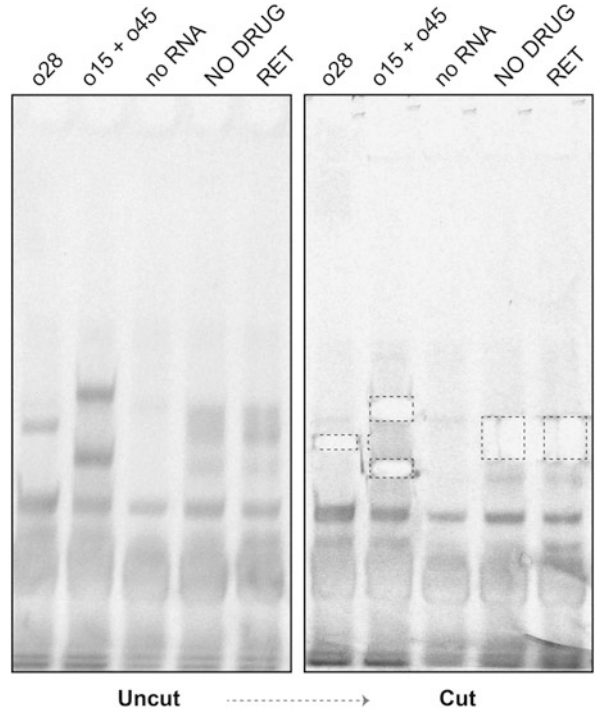

Fig. 6 Selection of cDNA products after reverse transcription. The cDNA products originated from control RNA oligos o15, o28, and o45 were used as size markers. The excised gel areas are indicated by dashed rectangles

the control RNA oligos, we recommend excising only the cDNA band corresponding to o15 (*see* **Note 11**).

13. Recover the size-selected cDNA products as follows: Place the gel slices in a 0.5 mL tube whose bottom has been pierced with a 20 gauge needle. Nest the tube into a 2 mL tube. Spin at 20,000 × *g* for 2 min or until most of the gel has extruded into the collection tube.
14. Add 500 μL DNA elution buffer. Shake overnight in a thermomixer at 1000 rpm at 25 °C.
15. Briefly spin samples and transfer the gel suspension to a Spin-X column using a wide bore pipette tip. Spin at 20,000 × *g* for 3 min. Transfer the eluate to a fresh tube and keep on ice.
16. Perform a second elution step by adding to the Spin-X column 200 μL of DNA elution buffer preheated to 70 °C. Shake Spin-X column in a thermomixer for 5 min at 1400 rpm at 70 °C. Spin at 20,000 × *g* for 3 min and combine the eluate with that obtained in **step 15**.

17. Add to the extracted cDNA 2 μL of Glycoblue and 750 μL of 100% isopropanol and vortex. Chill at −80 °C for 30 min.
18. Pellet cDNA at 20,000 × *g* for 1 h at 4 °C in microfuge. Aspirate supernatant and wash the pellet with 800 μL of ice-cold 80% ethanol as described in **step 13** of Subheading 3.7. Air-dry cDNA pellet for 5 min.
19. Resuspend the pellet in 15 μL of Tris solution.

3.11 Circularization of cDNA

1. Prepare circularization master mix. For one reaction:

10× CircLigase buffer	2 μL
1 mM ATP	1 μL
50 mM $MnCl_2$	1 μL

2. Add 4 μL of circularization master mix to 15 μL of cDNA samples (*see* Subheading 3.10). Add 1 μL of CircLigase.
3. Incubate at 60 °C for 1 h.
4. Heat inactivate CircLigase by incubating the reaction at 80 °C for 10 min. Chill on ice.
5. Clean the circularized cDNA (*see* **Note 13**) by adding 500 μL DNA elution buffer, 2 μL of Glycoblue, and 600 μL of isopropanol. Incubate at 80 °C for 30 min. Spin at 20,000 × *g* for 1 h at 4 °C in a microfuge. Aspirate supernatant and wash the pellet with ice-cold 80% ethanol as described in **step 14** of Subheading 3.7. Air-dry the pellets for 5 min.
6. Resuspend the pellet in 10 μL of Tris solution.

3.12 PCR Amplification

Pilot PCR:

1. Prepare the PCR amplification mix. For the following single reaction, we have used the primer pair AM-i51/AM-i71 (also *see* Fig. 2) as an example (*see* **Notes 14–15**):

5× HF buffer	16.7 μL
10 mM dNTP	1.7 μL
100 μM AM-i51 primer	0.4 μL
100 μM AM-i71 primer	0.4 μL
ddH_2O	58.8 μL
HF Phusion	0.8 μL

2. Add 4.5 μL of Tris solution to 0.5 μL of circularized cDNA (*see* Subheading 3.11). Add 79.2 μL of PCR amplification mix and mix by vortexing.
3. Aliquot 17 μL of the PCR mix into four separate PCR tubes.
4. Set up 12 cycles of the following PCR program:

Initial denaturation	30 s at 98 °C
Denaturation	10 s at 98 °C
Annealing	10 s at 65 °C
Extension	5 s at 72 °C

5. Sequentially remove individual PCR tubes after 6, 8, 10, or 12 cycles and place on ice.
6. Add 3.5 μL of 6× DNA loading dye to each tube.
7. Prepare 10 bp DNA ladder:

10 bp ladder	1 μL
Tris solution	9 μL
6× DNA loading dye	2 μL

8. Set up an 8% TBE gel in 1× TBE.
9. Load samples and run for 55 min at 180 V.
10. Stain the gel with 5 μL of SYBR Gold in 50 mL of 1× TBE for 5 min and visualize.

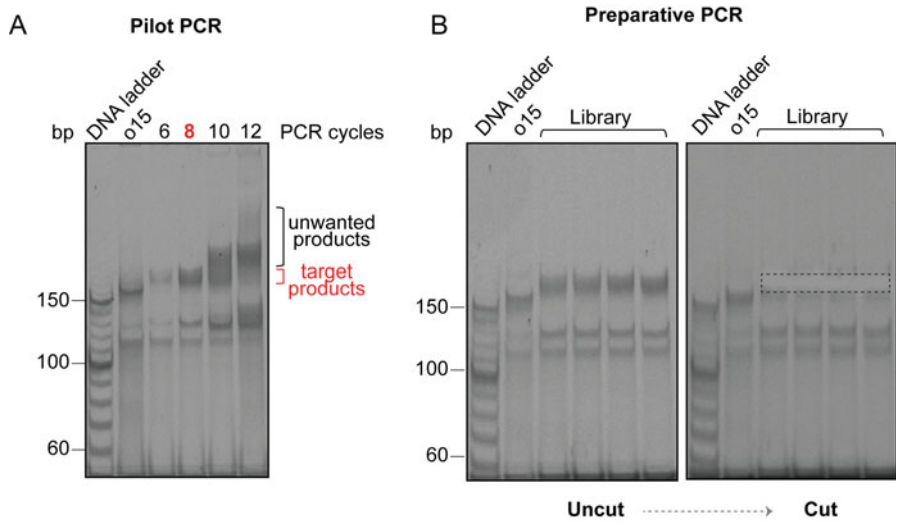

Fig. 7 (**a**) Pilot PCR analysis for determining the optimal number of cycles required for generation of the PCR library. In this example, the eight-cycle amplification (indicated in red) produced the desired result because it yielded a sufficient amount of the target products but lacked the undesirable higher molecular weight fragments appearing with increased number (10 and 12) of cycles. (**b**) Preparative gel for isolation of the PCR fragments obtained after eight amplification cycles. The excised gel area is indicated with a dashed rectangle

11. Identify the number of PCR cycles that results in high yield of the DNA fragments with the target size range of 150–170 bp but the lack of high molecular weight products (Fig. 7).
12. Preparative PCR: Set up reactions for preparative PCR exactly as described in **steps 1–3**. Run the PCR reaction for the optimal number of cycles determined in **step 11**.
13. Run the samples in an 8% TBE gel as described in **steps 6**, **9** and **10**, along with the 10 bp DNA ladder (**step 7**).
14. Excise the desired bands (Fig. 7), and recover the double-stranded DNA products as described in **steps 13–18** of Subheading 3.10.
15. Resuspend the pellet in 10 μL of Tris solution.

3.13 Preparing Samples for Next Generation Sequencing

1. Quantify samples using 1 μL of the prepared DNA (*see* Subheading 3.12) mixed with 2 μL of ddH_2O using Agilent TapeStation system or an equivalent platform.
2. Depending on the number of samples and sequencing platform to be used, samples can be sequenced individually or combined into a single sample. The latter is possible if different pairs of PCR primers listed in Subheading 2.12 were used for each sample in Subheading 3.12. We routinely combine 4–8 samples in the same tube so that the final amount of DNA is 10 nmol. This is enough for a single lane on Illumina HiSeq 4000 or NextSeq platforms (*see* **Note 15**). We usually request an SR75 or SR100 runs (single-end reading of 75 or 100 bases, respectively). In some instances, we have obtained sufficiently good results with SR50 runs (single-end 50 bases sequencing). We usually target to obtain 40–60 million raw reads per sample.

3.14 Computational Processing of Ribosome Profiling Reads

The data processing and analysis is performed using custom scripts and publicly available software packages available in the Galaxy platform [47]. Main steps of the recommended data processing flow are as follows:

1. Remove the adapter sequence from the raw sequencing reads using Cutadapt algorithm [48].
2. Align the processed reads to ribosomal and other non-coding RNA sequences using Bowtie algorithm [49] (parameters: `-n 1 -l 20 -m 1`), and discard them. Align remaining reads to genome using the same parameters.
3. Assign the 15th nucleotides upstream of the 3′ end of each read as the first nucleotide of the P-site codon. Divide the number of reads assigned to each genomic position by the total number of mapped reads divided by 1,000,000. This normalization results in reads per million (RPM) value for every nucleotide and can

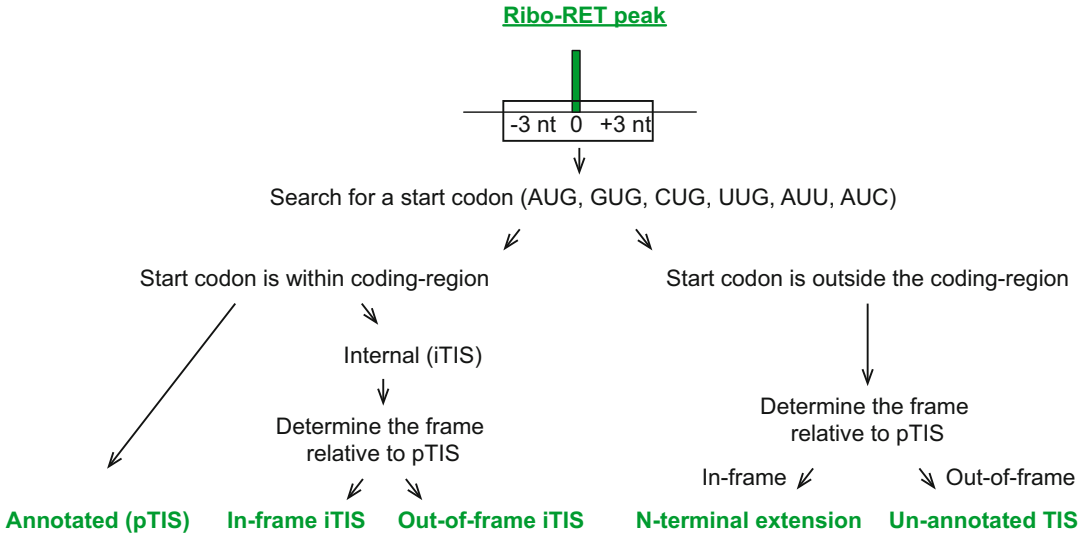

Fig. 8 Schematics of the computational algorithm to find genome-wide start sites using Ribo-RET peaks

be assembled in a wiggle track format (WIG file). Wig files are used to visualize the Ribo-RET peaks in a genome browser (e.g., MochiView [50]) and to analyze genome-wide start sites using custom scripts (*see* **Note 16**).

4. Identify potential start sites by searching for a start codon (AUG, GUG, CUG, UUG, AUU, AUC) within three nucleotides upstream or downstream of the Ribo-RET peaks. For annotated start sites, we define "Ribo-RET peak" as any read in the Ribo-RET data that has a density of >1 RPM. However, for different analysis, such as alternative TIS searching, we suggest using more stringent cut-off for Ribo-RET peak definition (*see* **step 5**).
5. If the start codon is within the annotated coding regions, we classify the corresponding site as in-frame or out-of-frame iTIS. If TIS identifies a start codon upstream of the genes, we classify it as "N-terminal extension" or "upstream TIS" if no stop codon is present between such TIS and the TIS of the annotated ORF. For the Ribo-RET peaks that are outside of the annotated ORFs and are out of frame relative to the first downstream ORF or that are in frame with the downstream ORF but a stop codon is present prior to the annotated start site, we consider them as potential TISs of the unannotated novel ORFs (Fig. 8). For pTIS classification, we only consider the peaks with minimum of 1 RPM. For alternative start sites, however, we use more stringent cut-off (RPM > 5) (*see* **Note 17**).

4 Notes

1. Proper practices for handling of radioactive material and waste should be followed for the procedures described in Subheading 3.1, **steps 3–11**.
2. We recommend to not autoclave or filter sterilize the complete MOPS medium as these procedures may deplete it from specific components [51].
3. Because of the heterogeneity in activity of different lots of MNase, we recommend to carry out the following steps prior to performing the Ribo-RET experiments: (i) pool together several different lots of the enzyme; (ii) prepare a mock lysate from cells not treated with RET by following the procedure outlined in Subheadings 3.2–3.4; carry out the MNase treatment (**steps 9** and **10** of Subheading 3.4) with different number of units of the enzyme (include an aliquot of lysate not treated with MNase); (iii) assess the MNase-mediated disruption of polysomes by the fractionation procedure described in Subheading 3.5. Choose the amount of MNase that converts polysomes into monosomes without affecting the 70S monosome peak. Prepare, aliquot, and store the MNase stock solution (**step 3** of Subheading 2.4) according to this optimization procedure.
4. As an alternative to radioactive metabolic labeling, cell-permeable methionine analogs such as 4-azido-L-homoalanine (L-AHA) or L-homopropargylglycine (HPG) can be used [52, 53].
5. It is recommended to first optimize the concentration of RET by exposing cells for the same amount of time (e.g., 2 min) to different concentrations of the drug and then carry out the time-course experiment described in Subheading 3.1 at the chosen concentration of RET. If optimization of RET treatment by metabolic labeling cannot be performed, exposure of cells to 100× MIC_{RET} for 5 min will likely afford a nearly 100% inhibition of protein synthesis [30]. Longer exposure times of bacterial cells to RET are not recommended as they can lead to undesirable secondary effects [54, 55].
6. For the procedure described here, we found that 150 mL of an early exponential *E. coli* BL21 culture is the optimal volume to be filtered (*see* **Note 7**). The optimal conditions for the collection/filtering steps may vary between strains or growth characteristics of the cultures.
7. It is important to keep the filtering time of the cells as brief as possible in order to preserve the position of ribosomes on mRNA [51]. For shortening the filtering time, (i) pay attention

that the shiny side of the filter is facing up; (ii) swirl the filtering flask while the culture is going through the filter disc; (iii) after completion of filtration, the filter can be transferred quickly onto a chilled glass plate to more rapidly and evenly scrape off the cell pellet (*see* also **Note 8**).

8. We highly recommend pre-optimizing the filtering step using a mock culture. Strains of some bacterial species may not be filterable, and thus other collection methods (such as centrifugation) may be considered. Alternatively, with the availability of the appropriate mixer mill equipment, the cell collection conditions could be adjusted for the filtering step to be skipped altogether [51].
9. Comparing MNase-untreated samples from the control and RET-treated cultures shows whether the optimized RET treatment (*see* Subheading 3.1) resulted in the expected enrichment of monosomes and concomitant depletion of polysomes (Fig. 4). Furthermore, including MNase-untreated samples of the control cell culture helps to evaluate the completeness of conversion of polysomes to monosomes following MNase treatment.
10. The remaining lysates can be stored at −80 °C. If desired, an aliquot of the lysates can be saved separately for RNA-seq.
11. We found it beneficial to separately process each of the RNA control oligos o15, o28, and o45 (*see* Subheading 2.7) in parallel to the experimental samples throughout the procedures described in Subheadings 3.7 to 3.10 (**step 12**). Processing the oligos separately allows for loading them in different lanes in the size-selection gels (Fig. 5 and Fig. 6) which facilitates a more accurate selection of the desired size range of the bands in the experimental samples. Starting from **step 13** of Subheading 3.10, through all the steps in Subheading 3.11, and up to **step 13** of Subheading 3.12, we process only the control oligo o15, as it is the most helpful marker to evaluate the size of PCR amplification products (*see* Subheading 3.11).
12. If the number of samples is less than six, the same linker can be ligated to RNA from all the samples because multiplexing can be achieved by using different pairs of PCR primers for each sample (*see* Subheading 3.12). If the number of samples exceeds six, it is beneficial to use different ligation linkers, each contributing its own individual barcode (*see* Subheading 2.9).
13. Cleaning circularized cDNA is optional, but we routinely carry out this step in order to improve reproducibility of the subsequent steps.
14. Using different PCR primer pairs (**item 3** of Subheading 2.12) for each sample allows for combining them together for

sequencing the pooled samples a single lane of the Illumina sequencing platforms (also *see* **Note 15**).

15. Inform the sequencing facility that your library is double-barcoded, and indicate the Illumina barcodes that were present in the primers used in Subheading 3.12.
16. We recommend confirming global RET-induced ribosome arrest at start codons by generating a metagene plot of averaged Ribo-RET occupancy in the vicinity of all start codons.
17. We observed that the absolute height (RPM value) of the RET peaks at the TISs of the annotated genes does not correlate with the expression of a gene in no-drug conditions and may vary between experiments. Therefore, when comparing RET peaks in different conditions, data from replicates should be diligently analyzed to ensure reproducibility. For example, for our alternative TIS assignment in the *E. coli* genome [30], we only considered the sites that were common in two different strains of *E. coli*. Depending on the research question and the bacterial strain of interest, different thresholds and adjustments to this computational analysis should be considered.

Acknowledgments

The work in Mankin/Vazquez-Laslop laboratory is supported by grants from the National Institutes of Health (R35 GM127134, to A.S.M.) and National Science Foundation (MCB 1951405, to N. V.-L.).

References

1. Laursen BS, Sorensen HP, Mortensen KK, Sperling-Petersen HU (2005) Initiation of protein synthesis in bacteria. Microbiol Mol Biol Rev 69:101–123
2. Ramakrishnan V (2002) Ribosome structure and the mechanism of translation. Cell 108:557–572
3. Rodnina MV (2018) Translation in prokaryotes. Cold Spring Harb Perspect Biol 10(9): a032664. https://doi.org/10.1101/cshperspect.a032664
4. Villegas A, Kropinski AM (2008) An analysis of initiation codon utilization in the Domain Bacteria - concerns about the quality of bacterial genome annotation. Microbiology 154:2559–2661
5. Shine J, Dalgarno L (1975) Determinant of cistron specificity in bacterial ribosomes. Nature 254:34–38
6. Gualerzi CO, Pon CL (2015) Initiation of mRNA translation in bacteria: structural and dynamic aspects. Cell Mol Life Sci 72 (22):4341–4367
7. Moll I, Grill S, Gualerzi CO, Blasi U (2002) Leaderless mRNAs in bacteria: surprises in ribosomal recruitment and translational control. Mol Microbiol 43:239–246
8. Nakagawa S, Niimura Y, Miura K, Gojobori T (2010) Dynamic evolution of translation initiation mechanisms in prokaryotes. Proc Natl Acad Sci U S A 107:6382–6387
9. Baez WD, Roy B, McNutt ZA, Shatoff EA, Chen S, Bundschuh R, Fredrick K (2019) Global analysis of protein synthesis in Flavobacterium johnsoniae reveals the use of Kozak-like sequences in diverse bacteria. Nucleic Acids Res 47:10477–10488
10. Li GW, Burkhardt D, Gross C, Weissman JS (2014) Quantifying absolute protein synthesis

rates reveals principles underlying allocation of cellular resources. Cell 157:624–635

11. Woolstenhulme CJ, Guydosh NR, Green R, Buskirk AR (2015) High-precision analysis of translational pausing by ribosome profiling in bacteria lacking EFP. Cell Rep 11:13–21
12. de Smit MH, van Duin J (1990) Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. Proc Natl Acad Sci USA 87:7668–7672
13. Espah Borujeni A, Channarasappa AS, Salis HM (2014) Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. Nucleic Acids Res 42:2646–2659
14. Kozak M (2005) Regulation of translation via mRNA structure in prokaryotes and eukaryotes. Gene 361:13–37
15. Vimberg V, Tats A, Remm M, Tenson T (2007) Translation initiation region sequence preferences in *Escherichia coli*. BMC Mol Biol 8:100
16. Osterman IA, Evfratov SA, Sergiev PV, Dontsova OA (2013) Comparison of mRNA features affecting translation initiation and reinitiation. Nucleic Acids Res 41:474–486
17. Bienvenut WV, Giglione C, Meinnel T (2015) Proteome-wide analysis of the amino terminal status of Escherichia coli proteins at the steady-state and upon deformylation inhibition. Proteomics 15:2503–2518
18. Impens F, Rolhion N, Radoshevich L, Becavin C, Duval M, Mellin J, Garcia Del Portillo F, Pucciarelli MG, Williams AH, Cossart P (2017) N-terminomics identifies Prli42 as a membrane miniprotein conserved in Firmicutes and critical for stressosome activation in Listeria monocytogenes. Nat Microbiol 2:17005
19. Yuan P, D'Lima NG, Slavoff SA (2018) Comparative membrane proteomics reveals a non-annotated E. coli heat shock protein. Biochemistry 57:56–60
20. Gao T, Yang Z, Wang Y, Jing L (2010) Identifying translation initiation sites in prokaryotes using support vector machine. J Theor Biol 262:644–649
21. Makita Y, de Hoon MJ, Danchin A (2007) Hon-yaku: a biology-driven Bayesian methodology for identifying translation initiation sites in prokaryotes. BMC Bioinformatics 8:47
22. Salzberg SL, Delcher AL, Kasif S, White O (1998) Microbial gene identification using interpolated Markov models. Nucleic Acids Res 26:544–548
23. Lomsadze A, Gemayel K, Tang S, Borodovsky M (2018) Modeling leaderless transcription and atypical genes results in more accurate gene prediction in prokaryotes. Genome Res 28:1079–1089
24. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2017) GenBank. Nucleic Acids Res 45(D1): D37–D42
25. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119
26. Delcher AL, Bratke KA, Powers EC, Salzberg SL (2007) Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics 23:673–679
27. Weaver J, Mohammad F, Buskirk AR, Storz G (2019) Identifying small proteins by ribosome profiling with stalled initiation complexes. MBio 10(2):e02819
28. Baek J, Lee J, Yoon K, Lee H (2017) Identification of unannotated small genes in salmonella. G3 7:983–989
29. Meydan S, Vázquez-Laslop N, Mankin AS (2019) Genes within genes in bacterial genomes. In: Storz G, Papenfort K (eds) Regulating with RNA in bacteria and archaea. ASM Press, Washington, DC, pp 133–154
30. Meydan S, Marks J, Klepacki D, Sharma V, Baranov PV, Firth AE, Margus T, Kefi A, Vazquez-Laslop N, Mankin AS (2019) Retapamulin-assisted ribosome profiling reveals the alternative bacterial proteome. Mol Cell 74:481–493
31. 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
32. Ndah E, Jonckheere V, Giess A, Valen E, Menschaert G, Van Damme P (2017) REPARATION: ribosome profiling assisted (re-) annotation of bacterial genomes. Nucleic Acids Res 45:e168
33. Giess A, Jonckheere V, Ndah E, Chyzynska K, Van Damme P, Valen E (2017) Ribosome signatures aid bacterial translation initiation site identification. BMC Biol 15:76
34. Schrader JM, Zhou B, Li GW, Lasker K, Childers WS, Williams B, Long T, Crosson S, McAdams HH, Weissman JS, Shapiro L (2014) The coding and noncoding architecture of the Caulobacter crescentus genome. PLoS Genet 10: e1004463

35. Gao X, Wan J, Liu B, Ma M, Shen B, Qian SB (2015) Quantitative profiling of initiating ribosomes in vivo. Nat Methods 12:147–153
36. Lee S, Liu B, Lee S, 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
37. 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
38. 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
39. Nakahigashi K, Takai Y, Kimura M, Abe N, Nakayashiki T, Shiwa Y, Yoshikawa H, Wanner BL, Ishihama Y, Mori H (2016) Comprehensive identification of translation start sites by tetracycline-inhibited ribosome profiling. DNA Res 23:193–201
40. Davidovich C, Bashan A, Auerbach-Nevo T, Yaggie RD, Gontarek RR, Yonath A (2007) Induced-fit tightens pleuromutilins binding to ribosomes and remote interactions enable their selectivity. Proc Natl Acad Sci U S A 104:4291–4296
41. Poulsen SM, Karlsson M, Johansson LB, Vester B (2001) The pleuromutilin drugs tiamulin and valnemulin bind to the RNA at the peptidyl transferase centre on the ribosome. Mol Microbiol 41:1091–1099
42. Schlunzen F, Pyetan E, Fucini P, Yonath A, Harms JM (2004) Inhibition of peptide bond formation by pleuromutilins: the structure of the 50S ribosomal subunit from *Deinococcus radiodurans* in complex with tiamulin. Mol Microbiol 54:1287–1294
43. Paukner S, Riedl R (2017) Pleuromutilins: potent drugs for resistant bugs-mode of action and resistance. Cold Spring Harb Perspect Med 7:a027110
44. Paukner S, Gelone SP, Arends SJR, Flamm RK, Sader HS (2019) Antibacterial activity of lefamulin against pathogens most commonly causing community-acquired bacterial pneumonia: SENTRY antimicrobial surveillance program (2015-2016). Antimicrob Agents Chemother 63:e02161–e02118
45. Oh E, Becker AH, Sandikci A, Huber D, Chaba R, Gloge F, Nichols RJ, Typas A, Gross CA, Kramer G, Weissman JS, Bukau B (2011) Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. Cell 147:1295–1308
46. McGlincy NJ, Ingolia NT (2017) Transcriptome-wide measurement of translation by ribosome profiling. Methods 126:112–129
47. Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Cech M, Chilton J, Clements D, Coraor N, Eberhard C, Gruning B, Guerler A, Hillman-Jackson J, Von Kuster G, Rasche E, Soranzo N, Turaga N, Taylor J, Nekrutenko A, Goecks J (2016) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res 44(W1): W3–W10
48. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17:10–12
49. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25
50. Homann OR, Johnson AD (2010) MochiView: versatile software for genome browsing and DNA motif analysis. BMC Biol 8:49
51. Mohammad F, Green R, Buskirk AR (2019) A systematically-revised ribosome profiling method for bacteria reveals pauses at single-codon resolution. elife 8:e42591
52. Dieterich DC, Link AJ, Graumann J, Tirrell DA, Schuman EM (2006) Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). Proc Natl Acad Sci U S A 103:9482–9487. https://doi.org/10.1073/pnas.0601637103
53. Oulhen N, Swartz SZ, Laird J, Mascaro A, Wessel GM (2017) Transient translational quiescence in primordial germ cells. Development 144:1201–1210
54. Felden B, Cattoir V (2018) Bacterial adaptation to antibiotics through regulatory RNAs. Antimicrob Agents Chemother 62: e02503–e02517
55. Steinchen W, Bange G (2016) The magic dance of the alarmones (p)ppGpp. Mol Microbiol 101:531–544

Chapter 3

Genome-Wide Ribosome Profiling of the *Plasmodium falciparum* Intraerythrocytic Developmental Cycle

Florence Caro and Joseph L. DeRisi

Abstract

Monitoring whole-genome translation and mRNA ribosome occupancy in vivo using ribosome profiling has proven to be a powerful tool for discovery of gene expression regulation, mechanisms of translation, and new open reading frames, in a wide range of different cell types in different organisms. Here we describe its application to the malaria parasite, *Plasmodium falciparum*. We present methods for intact polysome purification from parasite cultures, polysome digestion, monosome purification, ribosome footprint nucleic acid extraction, and Illumina library preparation.

Key words Ribosome profiling, Translation, Genome-wide, Malaria, *Plasmodium falciparum*, Deep sequencing

1 Introduction

Upon invasion of human red blood cells (RBCs), the malaria parasite *Plasmodium falciparum*, grows, divides, and ultimately ruptures the cell to propagate into other uninfected cells. It is during this intraerythrocytic developmental cycle (IDC) that all clinical manifestations of the disease occur, making it an important target for therapeutic intervention. Measuring the IDC transcriptome revealed that any given gene is expressed only once per cycle and "just in time" to exert its encoded function [1]. While the transcriptional machinery is conserved, only a single family of 27 specific transcription factors has been described to date [2]. This dearth of transcription factors suggests that regulation of expression of the roughly 5500 *P. falciparum* genes likely occurs at the epigenetic or posttranscriptional levels, for example, at the level of translation. Translation is the process by which a ribosome, guided by an mRNA template, synthesizes a protein. Timely and plastic regulation of translation allows cells to maintain or adjust their states under dynamically changing environments. Thus, determining which *P. falciparum* genes are translated and when adds an

Vyacheslav M. Labunskyy (ed.), *Ribosome Profiling: Methods and Protocols*, Methods in Molecular Biology, vol. 2252, https://doi.org/10.1007/978-1-0716-1150-0_3,

important resource for understanding the molecular physiology of this parasite during its life cycle transitions [3]. Ribosome profiling, the deep sequencing of ribosome-protected mRNA fragments (ribosome footprints), provides a powerful tool for the in vivo and genome-wide monitoring of translation. This method begins with the collection of a cell lysate containing polysomes, two or more ribosomes in complex with an mRNA, that are then subjected to nuclease treatment to digest the regions of mRNA that are not protected by the ribosome and converting them to monosomes (one ribosome protecting a ~30 nt of mRNA). Mock or nuclease-treated cell lysates are loaded onto sucrose gradients to separate monosomes and/or polysomes by ultracentrifugation according to their sedimentation coefficient. The quality of the sample can be assessed by flowing gradients past a detector of a continuous UV spectrophotometer to obtain a polysome profile. Polysomes readily detected in the untreated samples collapse into a single monosome peak in the nuclease-treated samples (Fig. 1). Finally, the ~30 nt ribosome footprints isolated from the sucrose gradient fractions containing the monosome peak are converted to a library for next-generation sequencing and mapped to the corresponding genome. The density of ribosome footprints on any given mRNA is a direct proxy for the rate of protein synthesis. Furthermore, the distribution of ribosome footprints reveals the identity of the translated product and thus can be used to annotate coding regions. In order to determine translational efficiency (the ratio of ribosome footprint density to mRNA abundance), total RNA is isolated in parallel to construct libraries for measuring mRNA abundance by mRNA-seq. This chapter describes the application of this technique to the *P. falciparum* IDC stages, including detailed procedures and notes.

2 Materials

2.1 Preparation of Parasite Lysates

1. Growth medium: Roswell Park Memorial Institute (RPMI) 1640 medium supplemented with 0.25% Albumax II, 2 g/L sodium bicarbonate, 0.1 mM hypoxanthine, 25 mM HEPES (pH 7.4), and 50 μg/L gentamicin.
2. Phosphate-buffered saline (PBS).
3. Hyperflasks hold large culture volumes (~500 mL) using a multilayer gas permeable growing surface for efficient gas exchange.
4. Microscope slides.
5. Giemsa stain.
6. 500 mL polypropylene centrifuge tubes.
7. 50 mL round bottom polypropylene tubes.

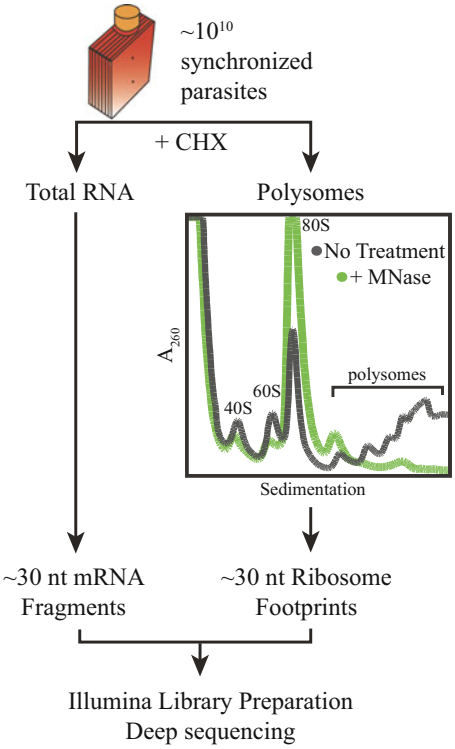

Fig. 1 Schematic of ribosome profiling of *P. falciparum*. Synchronized parasite cultures maintained in hyperflasks that hold a large volume of culture are treated with CHX before being processed for either total RNA or polysome extraction. Micrococcal nuclease (MNase)-treated or untreated parasite lysates are sedimented by sucrose gradient ultracentrifugation. A_{260} spectrophotometric monitoring of gradient fractions reveals polysomes in the untreated (gray) samples that have collapsed to a single monosome peak (80S, green) in the nuclease-treated sample. 40S and 60S denote the small and large ribosomal subunit peaks, respectively. Ribosome footprints (~30 nt) derived from the monosome peak or chemically fragmented polyA-purified mRNA (~30 nt) are used to build libraries for Illumina deep sequencing

8. Cycloheximide (CHX): Dissolve in dimethyl sulfoxide (DMSO) to 100 mg/mL and store at −80 °C.
9. Liquid nitrogen.
10. Lysis buffer: 0.2% saponin, 100 μg/mL CHX, in 1× PBS.
11. Resuspension buffer: 100 μg/mL CHX in 1× PBS.
12. 3× parasite lysis buffer (PLB): 45 mM KOAc, 30 mM Tris–HCl pH 7.4, 1.5 mM dithiothreitol (DTT), 1.5% Triton X-100, and CHX 100 μg/mL, in RNAse-free H_2O.
13. Liquid nitrogen.
14. 18-gauge needle.

15. Retsch MM400 Mixer mill with screw top grinding jars and stainless steel grinding balls.
16. Nonstick RNase-free microfuge tubes 2 mL.
17. Spatula.
18. Styrofoam box.

2.2 Micrococcal Nuclease Digestion

1. Nanodrop.
2. Micrococcal nuclease, 2,000,000 Units/mL.
3. 1× micrococcal nuclease buffer: 50 mM Tris–HCl, 5 mM $CaCl_2$.
4. Superase-In, 20 Units/μL.
5. Bovine serum albumin (BSA) 10 mg/mL (100×) stock.
6. Rotating platform.

2.3 Sucrose Gradients

1. Open top polyclear ultracentrifuge tubes 9/16 × 3 ½ inches.
2. Nonstick RNase-free microfuge tubes 2 mL.
3. Nonstick RNase-free microfuge tubes 1.5 mL.
4. BioComp Gradient Master Model EM-1 Econo that includes a gradient making station, a gradient fractionator, and a UV Monitor.
5. Gradient fraction collector.
6. Sucrose gradient buffer (SGB): 140 mM KOAc, 15 mM MgOAc, 20 mM Tris–HCl pH 7.4, 0.5 mM DTT, 20 Units/mL Superase-In, and 100 μg/mL CHX, in RNAse-free H_2O.
7. 10% sucrose: Add 40 mL of SGB to 5 g sucrose, mix well, and bring up to 50 mL.
8. 50% sucrose: Add 35 mL of SGB to 25 g sucrose, mix well, and bring up to 50 mL.
9. Ultracentrifuge.

2.4 RNA Isolation from Monosome Sucrose Gradient Fractions

1. Thermomixer.
2. Tabletop centrifuge.
3. Nonstick RNase-free microfuge tubes 1.5 mL.
4. RNAse-free H_2O.
5. 20% Sodium dodecyl sulfate (SDS).
6. Acid phenol: Chloroform 5:1 pH 4.5.
7. Chloroform.
8. 3 M NaOAc pH 5.5.
9. Isopropanol.
10. GlycoBlue coprecipitant.
11. Dry ice.

12. 70% Ethanol (EtOH).
13. 10 mM Tris–HCl pH 8.0.

2.5 Total RNA Isolation from iRBCs

1. Thermomixer.
2. Tabletop centrifuge.
3. Phase lock heavy gel 15 mL tubes.
4. Nonstick RNase-free microfuge tubes 1.5 mL.
5. TRIzol.
6. Acid phenol:chloroform 5:1 pH 4.5.
7. Chloroform.
8. 3 M NaOAc pH 5.5.
9. Isopropanol.
10. GlycoBlue coprecipitant.
11. Dry ice.
12. 70% EtOH.
13. 10 mM Tris–HCl pH 7.0.
14. 2× TBE-Urea sample loading buffer.
15. 15% TBE-Urea gel.
16. SYBR Gold.

2.6 polyA+ mRNA Purification

1. Thermomixer.
2. DynaMag magnet.
3. Nonstick RNase-free microfuge tubes 1.5 mL.
4. Dynabeads Oligo$(dT)_{25}$.
5. Superase-In, 20 U/uL.
6. RNase-free H_2O.
7. 2× binding buffer: 20 mM Tris–HCl pH 7.5, 1.0 M LiCl, 6.7 mM EDTA.
8. 1× binding buffer: dilute 2× binding buffer with RNAse-free H_2O.
9. Wash buffer B: 10 mM Tris–HCl pH 7.5, 0.15 M LiCl, 1 mM EDTA.
10. 10 mM Tris–HCl pH 8.0.

2.7 polyA+ mRNA Zn-mediated Fragmentation

1. Tabletop centrifuge.
2. Non-stick RNase-free microfuge tubes 1.5 mL.
3. 10× Zn-fragmentation buffer: 100 mM $ZnCl_2$ in 100 mM Tris–HCl pH 7.0.
4. RNAse-free H_2O.
5. 0.5 M EDTA.

6. Isopropanol.
7. GlycoBlue.
8. 3 M NaOAc pH 5.5.
9. 70% EtOH.
10. 10 mM Tris–HCl pH 8.0.
11. Dry ice.

2.8 Fragmented mRNA and Ribosome Footprint Size Selection

1. Novex 15% TBE-Urea Gels.
2. 1× Novex TBE Running Buffer.
3. 2× Novex TBE-Urea Sample Buffer.
4. Costar Centrifuge-X column.
5. RNAse-free H_2O.
6. SYBR Gold.
7. Razor blade.

2.9 Rapid Gel Extraction

1. Nonstick RNAse-free microfuge tube 0.5 mL.
2. Nonstick RNAse-free microfuge tube 1.5 mL.
3. 21-gauge needle.
4. Thermomixer.
5. Tabletop centrifuge.
6. RNAse-free H_2O.
7. Razor blade.
8. Costar Centrifuge-X column.
9. GlycoBlue.
10. 3 M NaOAc pH 5.5
11. Isopropanol.
12. 70% EtOH.
13. Dry ice.
14. 10 mM Tris–HCl pH 8.0.

2.10 3′ End Dephosphorylation

1. Thermomixer.
2. Tabletop centrifuge.
3. T4 polynucleotide kinase (PNK).
4. 1× T4 PNK Buffer: 70 mM Tris–HCl, 10 mM $MgCl_2$, and 5 mM DTT
5. RNAse-free H_2O.
6. Superase-In.
7. PNK reaction mix: 33 µL RNAse-free H_2O, 5 µL 10× PNK buffer, 1 µL Superase-In, and 1 µL PNK.

8. GlycoBlue.
9. 3 M NaOAc pH 5.5
10. Isopropanol.
11. 70% EtOH.
12. Dry ice.
13. 10 mM Tris–HCl pH 8.0.

2.11 Linker Ligation

1. Thermomixer.
2. 50% w/v polyethylene glycol (PEG) MW 8000: Mix 7 g RNAse-free PEG (MW 8000) with RNAse-free H_2O to a final 14 mL volume. Mix in a rotator at room temperature overnight. Following resuspension, purify and de-gas by forcing the solution through a sterile 0.2 μm cellulose nitrate filter (this is a slow process). Store solution in tightly capped 1 mL aliquots. PEG stock must be remade every month for maximum reaction efficiency.
3. T4 RNA ligase 2, truncated (NEB).
4. 10× T4 RNA ligase 2, truncated ligase buffer (NEB).
5. DMSO.
6. Superase-In, 20 U/uL.
7. Linker ligation reaction mix: 8 μL 50% PEG, 2 μL 10× RNA ligase buffer, 2 μL DMSO, 1 μL 20 U/μL Superase-In, and 1 μL T4 RNA ligase 2 truncated.
8. GlycoBlue.
9. 3 M NaOAc pH 5.5
10. Isopropanol.
11. 70% EtOH.
12. Dry ice.
13. 10 mM Tris–HCl pH 8.0.

2.12 Size Selection of the Linker-ligated Product

1. Thermomixer.
2. Razor blade.
3. Novex 10% TBE-Urea Gels.
4. 1× Novex TBE Running Buffer.
5. 2× Novex TBE-Urea Sample Buffer.
6. RNAse-free H_2O.
7. SYBR Gold.
8. 10 mM Tris–HCl pH 8.0.

2.13 Ribosomal RNA Subtraction

1. Thermomixer.
2. Tabletop centrifuge.

Table 1
Ribosomal RNA subtraction oligo sequences

Oligo	IDT oligo sequence	Target locus	Target sequence
asDNA1b	/5Biosg/TGTTACTTCTTTGTTA TAATTCCTT	MAL7_18S	AAGGAATTATAACAAAGAAG TAACA
asDNA2b	/5Biosg/CATATATAATTTCTCTTTTA CATTAG	MAL7_28Sa	CTAATGTAAAAGAGAAATTA TATATG
asDNA3b	/5Biosg/TGGTATCGGTAATCCGCTT TAGCG	PF08_tmp	CGCTAAAGCGGATTACCGA TACCA
asDNA4b	/5Biosg/CCGGTATTGTATG CAAAGTGG	MAL5_28S	CCACTTTGCATACAA TACCGG

3. DynaMag magnet.
4. Nonstick RNAse-free microfuge tube 1.5 mL.
5. Resuspend the HPLC-purified 5′ biotinylated DNA oligos listed in Table 1 in RNAse-free H_2O at 20 pmol/μL and store at −80 °C. These oligos include the standard IDT C6 linker between the biotin moiety and the DNA sequence.
6. Ultrapure 20× saline-sodium citrate (SSC).
7. Superase-In, 20 U/uL.
8. Dynabeads MyOne Streptavidin C1.
9. 2× Dynabeads MyOne C1 B&W buffer: 10 mM Tris–HCl pH 7.5, 1 mM EDTA, 2 M NaCl, and 0.01% Tween 20.
10. Dynabeads MyOne C1 Solution A: 0.1 M NaOH and 0.05 M NaCl.
11. Dynabeads MyOne C1 Solution B: 0.1 M NaCl.
12. RNAse-free H_2O.
13. GlycoBlue.
14. 3 M NaOAc pH 5.5.
15. Isopropanol.
16. 70% EtOH.
17. Dry ice.
18. 10 mM Tris–HCl pH 8.0.

2.14 Reverse Transcription

1. RNAse-free H_2O.
2. 1 M NaOH.
3. 1 M HCl.

4. To prepare 500 µL 80% AT dNTPs, mix 20 µL dATP 100 mM, 20 µL dTTP 100 mM, 5 µL dCTP 100 mM, 5 µL dGTP 100 mM, and 450 µL RNAse-free H_2O.
5. RT reaction mix: 4 µL 5× First Strand Buffer, 1 µL 80% AT dNTPs, 1 µL Superase-In, 1 µL 0.1 M DTT, and 1 µL Superscript III Reverse Transcriptase.
6. Razor blade.
7. Novex 10% TBE-Urea Gels.
8. 1× Novex TBE Running Buffer.
9. 2× Novex TBE-Urea Sample Buffer.
10. SYBR Gold.
11. 10 mM Tris–HCl pH 8.0.

2.15 Circularization

1. Thermomixer.
2. Circularization mix: 2 µL 10× CircLigase buffer, 1 µL 1 mM ATP, 1 µL 50 mM $MnCl_2$, and 1 µL CircLigase ssDNA Ligase.

2.16 PCR Amplification

1. PCR machine.
2. PCR tubes.
3. PCR master mix: 20 µL 5× Phusion HF buffer, 2 µL dNTPs 10 mM (50% AT), 5 µL 10 uM oNTI231, 5 µL 10 µM indexed primer (choose a different index primer to barcode each sample), 62 µL RNAse-free H_2O, and 1 µL Phusion polymerase.
4. Index primer sequences for oCJ196, oCJ197, oCJ198, oCJ199, index-1, index-23, index-2, and index-9 are listed in Table 2.
5. 6× DNA loading dye.
6. RNAse-free H_2O.
7. Novex 8% TBE Gels.
8. 1× Novex TBE Running Buffer.
9. SYBR Gold.
10. 10 mM Tris–HCl pH 8.0.
11. Razor blade.
12. Nonstick RNAse-free microfuge tube 0.5 mL.
13. Nonstick RNAse-free microfuge tube 1.5 mL.
14. 21-gauge needle.
15. Thermomixer.
16. Tabletop centrifuge.
17. DNA elution buffer: 0.3 M NaCl, 1 mM Tris–HCl pH 8.0, and 1 mM EDTA.
18. GlycoBlue.

Table 2
Oligo sequences

Oligo	Sequence	Type	Notes
Linker-1[a]	5′ /5rApp/CTGTAGGCACCATCAAT/3ddC/	DNA	/5rApp/T4 RNA Ligase ligates this pre-adenylated oligo to the 3′-OH of a second single-stranded sequence in the absence of ATP /3ddC/Dideoxycytidine (ddC) is a 3′ chain terminator that prevents 3′ extension by DNA polymerases
oGAB11[a]	5′AGU CAC UUA GCG AUG UAC ACU GAC UGU G/3Phos/	RNA	Use as 28-mer ligation control /3Phos/ 3′ phosphate
oNTI202	5′CGACAGGTTCAGAGTTCTACAGTCCGACGATC	DNA	Use as your sequencing primer on clustered flow cells
oNTI231	5′ CAAGCAGAAGACGGCATACGA	DNA	Forward PCR primer Use for product amplification from circularized DNA libraries
oCJ200[a]	/5Phos/GATCGTCGGACTGTAGAACTCTGAACCTGTCG/iSp18/CAAGCAGAAGACGGCATACGAGATATTGATGGTGCCTACAG	DNA	Primer for reverse transcription of linker-ligated samples /5Phos/ 5′ phosphate /iSp18/18-atom hexa-ethyleneglycol spacer
oCJ196	AATGATACGGCGACCACCGAGATCTACACGATCGGAAGAGCACACGTCTGAACTCCAGTCAC [agttcc]CGACAGGTTCAGAGTTC	DNA	Reverse indexed PCR primer

oCJ197	AATGATACGGCGACCACCGAGATCTACACGATCGGAAGAGCACACGTCTGAACTCCAGTCAC [cgatgt]CGACAGGTTCAGAGTTC	DNA	Reverse indexed PCR primer
oCJ198	AATGATACGGCGACCACCGAGATCTACACGATCGGAAGAGCACACGTCTGAACTCCAGTCAC [gctgaa]CGACAGGTTCAGAGTTC	DNA	Reverse indexed PCR primer
oCJ199	AATGATACGGCGACCACCGAGATCTACACGATCGGAAGAGCACACGTCTGAACTCCAGTCAC [ttaggc]CGACAGGTTCAGAGTTC	DNA	Reverse indexed PCR primer
Index-1	AATGATACGGCGACCACCGAGATCTACACGATCGGAAGAGCACACGTCTGAACTCCAGTCAC [atcacg]CGACAGGTTCAGAGTTC	DNA	Reverse indexed PCR primer
Index-23	AATGATACGGCGACCACCGAGATCTACACGATCGGAAGAGCACACGTCTGAACTCCAGTCAC [gagtgg]CGACAGGTTCAGAGTTC	DNA	Reverse indexed PCR primer
Index-2	AATGATACGGCGACCACCGAGATCTACACGATCGGAAGAGCACACGTCTGAACTCCAGTCAC [cgatgt]CGACAGGTTCAGAGTTC	DNA	Reverse indexed PCR primer
Index-9	AATGATACGGCGACCACCGAGATCTACACGATCGGAAGAGCACACGTCTGAACTCCAGTCAC [gatcag]CGACAGGTTCAGAGTTC	DNA	Reverse indexed PCR primer

[a]/modification/indicate Integrated DNA Technologies (IDT) oligo modification codes

19. 3 M NaOAc pH 5.5.
20. Isopropanol.
21. 70% EtOH.
22. Dry ice.
23. 10 mM Tris–HCl pH 8.0.

3 Methods

Because *P. falciparum* grows asynchronously under laboratory conditions, a culture of infected red blood cells (iRBCs) will contain a mixed population of parasites at every developmental stage. To capture the translation dynamics of a representative snapshot of the 48-h IDC, parasites have to first be synchronized, which is typically achieved using sorbitol treatment to differentially lyse late-stage parasites and enrich for early ring stage parasites [4]. In order to achieve the large number of synchronized parasites required for this protocol, large culture volumes must be subjected to multiple rounds of synchronization which takes weeks of round-the-clock care. The starting material used in this protocol are synchronous parasite cultures representing either early (rings), middle (early and late trophozoites), or late (schizonts) stages of parasite development. The first step involves releasing intact parasites from RBCs using saponin, an amphipathic glycoside that disrupts the RBC but not the parasite membrane. To freeze ribosomes in place during this and all subsequent steps, the translation elongation inhibitor cycloheximide (CHX) is added from the moment of culture harvest onward. Parasites frozen in liquid nitrogen are then lysed, and their contents loaded onto sucrose gradients for polysome fractionation and collection of ribosome footprints. The protocol requires roughly 12 h of continuous work before a stopping point is reached, from the moment of culture harvest to the collection of ribosome footprints. The RNA isolation, library preparation, sequencing, and data analysis steps require an additional 5–7 days.

3.1 Preparation of Parasite Lysates

1. Grow synchronized cultures in 500 mL of growth medium in hyperflasks at 37 °C, 5% O_2, and 5% CO_2, to a maximum 10–15% parasitemia at 5% hematocrit (HC). Harvest a minimum of ~10^{10} parasites on the day of the experiment (*see* **Note 1**).
2. Transfer the culture to 500 mL conicals. Use pre-warmed medium to rinse and transfer the remaining culture in the flask to as many conical tubes as necessary.
3. Centrifuge to collect iRBCs at 319 × *g* for 5 min at room temperature with no brake.

4. Prepare a Giemsa-stained thin blood smear of the harvested culture to count parasitemia (*see* **Note 2**).
5. Resuspend the culture in 500 mL of warm growth medium containing 100 μg/mL CHX. Gently mix the resuspended culture and incubate for 5 min at room temperature.
6. Centrifuge to collect iRBCs at 319 × *g* for 5 min at room temperature with no brake.
7. Aspirate the supernatant and pool packed culture volumes. Transfer 2 mL of packed culture volume to a 50 mL tube, flash freeze in liquid nitrogen, and store at −80 °C for subsequent total RNA extraction (*see* Total RNA extraction protocol) (*see* **Note 3**).
8. Prepare 50 mL of resuspension buffer ahead of time, leave 40 mL at room temperature, and place 10 mL on ice.
9. Transfer up to 20 mL packed culture volume to 50 mL polypropylene tubes. Use as many tubes as you need to transfer the whole packed culture volume.
10. Add 1 volume (20 mL) ice cold lysis buffer to each tube to bring the volume up to 40 mL (final 0.1% saponin). Mix gently by inversion five times and immediately place on ice. Work fast to allow saponin lysis to occur simultaneously in all tubes. Proceed immediately to the centrifugation step.
11. Centrifuge at 10,000 × *g* for 2 min at 4 °C and set brake at minimum level.
12. Carefully aspirate supernatant without disturbing the parasite pellet. Use ice cold resuspension buffer to resuspend and pool the parasite pellets. Do not dilute the parasite pellet past 2 mL. Work fast and keep samples always on ice. Do not freeze and proceed immediately to the next step (*see* **Note 4**).
13. Add 1 mL ice cold 3× PLB to 2 mL parasite suspension (*see* **Note 5**).
14. Fill a 50 mL tube with liquid nitrogen. Use a P1000 micropipette to slowly drip parasite suspension into the liquid nitrogen creating frozen droplets of cells. Keeping the frozen droplets small is important for fitting them inside the mixer mill's grinding jars (*see* **Note 6**).
15. Using an 18-gauge needle poke holes into the cap of the 50 mL tube. Cap the tube containing the frozen parasites and tilt to pour the liquid nitrogen out. Proceed immediately to the mixer mill step (*see* **Note 7**).
16. Place mixer mill grinding jars and balls into liquid nitrogen (let cool until boiling stops).
17. Take cooled items out of the liquid nitrogen and place a ball into bottom half of the jar.

18. Transfer the frozen parasite droplets into the bottom half of the jar and screw top on tightly. Cool jar in liquid nitrogen until boiling stops.
19. Loosen the grinding jar about one quarter turn, place it in the mixer mill, and grind parasites for 3 min at 15 Hz (*see* **Note 8**).
20. Tighten jar and chill immediately in liquid nitrogen along with a clean spatula. Cool until boiling stops.
21. Open the grinding chamber and recover the powder using the chilled metal spatulas, re-chilling them as needed. Transfer the powder to a conical tube placed inside a styrofoam box containing liquid nitrogen. Proceed immediately to the next step (*see* **Note 9**).
22. Thaw parasite powder on ice (*see* **Note 10**).
23. Transfer thawed lysate to as many 2 mL nonstick RNAse-free microfuge tubes as needed.
24. Centrifuge parasite lysate for 10 min at 16,000 × *g* at 4 °C.
25. Transfer the supernatant (~2 mL) to a 2 mL nonstick RNAse-free microfuge tube. Proceed immediately to the next step.

3.2 Micrococcal Nuclease Digestion

A nonspecific nuclease is used to digest the regions of mRNA not protected by the ribosome. Micrococcal nuclease is the RNAse of choice for *P. falciparum* since RNAse I was observed to over digest the sample, as manifested by degraded polysomes. From this step on, use RNAse-free reagents to protect the sample from degradation by contaminating nucleases.

1. Mix 1 μL of polysome extract with 9 μL PLB and determine RNA concentration of this tenfold dilution using a nanodrop.
2. Split the polysome extract into nuclease-treated (~1350 μL) and undigested control (~650 μL). Digest the polysome extract with 2.88 Units/μg (or 115 Units/OD) micrococcal nuclease in 1× micrococcal nuclease buffer and 1× BSA. Add 0.1 Unit/μg Superase-In to the untreated control.
3. Mix well and incubate on a rotating platform for 30 min at room temperature. Prepare sucrose gradients during this incubation step (*see* **Note 11**).

3.3 Sucrose Gradients

1. Prepare 12 mL sucrose density gradients in each of 6 ultracentrifuge tubes. For this fill each tube with ~6 mL of 50% sucrose solution and then gently add ~6 mL of 10% sucrose solution on top. Next use the Gradient Master preset program to mix the two sucrose layers and generate a 10–50% (wt/vol) sucrose gradient (*see* **Note 12**).
2. Gently load up to 350 μL of nuclease-treated or untreated polysome extract onto the top surface of the sucrose gradients (*see* **Note 13**).

3. Centrifuge at 35,000 rpm (151,263 × *g*) at 4 °C for 3 h.
4. Using an automated fractionation system, flow the gradient volume past a detector of a continuous UV spectrophotometer to monitor A_{254} (to obtain a polysome profile).
5. Collect gradient fractions of the nuclease treated samples in 1.5 mL nonstick RNase-free microfuge tubes. Note which tubes contain the monosome peak.

3.4 RNA Isolation from Monosome Sucrose Gradient Fractions

1. Prefill a 1.5 mL nonstick RNase-free microfuge tube with 1 volume acid phenol:chloroform 5:1 pH 4.5 (Phe:Chl) and pre-warm it to 65 °C using a Thermomixer.
2. To each sucrose gradient fraction containing the monosome peak, add RNAse-free H_2O to a final 600 μL volume.
3. To each sample, add 30 μL 20% SDS. Mix well and heat to 65 °C until SDS is dissolved.
4. Transfer the sample to the hot acid phenol and incubate 5 min at 65 °C mixing at 450 rpm.
5. Chill samples for 5 min on ice.
6. Centrifuge samples for 2 min at 16,000 × *g* at room temperature and immediately transfer the top, aqueous phase to a new 1.5 mL nonstick RNase-free microfuge tube.
7. Add 1 volume Phe:Chl and incubate 5 min at room temp mixing at 450 rpm.
8. Centrifuge samples for 2 min at 16,000 × *g* at room temperature and immediately transfer the top, aqueous phase to a new tube.
9. Add 1 volume chloroform and vortex 30 s at room temperature.
10. Centrifuge for 1 min at 16,000 × *g* and transfer the top, aqueous phase to a 1.5 mL non-stick RNAse-free microfuge tubes.
11. Add 0.1 vol 3 M NaOAc pH 5.5. Mix by flicking the tube.
12. Add 1.5 μL GlycoBlue and 1 volume isopropanol. Mix by flicking the tube.
13. Precipitate for at least 30 min on dry ice (or overnight at −80 °C).
14. Centrifuge for 30 min at 16,000 × *g* at 4 °C to pellet nucleic acids.
15. Visualize the pellet, carefully remove the supernatant, and add 800 μL 70% EtOH.
16. Centrifuge for 10 min at 16,000 × *g* at 4 °C.

17. Visualize the pellet and carefully remove the supernatant using a pipette. Air dry the pellet by leaving the tube open at room temperature for 5 min.
18. Resuspend in 20 µL 10 mM Tris–HCl pH 8.0 and store at −80 °C.

3.5 Total RNA Isolation from iRBCs

1. Preheat TRIzol to 65 °C and centrifuge 15 mL phase lock tubes (heavy gel) to pellet phase lock gel at 1500 × *g* for 2 min.
2. Take the iRBC pellet out of the −80 °C freezer (note that freeze-thawing causes iRBCs to lyse) and add 10 mL 65 °C Trizol on top and vortex for 5 s (*see* **Note 14**). Use a pipet-aid and a 10 mL pipette to resuspend the pellet until the solution is smooth (this can take up 10–20 min).
3. Add 2 mL chloroform and leave on ice for 5 min.
4. Centrifuge at 1500 × *g* for 10 min at 4 °C.
5. Transfer 4 mL aqueous phase to a 15 mL heavy gel phase lock tubes and add 4 mL acid phenol:chloroform 5:1 pH 4.3. Mix by inverting the tube and place immediately back on ice.
6. Centrifuge at 1500 × *g* for 10 min at 4 °C and decant up to 4 mL aqueous layer into a 15 mL heavy gel phase lock tube (pool samples if volumes allow).
7. Add 1 volume acid phenol:chloroform 5:1 pH 4.3, mix by inversion, and place tube immediately back on ice.
8. Centrifuge at 1500 × *g* for 10 min at 4 °C and decant aqueous layer into a 15 mL tube.
9. Add 1 volume chloroform.
10. Centrifuge at 1500 × *g* for 10 min at 4 °C and transfer aqueous layer to as many 2 mL nonstick RNase-free microfuge tubes as needed.
11. Add 1/10 volumes of 3 M NaOAc pH 5.5. Mix by flicking the tube.
12. Add 1 volume isopropanol and 2 µL GlycoBlue. Mix by flicking the tube.
13. Precipitate for at least 30 min on dry ice (or overnight at −80 °C).
14. Centrifuge at 16,000 × *g* for 30 min at 4 °C.
15. Visualize the pellet, carefully remove the supernatant, and add 1 mL 70% EtOH.
16. Centrifuge at 16,000 × *g* for 30 min at 4 °C.
17. Visualize the pellet and carefully remove the supernatant using a pipette. Air dry the pellet by leaving the tube open at room temperature for 5 min.
18. Resuspend in 20 µL 10 mM Tris–HCl pH 7.0 and store at −80 °C.

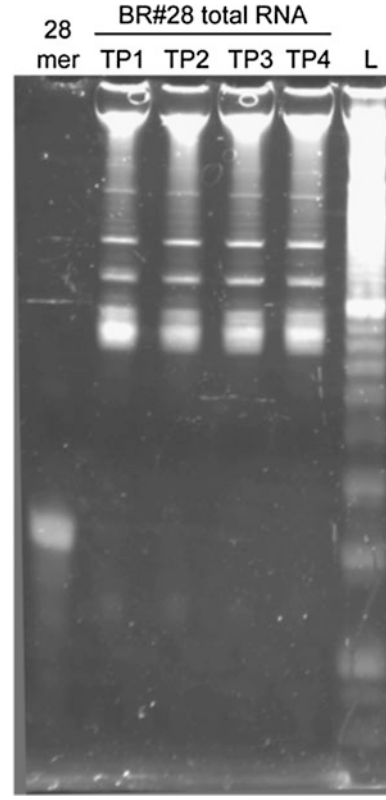

Fig. 2 *P. falciparum* total RNA quality assessment. Total RNA (0.5 μg) harvested from each parasite stage (Rings, TP1; early trophozoites, TP2; tate Trophozoites, TP3; schizonts, TP4), 10 bp DNA ladder (L), and the 28-mer sizing control, run on a 15% TBE-Urea gel for 65 min at 200 V. Discrete ribosomal RNA bands and no lower molecular weight smear are indicative of intact RNA

19. Run a gel to confirm that total RNA is not degraded (Fig. 2). For this mix, 0.5 μg of total RNA with 2× TBE-Urea sample loading buffer in a final volume of 20 μL.
20. Mix 0.5 μL 10 bp ladder with 9.5 μL RNAse-free H_2O and 10 μL 2× TBE-Urea denaturing loading buffer.
21. Denature all samples and controls at 75 °C for 2 min and place immediately on ice.
22. Pre-run a 10 well, 15% TBE-Urea, gel at 200 V for 10 min in 1× TBE running buffer and rinse urea from wells before loading samples.
23. Load gel with 20 μL sample making sure that every footprint sample is flanked by a 28-mer and 31-mer oligo sizing control.
24. Run gel at 200 V for 65 min.
25. Stain the gel in 60 mL 1× TBE running buffer plus 6 μL SYBR Gold for 5 min.

3.6 PolyA+ mRNA Purification

1. Resuspend Dynabeads Oligo$(dT)_{25}$ by vortexing and transfer 150 µL beads to a 1.5 mL nonstick RNase-free microfuge tube.
2. Collect beads by leaving tube on magnet for 30 s and then carefully remove the buffer in which they are suspended using a pipette. Immediately resuspend beads in 100 µL 1× binding buffer.
3. Repeat the wash in 100 µL 1× binding buffer and leave the beads on the magnet in this second binding buffer wash.
4. Take 150 µg total RNA and dilute to 50 µL with RNAse-free H_2O (*see* **Note 15**). Add 50 µL 2× binding buffer.
5. Denature RNA sample 2 min at 75 °C and then return to ice.
6. Add 1 µL Superase-In to each sample.
7. Remove binding buffer from beads and resuspend beads in RNA sample.
8. Incubate 5 min at room temp to bind RNA to beads.
9. Wash beads twice in 100 µL wash buffer B.
10. Ensure that all wash buffer is removed after the last wash and resuspend beads in 40 µL 10 mM Tris–HCl pH 8.0.
11. Elute RNA from beads by heating 2 min at 75 °C and then immediately placing tube on magnet for 30 s.
12. Immediately remove eluate to a new 1.5 mL nonstick RNase-free microfuge tubes and place tubes on ice.

3.7 PolyA+ mRNA Zn-mediated Fragmentation

Chemical fragmentation of RNA is highly sensitive to sample concentration, sample volume, and fragmentation incubation temperature. Therefore, it is advised to first optimize the Zn-mediated fragmentation conditions (Fig. 3). Alternatively, other concentration independent and isothermal methods, such as physical shearing, may be applied.

1. Mix 1 µg mRNA with 3 µL 10× Zn-fragmentation buffer and H_2O to final 30 µL volume.
2. Incubate 30 min at 75 °C (*see* **Note 16** and Fig. 3). Place immediately on ice and add 3 µL 0.5 M EDTA to stop the reaction.
3. Add 570 µL H_2O, 2 µL GlycoBlue. Mix by flicking the tube.
4. Add 60 µL 3 M NaOAc pH 5.5 and 660 µL isopropanol. Mix by flicking the tube.
5. Precipitate for at least 30 min on dry ice (or overnight at −80 °C) (*see* **Note 17**).
6. Centrifuge for 30 min at 13,000 × *g* at 4 °C.
7. Visualize the pellet, carefully remove the supernatant, and add 800 µL 70% EtOH.

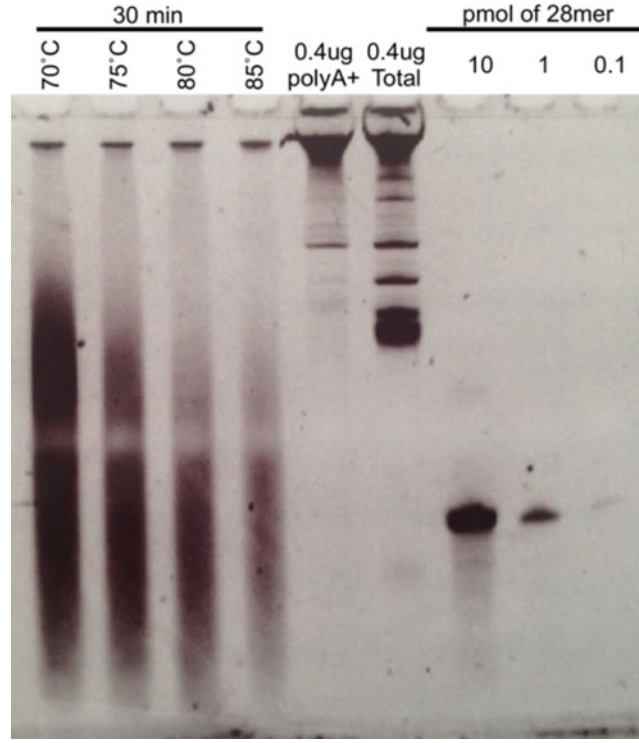

Fig. 3 *P. falciparum* total RNA optimization of Zn-mediated fragmentation. mRNA (1 μg) following Zn-mediated fragmentation for 30 min at 70, 75, 80, or 85 °C. Untreated mRNA control (0.4 μg polyA+) or total RNA (0.4 μg Total) were run in parallel for comparison. Samples were mixed with 10 ul 2× TBE-UREA sample loading buffer to a final volume of 20 ul with H_2O and loaded on 15% TBE-Urea gel run 65 min at 200 V along with different amounts of the 28-mer sizing control

8. Centrifuge for 15 min at 13,000 × *g* at 4 °C.
9. Visualize the pellet and carefully remove the supernatant using a pipette. Air dry the pellet by leaving the tube open at room temperature for 5 min (*see* **Note 18**).
10. Add 10 μL 10 mM Tris–HCl pH 8.0 and store at −80 °C.

3.8 Fragmented mRNA and Ribosome Footprint Size Selection

This protocol is adapted from Ingolia et al. Science 2009 [5], to be applied specifically to *P. falciparum*. However, since its inception, improvements to the ribosome profiling method have been developed that may be applied to the steps described below [6]. To reduce the introduction of technical errors, prepare mRNA and ribosome footprint libraries in parallel for each of the parasite time points.

1. Add 10 μL 2× TBE-UREA sample loading buffer to 10 μL Zn-fragmented mRNA and to 50 μg ribosome footprint sample in 10 μL 10 mM Tris–HCl pH 8.0 (*see* **Note 19**).

2. Mix 0.5 μL 10 bp ladder with 9.5 μL RNAse-free H_2O and 10 μL 2× TBE-UREA denaturing loading buffer.
3. Add 10 μL 2× Urea sample loading buffer to 1 μL 1 uM 28-mer and 31-mer in 9 μL RNAse-free H_2O.
4. Denature all samples and controls at 75 °C for 2 min in a PCR machine and place immediately on ice.
5. Pre-run a 10 well, 15% TBE-Urea, gel at 200 V for 10 min in 1× TBE running buffer and rinse urea from wells before loading samples.
6. Load gel with 20 μL sample making sure that every footprint sample is flanked by a 28-mer and 31-mer oligo sizing control.
7. Run gel at 200 V for 65 min.
8. Stain the gel in 60 mL 1× TBE running buffer plus 6 μL SYBR Gold for 5 min.
9. Use a new single-use individually sealed razor blade to cut each footprint and mRNA slices from gel and extract using the rapid gel extraction protocol (*see* **Note 20** and Fig. 4).

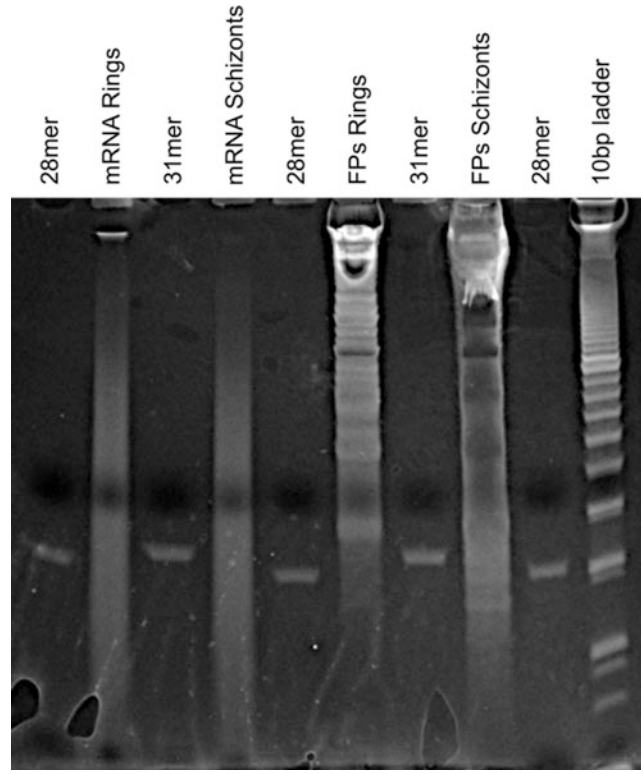

Fig. 4 Fragmented mRNA and ribosome footprint size selection. Fragmented mRNA and ribosome footprints (FP) from ring and schizont parasite stages were run alongside 28- and 31-mer sizing controls and a 10 bp ladder

3.9 Rapid Gel Extraction

1. Pierce a 0.5 mL nonstick RNAse-free microfuge tube with a 21-gauge needle and nest it inside a 1.5 mL nonstick RNAse-free microfuge tube. Cut both lids off.
2. Place gel slice inside the nested 0.5 mL tube and centrifuge the nested tubes for 3 min at 20,000 × *g* to force the gel through the needle holes. Shake any residual gel from the small tube into the larger tube.
3. Add 500 μL RNAse-free H_2O to the gel slices and incubate 10 min at 70 °C in a Thermomixer shaking at 400 rpm.
4. Use a clean razor blade to cut the tip of a P1000 tip and use it to pipet the gel mixture to a Costar Centrifuge-X column.
5. Centrifuge for 3 min at 20,000 × *g* to recover the elution mixture free of gel debris.
6. Transfer eluate to a new 1.5 mL nonstick RNAse-free microfuge tube.
7. Add 1.5 μL GlycoBlue and 55 μL NaOAc pH 5.5 and mix well.
8. Add 0.55 mL isopropanol and mix well.
9. Precipitate for at least 30 min on dry ice (or overnight at −80 °C).
10. Centrifuge for 30 min at 20,000 × *g* at 4 °C.
11. Visualize the pellet, carefully remove the supernatant, and add 800 μL 70% EtOH.
12. Centrifuge for 15 min at 13,000 × *g* at 4 °C.
13. Visualize the pellet and carefully remove the supernatant using a pipette. Air dry the pellet by leaving the tube open at room temperature for 5 min.
14. Resuspend the pellet in 10 μL 10 mM Tris–HCl pH 8.0.

3.10 3′ End Dephosphorylation

1. Setup two reactions using 1 μL 10uM oGAB11 (10 pmol) in 9 μL RNAse-free H_2O. Include oligo oGAB11 as a control from now on and in all subsequent steps.
2. Denature mRNA samples, ribosome footprint samples, and oGAB11 control for 2 min at 75 °C in a Thermomixer and place on ice.
3. Add 40 μL PNK reaction mix to each sample and mix well. Incubate at 37 °C for 1 h.
4. Incubate at 75 °C for 10 min to heat inactivate the enzyme.
5. Add 450 μL RNAse-free H_2O, 1.5 μL GlycoBlue, and 50 μL 3 M NaOAc pH 5.5 to reaction tube and mix well.
6. Add 600 μL isopropanol and mix well.
7. Precipitate for at least 30 min on dry ice (or overnight at −80 °C).

8. Centrifuge 30 min at 13,000 × *g* at 4 °C.
9. Visualize the pellet, carefully remove the supernatant, and add 800 µL 70% EtOH.
10. Centrifuge 15 min at 13,000 × *g* at 4 °C.
11. Visualize the pellet and carefully remove the supernatant using a pipette. Air dry the pellet by leaving the tube open at room temperature for 5 min.
12. Resuspend the pellet in 4.5 µL 10 mM Tris–HCl pH 8.0.

3.11 Linker Ligation

1. Add 1.5 µL Linker-1 at 0.5 µg/µL to 4.5 µL dephosphorylated sample or the oGAB11 control. Include a no ligase oGAB11 control.
2. Denature samples for 2 min at 75 °C in a Thermomixer and place immediately on ice.
3. Add 14 µL linker ligation reaction mix to each sample. Mix by flicking the tube but avoid creating bubbles.
4. Incubate at 37 °C for 2.5 h.
5. Add 338 µL RNAse-free H_2O and mix well.
6. Add 1.5 µL GlycoBlue and 40 µL 3 M NaOAc pH 5.5 and mix well.
7. Add 500 µL isopropanol and mix well.
8. Precipitate for at least 30 min on dry ice (or overnight at −80 °C).
9. Centrifuge for 30 min at 13,000 × *g* at 4 °C.
10. Visualize the pellet, carefully remove the supernatant, and add 800 µL 70% EtOH.
11. Centrifuge for 15 min at 13,000 × *g* at 4 °C.
12. Visualize the pellet and carefully remove the supernatant using a pipette. Air dry the pellet by leaving the tube open at room temperature for 5 min.
13. Resuspend samples in 10 µL 10 mM Tris–HCl pH 8.0.

3.12 Size Selection of the Linker-ligated Product

1. Pre-run a 10 well, 10% TBE-Urea, gel at 200 V for 10 min in 1× TBE running buffer.
2. Add 10 µL 2× TBE-Urea sample loading buffer to the samples from the previous step.
3. Mix 0.5 µL 10 bp ladder with 9.5 µL RNAse-free H_2O and 10 µL 2× TBE-Urea denaturing loading buffer.
4. Denature samples and controls for 2 min at 75 °C. Place immediately on ice.
5. Rinse urea from wells before loading 20 µL of sample per lane. Run at 200 V for 50 min.

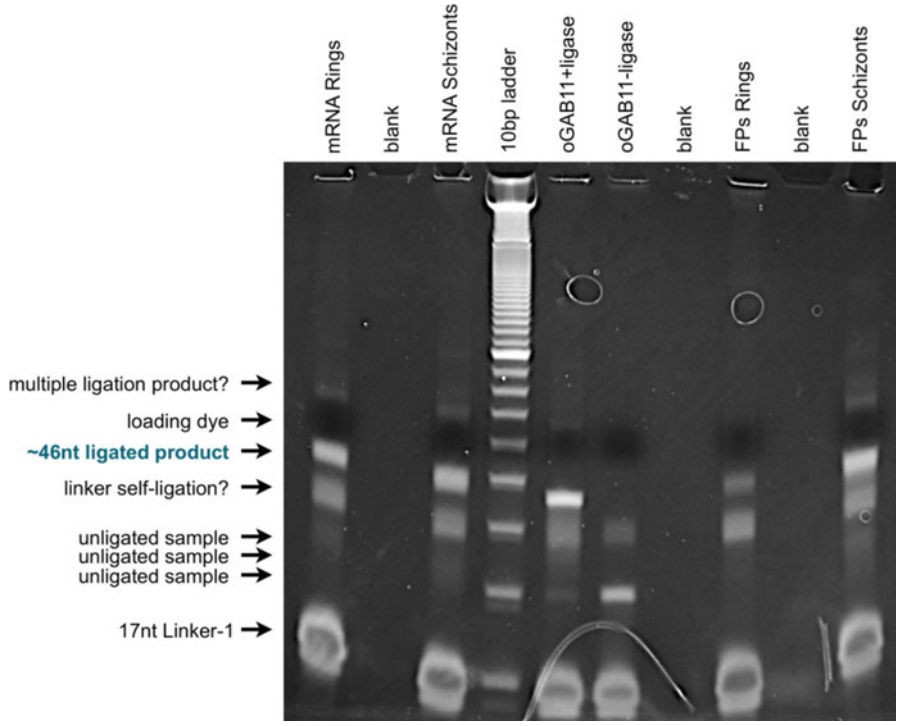

Fig. 5 Size selection of linker-ligated product. Use oGAB11 ligated control band (bright and just below the 40 bp ladder marker) as size reference to cut out the correct bands of your samples (indicated in blue). The ligated product is around 46 nt in size (17 nt Linker-1 + ~29mer). Note that the 10 bp DNA ladder gives only an approximate estimation of size since this is dsDNA and samples are ssRNA

6. Stain the gel in 60 mL 1× TBE running buffer plus 6 μL SYBR Gold for 5 min.
7. Using a clean razor, cut the linker-ligated product and oGAB11 control (Fig. 5).
8. Gel extract using the rapid gel extraction protocol.
9. Resuspend in 10 μL 10 mM Tris–HCl pH 8.0.

3.13 Ribosomal RNA Subtraction

This step removes four rRNA molecules found in great abundance in the *P. falciparum* ribosomal footprint samples and can be applied to the mRNA sample as well. The oligos were designed to uniquely map the most highly abundant sequences. Including additional oligos targeting other rRNA sequences may further decrease the 60% rRNA still detected after this depletion step.

1. Mix 3 μL of the biotinylated oligos asDNA1b, 2b, 3b, and 4b (at 20 pmol/μL) with 3 μL 20× SSC and 4 μL RNAse-free H_2O (*see* **Note 21**).
2. Add 19 μL biotinylated oligo mix to 10 μL sample from the previous step.

3. Incubate 2 min at 70 °C in a Thermomixer. Place immediately on ice.
4. Add 1 μL Superase-In and incubate 15 min at 37 °C in a Thermomixer shaking at 400 rpm.
5. Incubate for 5 min on ice.
6. Vortex to resuspend Dynabeads MyOne C1 and transfer 150 μL beads to a 1.5 mL nonstick RNase-free microfuge tube. Prepare one tube for each of your samples.
7. Wash Dynabeads three times in 150 μL B&W buffer 1× + 0.01% Tween.
8. Wash Dynabeads two times in 150 μL Solution A.
9. Wash Dynabeads two times in 150 μL Solution B.
10. Resuspend Dynabeads in 30 μL B&W buffer 2× + 0.01% Tween.
11. Add sample to washed Dynabeads and incubate in a Thermomixer at 400 rpm for 15 min at room temperature.
12. Place tubes on magnet and remove supernatant (~60 μL) to a new 1.5 mL nonstick RNAse-free microfuge tube.
13. Add 440 μL RNAse-free H_2O, 1.5 μL GlycoBlue, and 60 μL 3 M NaOAc pH 5.5 and mix well.
14. Add 500 μL isopropanol and mix well.
15. Precipitate for at least 30 min on dry ice (or overnight at −80 °C).
16. Centrifuge for 30 min at 13,000 × *g* at 4 °C.
17. Visualize the pellet, carefully remove the supernatant, and add 800 μL 70% EtOH.
18. Centrifuge for 15 min at 13,000 × *g* at 4 °C.
19. Visualize the pellet and carefully remove the supernatant using a pipette. Air dry the pellet by leaving the tube open at room temperature for 5 min.
20. Resuspend in 10 μL 10 mM Tris–HCl pH 8.0.

3.14 Reverse Transcription

1. Add 2 μL 2.5 μM oCJ200 RT primer to 10 μL sample from the previous step.
2. Split linker-ligated oGAB11 into 2 tubes each with 5 μL linker-ligated oGAB11 and 5 μL RNAse-free H_2O, to use as plus and minus RT controls.
3. Add 8 μL RT reaction mix to each sample tube and mix well.
4. Incubate at 48 °C for 30 min.
5. Add 1.8 μL 1 M NaOH, mix well, and incubate at 98 °C for 20 min.

6. Add 1.8 µL 1 M HCl, mix well, and place tube on ice.
7. Add 20 µL 2× TBE-Urea sample loading buffer.
8. Mix 0.5 µL 10 bp ladder with 9.5 µL RNAse-free H_2O and 10 µL 2× TBE-Urea denaturing loading buffer.
9. Denature samples at 95 °C for 3 min prior to loading and place on ice.
10. Pre-run a 10 well 10% TBE-Urea gel at 200 V for 10 min in 1× TBE running buffer.
11. Load samples and run samples at 200 V for 65 min.
12. Stain the gel in 60 mL 1× TBE running buffer plus 6 µL SYBR Gold for 5 min.
13. Using a clean razor, cut out the band corresponding to ligated product (Fig. 6).
14. Gel extract using the rapid gel extraction protocol. Remember to gel extract the oGAB11 plus RT control as well.
15. Resuspend in 15 µL 10 mM Tris–HCl pH 8.0.

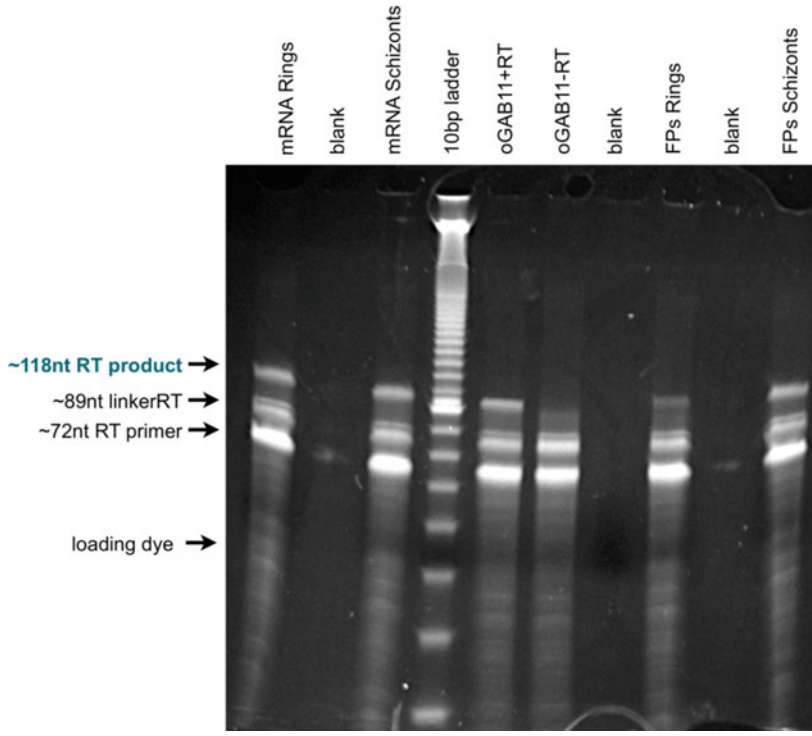

Fig. 6 RT product selection. Use oGAB11 + RT control band as size reference to cut out the correct bands of your samples (indicated in blue). The ligated product is around 118 nt in size (17 nt Linker-1 + ~29mer + ~72 nt oCJ200 RT primer). Note that the 10 bp DNA ladder gives only an approximate estimation of size since this is dsDNA and samples are ssDNA

3.15 Circularization

1. Add 5 μL circularization mix to each sample from the previous step.
2. Incubate at 60 °C for 60 min in a Thermomixer.
3. Incubate at 80 °C for 10 min to heat inactivate enzyme. Keep tube on ice if proceeding directly to product amplification or store at −20 °C indefinitely.

3.16 PCR Amplification

1. Add 95 μL PCR master mix to 5 μL circles from the previous step (*see* **Note 22**).
2. Set up 5 PCR tube strips and transfer a 16.7 μL aliquot of the PCR mixture into one tube in each strip.
3. Perform PCR amplification with varying numbers of cycles (6, 8, 10, 12, and 14) by placing all strip tubes in the PCR machine and starting a program with the conditions given below.

Initial denaturation	98° C, 30 s
Denaturation	98° C, 10 s
Annealing	60° C, 10 s
Extension	72° C, 5 s (return to 10 s denaturation step)

4. Remove strips successively at the very end of the extension step after 6, 8, 10, and 12 extension cycles, leaving the last strip in the PCR machine until the end of cycle 14.
5. Add 3.4 μL 6× DNA loading dye to each PCR tube. Do NOT denature samples before loading gel since now you have a dsDNA library.
6. Mix 1 μL 10 bp ladder with 15.7 μL RNAse-free H_2O and 3.4 μL 6× DNA loading dye.
7. Load samples on a 10 well 8% TBE non-denaturing gel (no urea) and run for 55 min at 200 V.
8. Stain the gel in 60 mL 1× TBE running buffer plus 6 μL SYBR Gold for 5 min.
9. Visualize and excise PCR product band from the gel (Fig. 7).
10. Pierce a 0.5 mL nonstick RNAse-free microfuge tube with a 21-gauge needle and nest it inside a 1.5 mL nonstick RNAse-free microfuge tube. Cut both lids off.
11. Place gel slice inside the nested 0.5 mL tube and centrifuge the nested tubes for 3 min at 20,000 × *g* to force the gel through the needle holes. Shake any residual gel from the small tube into the larger tube.
12. Soak the gel in 0.7 mL DNA elution buffer at room temperature overnight in a Thermomixer at 1400 rpm.

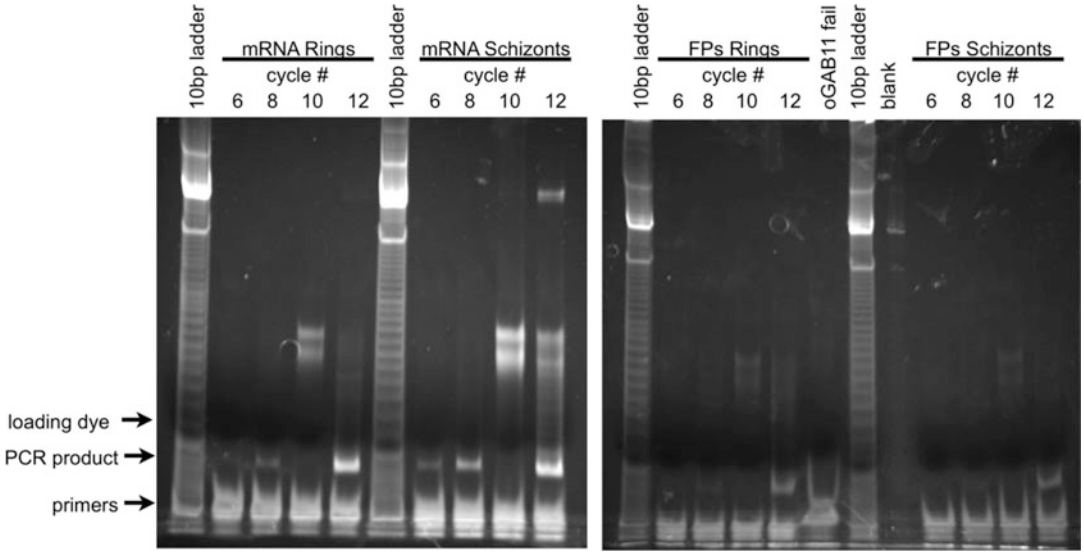

Fig. 7 PCR product selection. The indicated correct PCR ~ 175 bp product increases in intensity with cycle number. Note that the 10-cycle tube strip was taken out of the PCR machine too late as it was already at denaturing temperatures. This sample can therefore not be used

13. Use a clean razor blade to cut the tip of a P1000 tip and use it to pipet the gel mixture to a Costar centrifuge-X column.
14. Centrifuge for 3 min at 20,000 × *g* to recover the elution mixture free of gel debris.
15. Transfer eluate to a 1.5 mL nonstick RNAse-free microfuge tube, add 1.5 μL GlycoBlue, and mix well.
16. Add 0.7 mL isopropanol and mix well.
17. Precipitate for at least 30 min on dry ice (or overnight at −80 °C).
18. Centrifuge for 30 min at 13,000 × *g* at 4 °C.
19. Visualize the pellet, carefully remove the supernatant, and add 800 μL 70% EtOH.
20. Centrifuge for 15 min at 13,000 × *g* at 4 °C.
21. Visualize the pellet and carefully remove the supernatant using a pipette. Air dry the pellet by leaving the tube open at room temperature for 5 min.
22. Resuspend the pellet in 10 μL 10 mM Tris–HCl pH 8.0.
23. Determine library quality by measuring concentration and size distribution.

3.17 Sequencing and Data Analysis

Sequence libraries using a single-end 50 bp run on an Illumina platform. Split the FASTQ sequence data files into each of the indexes used for each library. Quality filter ribosome footprints and mRNA sequencing reads and trim to remove the library

adapter sequence (which begins with CTGTAGGCACCA) (*see* [6] for more details). Next, remove the first two nucleotides from the 5′ end of each read, as they frequently represent untemplated additions during reverse transcription (this causes loss of codon resolution). Align trimmed sequencing reads to *P. falciparum* rRNA, tRNA, and human red blood cell transcripts and collect the unaligned reads. Align the collected unaligned reads uniquely, and allowing no mismatches, to the *P. falciparum* reference genome using Bowtie [7] (*see* **Note 23**). Both mRNA and ribosome footprint rpkM [8] are calculated excluding the first 50 bases of each gene to eliminate bias introduced by the observed ribosome accumulation peak near the start codon. Translational efficiencies for each gene are calculated as the ratio of ribosome footprint rpkM to mRNA rpkM. Genome browser data tracks can be visualized in MochiView [9] or other visualization software.

4 Notes

1. Sample yields (detailed in Table 3) obtained from each stage increase as the parasite progresses through its life cycle, with ring, and schizont stage yields being the lowest and highest, respectively. Because hyperflask cultures have a volume, hematocrit, and parsitemia limit, multiple hyperflasks of ring stage parasites need to be collected in order to reach 10^{10}. To achieve a highly synchronous population, parasites are synchronized by two consecutive sorbitol treatments for three generations for a total of six treatments. Additionally, hyperflask cultures require media changes every 6–8 h minimum. The maximum invasion, point at which half of the culture is either rings or schizonts, is defined as hour zero, and independent time points are harvested 11, 21, 31, and 45 h later, for rings, early and late trophozoites, and schizonts, respectively. Executing every one of these steps as well as catching hour zero to harvest the desired timepoint require round the clock care for multiple weeks at a time.
2. If needed, use a hemocytometer to obtain an accurate count of the total cells harvested.
3. Yield of total RNA depends on the stage and number of parasites harvested; *see* Table 3. If more culture is needed, transfer additional 2 mL aliquots to as many 50 mL falcon tubes as needed. It is important that the packed culture volume is exactly 2 mL to be able to later estimate parasite number.
4. Freezing the parasite pellet at this step causes the collapse of polysomes.
5. Keeping track of the volume of the parasite suspension in this step is useful to estimate yields.

Table 3
Example of yields per parasite stage

Parameter	Rings	Schizonts	Notes
Total # of RBCs (infected and uninfected)	5.6×10^{11}	1.6×10^{11}	Hemocytometer count
% Parasitemia	8.2	6.6	Blood smear count
# Uninfected cells	9417	10,108	
# Rings	843	57	
# Trophozoites	0	2	
# Schizonts	0	656	
Total # parasites harvested	4.6×10^{10}	1.0×10^{10}	
Packed culture volume (mL)	72	20	Volume of pelleted RBCs harvested for lysis
# Parasites/mL of packed culture volume	6.4×10^{8}	5.2×10^{8}	
Packed culture volume allocated for total RNA extraction (mL)	16	4	
Packed culture volume allocated for ribosome profiling (mL)	56	16	
# Parasites harvested for total RNA extraction	1.0×10^{10}	2.1×10^{9}	
# Parasites harvested for ribosome profiling	3.6×10^{10}	8.3×10^{9}	
Average total RNA yield per 2 mL packed culture (μg)	118.60	840	
Total RNA per parasite (μg)	9.3×10^{-8}	8.1×10^{-7}	Note ~ 7-fold more RNA in schizonts than rings
Lysate volume loaded onto 1 gradient (ul)	350	360	6 gradients total – 4 loaded with nuclease-treated and 2 with the untreated control
# Parasites loaded onto 1 gradient	6.3×10^{9}	1.5×10^{9}	
Average footprint RNA yield per gradient (μg)	3.7	72.0	
Footprint RNA yield per parasite (μg)	5.8×10^{-10}	4.8×10^{-8}	83-fold more footprint RNA per schizont
Footprint RNA used for library prep (μg)	14.6	50.0	
# Parasites used for footprint library prep	2.5×10^{10}	1.0×10^{9}	
Total RNA used for library prep (μg)	150	150	
# Parasites used for mRNA library prep	1.6×10^{9}	1.8×10^{8}	

6. The milling step of this protocol was optimized using the Retsch MM400 Mixer mill. While other mills may work, a key point is to keep the parasite suspension frozen in liquid nitrogen during lysis, since polysome integrity was affected when parasites were lysed using a dounce on ice. The working volume in the 10 mL mixer mill grinding jar is 3 mL, but 2 and 4 mL may also work. If using different final volumes, make sure to adjust the PLB concentration accordingly.
7. Leaving frozen parasites in liquid nitrogen overnight affects polysome integrity.
8. The mixer mill jar size and milling frequency affects outcome, and excess milling can disintegrate polysomes. Jars have a tendency to lock up, making it difficult to unscrew them at the end of breaking cycles. To avoid this, loosen jar about a quarter of a turn when taking them out of liquid nitrogen and tighten them before returning them to lN_2.
9. Storing parasite powder at −80 °C or in liquid nitrogen at this step affects polysome integrity.
10. This step takes ~1–2 h and is amenable to optimization. Quick thawing at higher temperatures may cut down time.
11. Sucrose gradients should be made not more than 1 h before needed.
12. While the gradient making step of this protocol was optimized using the BioComp Gradient, any standard method to create a linear sucrose gradient can likely be used.
13. Remove up to 50 μL from the top of the gradient to hold more sample if necessary.
14. Do not vortex for more than 5 seconds to avoid shearing the mRNA.
15. A typical mammalian cell contains ~10–30 pg RNA of which 1–5% is mRNA. This proportion holds true for *P. falciparum* where ~150 μg Total RNA typically yield ~2.5 μg or 1.7% polyA+ mRNA.
16. The 75 °C treatment was chosen due to its enrichment of ~30 nt fragments, but treatment at 80 °C could also work. The optimal amount of 28-mer sizing control loading is 1 pmol.
17. Purifying RNA on filter columns in this step is not advised since these might introduce bias in fragment sizes.
18. Avoid over-drying the pellet because it becomes hard to resuspend.
19. As low as 14 μg ribosome footprints can be successfully processed.

20. Discrete bands present in the ribosome footprint samples likely correspond to rRNA fragments. Cut ribosome footprint bands out between just below the bottom of the 28-mer and just above the top of the 31-mer bands. Sizing is less critical here for the mRNA fragment sizes. Cut two slices between 35–50 nt and 50–70 nt. Keep working with the 35–50 nt slice and keep larger slice at −80 °C as a backup.
21. The antisense, biotinylated, DNA oligo sequences are listed in Table 1. They have a 5′ biotin modification with the standard linker included from IDT (C6) and should be HPLC purified and stored in 20 pmol/μL stocks in RNAse-free H_2O at −80 °C.
22. In this case, equal amounts of dNTPs are used instead of 80% dNTPs because the addition of Illumina sequences renders the library construct minimum 46% GC.
23. We found that during the reverse transcription reaction, the RT adds one or two untemplated bases to the read which is why we chose to trim two additional bases off of each read, thus loosing codon resolution.

References

1. Bozdech Z, Llinás M, Pulliam BL, Wong ED, Zhu J, DeRisi JL (2003) The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. PLoS Biol 1:e5
2. Silva EKD, Gehrke AR, Olszewski K, León I, Chahal JS, Bulyk ML, Llinás M (2008) Specific DNA-binding by apicomplexan AP2 transcription factors. Proc Natl Acad Sci 105:8393–8398
3. Caro F, Ahyong V, Betegon M, DeRisi JL (2014) Genome-wide regulatory dynamics of translation in the Plasmodium falciparum asexual blood stages. elife 3:e04106
4. Lambros C, Vanderberg JP (1979) Synchronization of Plasmodium falciparum erythrocytic stages in culture. J Parasitol 65:418–420
5. Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324:218 223
6. McGlincy NJ, Ingolia NT (2017) Transcriptome-wide measurement of translation by ribosome profiling. Methods 126:112–129
7. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25
8. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5:621 628
9. Homann OR, Johnson AD (2010) MochiView: versatile software for genome browsing and DNA motif analysis. BMC Biol 8:49

Chapter 4

Performing Ribosome Profiling to Assess Translation in Vegetative and Meiotic Yeast Cells

Emily Nicole Powers and Gloria Ann Brar

Abstract

Ribosome profiling, first developed in 2009, is the gold standard for quantifying and qualifying changes to translation genome-wide (Ingolia et al., Science, 2009). Though first designed and optimized in vegetative budding yeast, it has since been modified and specialized for use in diverse cellular states in yeast, as well as in bacteria, plants, human cells, and many other organisms (Ingolia et al. Science, 2009, reviewed in (Ingolia et al., Cold Spring Harb Perspect Biol, 2019; Brar and Weissman, Nat Rev Mol Cell Biol, 2015)). Here we report the current ribosome profiling protocol used in our lab to study genome-wide changes to translation in budding yeast undergoing the developmental process of meiosis (Brar et al., Science, 2012; Cheng et al., Cell, 2018). We describe this protocol in detail, including the following steps: collection and flash freezing samples, cell lysis and extract preparation, sucrose gradient centrifugation and monosome collection, RNA extraction, library preparation, and library quality control. Almost every step presented here should be directly applicable to performing ribosome profiling in other eukaryotic cell types or cell states.

Key words Translation, Meiosis, Yeast, Ribosome profiling, Gene expression, Cellular development

1 Introduction

Ribosome profiling is a quantitative assay used to report transcriptome-wide measurements of translation in a given cellular context ([1], reviewed in [2, 3]). This method, when performed in conjunction with mRNA sequencing, allows global identification of the transcriptional and translational regulatory changes between multiple cellular states ([1], reviewed in [2, 3]). The developmental process of meiosis is a context that depends on extensive gene regulation ([4, 5], reviewed in [6, 7]). This specialized cell division remodels diploid cells into specialized haploid gametes, called spores in yeast (reviewed in [6, 7]). The elaborate gene expression program that drives meiosis includes many cases of translation-level control, most of which were identified by ribosome profiling [4, 5, 8–11]. Interestingly, during meiosis, translation control

Vyacheslav M. Labunskyy (ed.), *Ribosome Profiling: Methods and Protocols*, Methods in Molecular Biology, vol. 2252, https://doi.org/10.1007/978-1-0716-1150-0_4,

coordinates the expression levels of sets of mRNAs, as well as changes the identities of some translated products (i.e., extended proteins, short proteins, and proteins initiated from non-AUG codons) [4, 12]. Our lab has used ribosome profiling to elucidate both of these types of changes to translation in meiosis and continues to use the ribosome profiling protocol for these purposes [4, 5]. Here we describe, in detail, our current protocol for using ribosome profiling to study changes to translation, in both vegetatively growing and meiotic yeast cells.

Successful ribosome profiling relies upon a rapid sample collection strategy to avoid perturbing the cellular state that you wish to measure, as well as a strategy (e.g., pharmacological agents) to halt ribosomes at the precise location on the mRNA that was being translated at the time of cell collection ([1], reviewed in [2, 3]). For traditional ribosome profiling, cycloheximide (CHX) is used to non-specifically stall elongating ribosomes [1]. This allows for collection of ribosome-protected fragments along the entire coding regions of translated genes, although it can also introduce positional artifacts [13–22]. Our standard protocol uses a brief CHX pre-treatment followed by rapid filtration and flash freezing of cells to maintain and measure the most accurate in vivo ribosomal positions [1, 4]. We note, however, that it is possible, and sometimes beneficial, to exclude a translation inhibitor pre-treatment of cells to avoid positional artifacts, depending on the downstream analysis desired. The disadvantage of excluding a translation inhibitor pre-treatment of cells is that ribosomes will continue to elongate between filtration and flash freezing which may result in the runoff of ribosomes from the 5′ end of open reading frames (ORFs) and a loss of the corresponding data for these regions ([1, 21–23], reviewed in [2]). However, it is well-known that translation inhibitor pre-treatments of cells can elicit technical artifacts in the precise ribosomal positions within coding regions. Furthermore, pre-treatment of cells with translation inhibitors can also cause biological changes to gene expression by inducing expression of ribosome biogenesis genes as a reaction to decreased translation [1, 4, 15, 17, 21, 23, 24]. Thus, one must weigh all the benefits and disadvantages of using a translation inhibitor pre-treatment when deciding to use one or not, depending on the specific desired downstream analyses to be performed.

Following rapid collection by filtration, cells are lysed cryogenically and the resulting extract is treated with RNAse I. The digested cell extract is then run on a sucrose gradient to isolate single ribosomes that protect ~30 nt mRNA fragments (also referred to as ribosome footprints (FPs)). Next, total RNA is extracted from the monosome fraction, and gel-based size selection is used to enrich for ribosome-protected mRNA fragments from the pool of mostly ribosomal RNA (rRNA). It is these fragments that are taken through a series of enzymatic steps to make a sequencing library [1, 4, 25].

Our standard protocol uses linker ligation in order to generate libraries out of the pool of mRNA FPs [13]. Though this method is often used for studies focused on gene expression quantification, it should be noted that ligation can produce bias in the FPs sequenced based on the nucleotide identity of the 3′ end of the FP. Thus, for studies focused on codon-level resolution of translation characteristics, care should be taken to use the most unbiased approaches possible in library preparation [22, 23, 25]. One alteration that can be used to alleviate linker ligation bias is to use a library of linker oligos, with random nucleotides on their 5′ end [22, 23]. Similarly, using a library of reverse transcription primers with varied nucleotides on their ends can help minimize any bias introduced during the following circularization step [22]. Following reverse transcription, linear sequencing libraries are created from the circular libraries using the minimal number of PCR cycles required to generate the desired amount of material.

After sequencing, the fragments are aligned to the yeast transcriptome to find the corresponding position of each ribosome. Footprints are then quantified per transcript to measure the amount of translation occurring on each transcript, or viewed on a genome browser or subjected to metagene analysis to look for changes to the regions being translated [1, 13, 24]. Comparison to a matched RNAseq sample prepared in parallel is used to determine whether gene expression changes are exerted at the transcriptional or translational level ([1, 4, 5], reviewed in [14]). The protocol described here can be applied to yeast cells in a variety of cellular states with only minor differences, based primarily on the total levels of translation and thus the amount of FPs collected in each state.

2 Materials

2.1 Media

Media should be made with water from a Milli-Q filtration system and sterilized prior to storing. Described percent compositions represent weight/volume calculations.

1. YPD: 1% yeast extract, 2% bacto peptone, 2% dextrose.
2. BYTA: 1% yeast extract, 2% bacto tryptone, 1% potassium acetate, 1.02% potassium phthalate.
3. SPO: 2% potassium acetate, 40 mg/L adenine, 40 mg/L uracil, 20 mg/L histidine, 20 mg/L leucine, 20 mg/L tryptophan, pH to 7.0 using acetic acid; bring to final volume.
4. YPG plates: 1% yeast extract, 2% bacto peptone, 3% glycerol, 2% agar.
5. YPD 4% plates: 1% yeast extract, 2% bacto peptone, 4% dextrose, 2% agar.

2.2 Base Reagents and Materials

All reagents and solutions should be prepared to be nuclease-free. Hazardous materials should be disposed of and used with proper safety precautions as specified by regulations. Reagents should be stored at room temperature unless otherwise indicated. Solutions are made in water unless otherwise noted.

1. 3 L sterile flasks for yeast culture.
2. Cycloheximide: 50 mg/mL in ethanol (500×); store at −20 °C, dispose of cycloheximide hazardous waste by following the appropriate regulations, and handle all materials containing cycloheximide with gloves (*see* **Note 1**).
3. Nuclease-free water.
4. 1 M Tris–HCl (pH 7.0).
5. 1 M Tris–HCl (pH 8.0).
6. 10 mM Tris–HCl (pH 7.0).
7. 10 mM Tris–HCl (pH 8.0).
8. 2 M KCl.
9. 1 M $MgCl_2$.
10. 20% Triton X-100.
11. 20- and 22-gauge needles.
12. 50 mL plastic, screw cap, nuclease-free tubes.
13. 2 mL screw cap, nuclease-free tubes.
14. 1.5 mL non-stick, nuclease-free tubes.
15. 0.5 mL non-stick, nuclease-free tubes.
16. Metal spatulas.
17. Liquid nitrogen (LN_2): Note that LN_2 can cause cryogenic burns and frostbite and can displace oxygen. It should be stored and used according to established safety regulations and handled using cryo-gloves in a well-ventilated area.
18. Styrofoam box: Ideally at least 6 inches each in height, width, and depth.
19. Plastic rack for 50 mL tubes (fits within styrofoam box).
20. Tube rack to hold ultracentrifuge tubes (small 15 mL tube racks work).
21. Filter membranes, 0.45 μm pore size, cellulose nitrate, to be used with the glass filtration apparatus for yeast cell harvesting (*see* **Note 2**).
22. Metal tweezers.
23. Metal tongs to hold mixer mill chambers.
24. Cryo-gloves.

25. 12 mL open-top, polyclear ultracentrifuge tubes and short rubber caps (*see* **Note 3**).
26. 1 M DTT: store at −20 °C, and handle on ice.
27. Ultra-pure sucrose.
28. 10 mL syringes with long metal tip attachments and SW 41 Ti marker block.
29. RNAse I: 100 U/μL; store at −20 °C, and handle on ice.
30. Acid phenol/chloroform/isoamyl alcohol (125:24:1), pH 4.5; store at 4 °C. Note that acid phenol is hazardous; use and dispose it of according to established safety regulations.
31. Chloroform: Note that chloroform is hazardous; use and dispose of it according to established safety regulations.
32. 20% sodium dodecyl sulfate (SDS).
33. 3 M NaOAc, pH 5.5.
34. Isopropanol.
35. 80% ethanol, ice-cold; store at −20 °C.
36. GlycoBlue; store at −20 °C (*see* **Note 4**).
37. 15% TBE-Urea polyacrylamide gels; store at 4 °C (*see* **Note 5**).
38. 10% TBE-Urea polyacrylamide gels; store at 4 °C (*see* **Note 5**).
39. 8% TBE polyacrylamide gels; store at 4 °C (*see* **Note 5**).
40. TBE-Urea sample buffer 2×; store at 4 °C.
41. SYBR Gold; store at −20 °C. Note that SYBR Gold is hazardous, use and dispose of it according to established safety regulations (*see* **Note 6**).
42. 1× TBE: 89 mM Tris, 89 mM Borate, 2 mM EDTA.
43. 10 bp ladder (optional).
44. Single-use razors or scalpels.
45. Centrifuge tube filters, cellulose acetate, 0.45 μm pore size.
46. SUPERase·In (20 U/μL); store at −20 °C, and handle on ice.
47. T4 polynucleotide kinase (New England Biolabs (NEB)); store at −20 °C, and handle on ice (*see* **Note 7**).
48. Truncated T4 RNA ligase 2 (NEB); store at −20 °C, and handle on ice (*see* **Note 7**).
49. Oligo Clean & Concentrator kit (Zymo Research) (*see* **Note 8**).
50. 3 M NaCl.
51. 0.5 M EDTA.
52. 10% Tween 20.
53. MyOne Streptavidin C1 Dynabeads (Thermo Fisher); store at 4 °C (*see* **Note 9**).

54. 20× SSC: 3 M sodium chloride, 300 mM sodium citrate.
55. 1 M NaOH.
56. 1 M HCl.
57. dNTPs: 10 mM dATP, 10 mM dCTP, 10 mM dGTP, 10 mM dTTP; store at −20 °C, and handle on ice.
58. Superscript III reverse transcriptase (Invitrogen); store at −20 °C, and handle on ice (*see* **Note 7**).
59. CircLigase II ssDNA ligase (Epicentre); store at −20 °C, and handle on ice (*see* **Note 7**).
60. 1 mM ATP; store at −20 °C, and handle on ice.
61. Phusion DNA polymerase (2000 U/mL NEB); store at −20 °C, and handle on ice (*see* **Note 7**).
62. 6× gel loading dye for non-denaturing polyacrylamide gels.
63. High Sensitivity D1000 ScreenTape (Agilent Technologies) (*see* **Note 10**).
64. High Sensitivity D1000 Reagents (Agilent Technologies) (*see* **Note 10**).

2.3 Non-standard Equipment or Facilities Required

1. 30 °C shaker for yeast cultures.
2. Glass filtration apparatus for yeast cell harvesting (*see* **Note 2**).
3. Mixer mill that can be cryogenically operated with 50 mL stainless steel cannisters (*see* **Note 11**).
4. Refrigerated centrifuge (fits 50 mL tubes).
5. Refrigerated microcentrifuge.
6. Ultracentrifuge with Beckman SW 41 Ti rotor and buckets.
7. Gradient station (*see* **Note 3**).
8. UV monitor for gradient station (*see* **Note 3**).
9. Light box for gel cutting.
10. Magnet rack for 1.5 mL tubes to use for steps involving Dynabeads.
11. Agilent Bioanalyzer, TapeStation, or equivalent (*see* **Note 10**).
12. Illumina HiSeq 4000.

2.4 Buffers and Solutions

All buffers listed are to be prepared in water unless otherwise noted.

1. Polysome lysis buffer: 20 mM Tris–HCl pH 8.0, 140 mM KCl, 1.5 mM $MgCl_2$, 100 μg/mL cycloheximide, 1% (v/v) Triton X-100; do not thaw and re-freeze buffer.
2. Polysome gradient buffer: 20 mM Tris–HCl pH 8.0, 140 mM KCl, 5 mM $MgCl_2$, 100 μg/mL cycloheximide, 500 μM DTT, 20 U/mL SUPERase·In; make fresh on ice.

3. 10% sucrose: 10% sucrose (w/v) in polysome gradient buffer; make fresh.
4. 50% sucrose: 50% sucrose (w/v) in polysome gradient buffer; make fresh.
5. Dynabead B&W buffer (2×): 10 mM Tris–HCl pH 7.5, 1 mM EDTA, 2 M NaCl, 0.01% Tween 20 (*see* **Note 9**).
6. Dynabead B&W buffer (1×): 5 mM Tris–HCl pH 7.5, 500 μM EDTA, 1 M NaCl, 0.01% Tween 20 (*see* **Note 9**).
7. Dynabead Solution A: 100 mM NaOH, 50 mM NaCl (*see* **Note 9**).
8. Dynabead Solution B: 100 mM NaCl (*see* **Note 9**).

2.5 Oligonucleotides

Example sequences for sizing oligos are given; however, the specific sequence of these oligos is not critical. The 28-mer RNA oligo and 31-mer RNA oligos described here are unpublished and were designed by Gloria Brar and Nick Ingolia. If sequences are changed for the ligated oligo, note that the reverse transcription primer must be designed to amplify the sequence used. Oligos for rRNA subtraction were designed by Nick Ingolia and Gloria Brar [13]. The linker and amplification primers described here were designed and validated by Calvin Jan and their sequences generously shared with our lab. Alternative linker and primer sequences can be used for library preparation, but they must be validated prior to use to ensure high efficiency of ligation and amplification, respectively. For all oligonucleotides listed that include specialized modifications, we described the modifications present and listed the modification codes used by IDT. Equivalent modifications by other oligonucleotide providers may likely be substituted; however, in all cases, any oligonucleotides used must be confirmed to be highly efficient and unbiased for sequencing library preparation. We dilute all oligos in water to the stated concentrations and store stocks at −20 °C.

1. 28-mer RNA control oligo: 5′AGUCACUUAGCGAUGUA CACUGACUGUG /3Phos/3′, oligo has a 3′ phosphate (3Phos) and can be used for FP sizing and as a positive control and sizing guide throughout library preparation, 10 μM, oligo is PAGE purified.
2. 31-mer RNA control oligo: 5′AUGUACACGGAGUCGAG CACCCGCAACGCGA/3Phos/3′, oligo has a 3′ phosphate (3Phos) and is used for FP sizing, 10 μM, oligo is PAGE purified.
3. Linker: 5′/5rApp/ GATCGGAAGAGCACACGT /3ddC/3′, oligo is 5′ adenylated (5rApp), terminated using a 3′ dideoxycytidine (3ddC), DNA, and is ligated onto FPs for library prep, 20 μM, oligo is HPLC purified.

4. asDNA1b: 5′/biosg/ GATCGGTCGATTGTGCACC 3′, DNA, 5′ biotin (biosg) with the standard linker from IDT (C6), and HPLC purified, used for rRNA subtraction from *S. cerevisiae* FPs, 20 pmol/μL.
5. asDNA2b: 5′/biosg/ CCGCTTCATTGAATAAGTAAA GAAAC 3′, DNA, 5′ biotin (biosg) with the standard linker from IDT (C6), and HPLC purified, used for rRNA subtraction of *S. cerevisiae* FPs, 20 pmol/μL.
6. asDNA3b: 5′/biosg/GACGCCTTATTCGTATCCATCTATA 3′, DNA, 5′ biotin (biosg) with the standard linker from IDT (C6), and HPLC purified, used for rRNA subtraction of *S. cerevisiae* FPs, 20 pmol/μL.
7. RT primer: 5′/5Phos/AGATCGGAAGAGCGTCGTGTAGG GAAAGAGTGT /iSp18/ GTGACTGGAGTTCA GACGTGTGCTCTTCCGATC 3′, DNA, internal spacer 18 from IDT (iSp18), this primer is used to reverse transcribe linker-ligated RNA FPs, 20 μM, oligo is PAGE purified.
8. PCR F primer: 5′AATGATACGGCGACCACCGAGATCTA CACTCTTTCCCTACACGACGCTCTTCCGATCT 3′, DNA, used to amplify all FP samples, 10 μM, oligo is PAGE purified.
9. Barcoding primers: 5′CAAGCAGAAGACGGCATACGAGAT **XXXXXXXX**GTGACTGGAGTTCAGACG3′, oligo is PAGE purified, reverse primer used to amplify and barcode samples with Illumina barcodes in place of "**XXXXXXXX**", we use the following barcodes: D701 index ATTACTCG, D702 index TCCGGAGA, D703 index CGCTCATT, D704 index GAGATTCC, D705 index ATTCAGAA, D706 index GAATTCGT, D707 index CTGAAGCT, D708 index TAATGCGC, D709 index CGGCTATG, D710 index TCCGCGAA, D711 index TCTCGCGC, and D712 index AGCGATAG. Example sequence of full primer sequence with the D701 index, 5′ CAAGCAGAAGACGGCATACGAGAT **CGAGTAAT**GTGACTGGAGTTCAGACG3′.

3 Methods

Carry out all methods at room temperature unless otherwise noted.

3.1 Yeast Growth and Sporulation Conditions (See Note 18)

Meiotic Samples

1. Day 1: Thaw a fresh patch of diploid yeast from the desired glycerol stock onto a YPG plate at ~5 pm. Grow patch at 30 °C for ~16 h.
2. Day 2: Patch yeast from the YPG plate to a YPD 4% plate at ~9 am; grow for ~8 h at 30 °C. At ~5 pm transfer a scoop of yeast into 10 mL of YPD liquid media. Grow cells, with shaking, for either ~24 h at room temperature or ~16 h at 30 °C.

3. Day 3: At ~5 pm measure the OD_{600} of the YPD culture. Dilute the culture such that it is in the accurate range of the spectrophotometer (1:20 usually works well), before measuring. Start a 100 mL culture of BYTA at 0.25 OD_{600}; grow overnight (12–16 h) at 30 °C, with shaking.
4. Day 4: At ~9 am measure the OD_{600} of the BYTA culture; calculate the amount of culture volume needed to start an SPO culture of 200 mL at a density of 1.9 OD_{600} (note that this is an excess quantity for one ribosome profiling sample). Pellet cells in BYTA (1100–2000 rcf for 2–2.5 min). Discard the supernatant, and resuspend cells with sterile Milli-Q water (wash should be greater than the volume of culture pelleted). Pellet cells in water (1100–2000 rcf for 2–2.5 min), and discard the supernatant. Resuspend pellet in 200 mL of SPO and move to a 3 L flask, shake culture at 30 °C, and record the time.
5. Take samples for staging (to assess DNA content, spindle morphology, etc.) as necessary (*see* **Note 13**). Ten minutes prior to the desired time of sample collection, measure 150 mL of SPO culture and transfer the excess volume to a new flask to monitor meiotic progression in parallel. Continue to shake both at 30 °C. Proceed to Subheading 3.2.

Vegetative Exponential Phase Samples

1. Day 1: Thaw yeast and start an overnight culture exactly as described for the meiotic sample protocol (**steps 1** and **2**).
2. Day 2: Measure the OD_{600} of the culture, and calculate the amount of culture volume needed to start a 450 mL culture at a cell density of 0.05 OD_{600}. Pipette the calculated volume of yeast into a 3 L flask with 450 mL of YPD. Place culture on a shaker and incubate at 30 °C until grown to ~0.6 OD_{600}. Measure growth periodically by taking and recording the OD_{600}. As cell density nears 0.6 OD_{600} (mid-exponential phase), move on to Subheading 3.2.

3.2 Cell Harvesting and Preparing Polysome Lysis Buffer (See Notes 19 and 20)

1. Label and prepare one 50 mL tube per sample by piercing 3–4 holes in the cap with a needle, and label one 2 mL tube/sample for the matched RNAseq sample (*see* **Note 12**).
2. Set up the filtration apparatus. Connect a filter flask to a vacuum pump, place a filter piece into the flask, and lay the nitrocellulose membrane on top. Place the collection beaker over the filter, and attach it to the filter piece with a clasp.
3. Fill a styrofoam box with LN_2, and place a 50 mL tube rack inside. Cool one large and one small metal spatula in liquid nitrogen with the handles sticking out. Fill the first labeled 50 mL tube with about 20–30 mL LN_2, and leave uncapped,

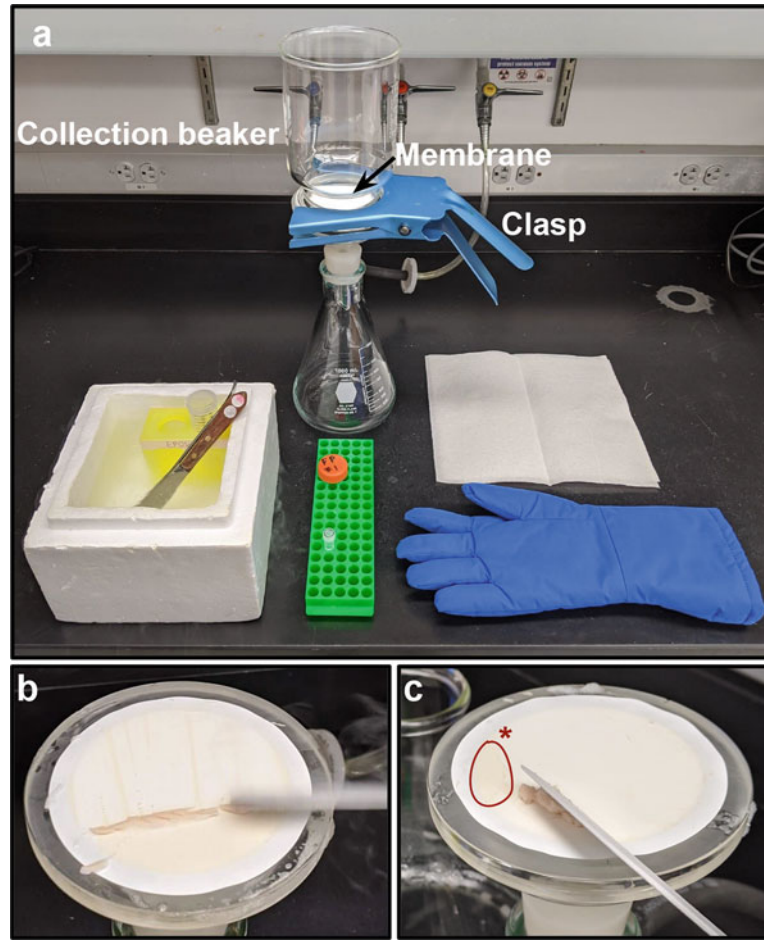

Fig. 1 The setup for ribosome profiling/mRNAseq sample collection. (**a**) A typical benchtop setup in preparation for filtration. A filter flask is connected to a vacuum pump as shown. The filter piece is placed into the filter flask with a cellulose nitrate membrane (labeled) laid on top. Next, the collection beaker (labeled) is placed on top of the filter piece and over the membrane and held in place using a clasp (labeled). A styrofoam box is close by and contains LN_2, two cooled metal spatulas (one large, one small), and a labeled uncapped 50 mL tube with LN_2 inside (with a labeled and pierced cap nearby). A labeled 2 mL tube is nearby for RNAseq sample collection. A cryo-glove is used to protect your hands from the cooled metal spatulas and LN_2. (**b**) An image of a LN_2-cooled metal spatula horizontally scraping yeast into a large pile for ribosome profiling collection. (**c**) A small patch of un-scraped yeast is circled as it is left for RNAseq sample collection; the larger heap of yeast will be scraped directly into LN_2 in the 50 mL tube for ribosome profiling

within the box of LN_2. Uncap the 2 mL tube for the same sample. *See* Fig. 1 for an example filtration apparatus and representative images of cell collection from the filter.

4. Take CHX stock (50 mg/mL) out of the freezer on ice. If CHX has precipitated out of solution, vortex until fully dissolved and then return to ice (*see* **Note 15**). Turn the vacuum pump on.
5. Add CHX to 100 μg/mL in the first culture (300 μL of 50 mg/mL stock for the 150 mL meiotic culture, 900 μL of 50 mg/mL stock for 450 mL of YPD culture); return shaking for 30 s (*see* **Note 15**). Carry flask quickly to filter apparatus, and pour culture into collection container. Place flask down, and as media is almost completely filtered, remove the clasp and collection top and begin scraping cells horizontally off the filter with the large cooled spatula. Scrape ~90% of the cells into a mound on the metal spatula. Plunge the cells, and if needed, the entire spatula directly into the LN_2-filled conical. Quickly recover the small patch of cells leftover into the 2 mL tube with the smaller cooled spatula, cap the tube, and drop into LN_2.
6. Place the 50 mL sample tube with LN_2 upright to allow venting from the cap in a −80 °C freezer, and store the 2 mL tube there as well. Allow LN_2 to completely evaporate off before moving on to sample lysis.
7. Dispose of the used membrane, and rinse or switch the collection container. Repeat **steps 2–7** for each remaining sample.
8. When sample collection is complete, rinse all the glassware and spatulas with Milli-Q water and allow to dry.
9. Next, make 5 mL of polysome lysis buffer (enough for two samples at 2–2.5 mL/sample). Mix 4.28 mL of water, 100 μL 1 M Tris–HCl (pH 8.0), 350 μL 2 M KCl, 7.5 μL 1 M $MgCl_2$, 10 μL 50 mg/mL CHX, and 250 μL 20% Triton-X 100 in a 50 mL tube on ice.
10. Label and prepare one 50 mL tube per sample exactly as done previously for sample tubes in **steps 1** and **3**. Uncap tubes, ensure buffer is mixed, and slowly dispense 2 mL of polysome lysis buffer directly into the LN_2 within each 50 mL tube. Small 5 mL serological pipettes work well to dispense buffer at a controlled rate into the LN_2 such that it freezes in droplets before contacting the sides of the tube. It is important that the frozen buffer is not stuck to the tubes for later use. Cap tubes with pierced caps, and allow LN_2 to freely evaporate while tubes are sitting upright in the −80 °C freezer.
11. When LN_2 has completely evaporated off both samples and buffer aliquots, move on to Subheading 3.3. Alternatively, samples can be stored at −80 °C indefinitely.

3.3 Yeast Cell Lysis (See Notes 11 and 21)

1. Record which sample will go in each mill chamber prior to starting, place chambers and balls into a LN_2-filled styrofoam box using tongs, and place a small 50 mL tube rack into the box as well. Let chambers cool until boiling stops (wait until large bubbles have ceased and only small ones persist).

2. Collect each sample/buffer tube from the freezer, and place into the LN_2 to prevent thawing. Loosen pellets by softly hitting each tube with metal tongs, and pour one aliquot of the polysome lysis buffer pellets prepared in Subheading 3.2, **step 9** into each tube with a frozen sample pellet.
3. Use metal tongs to remove both halves of chamber one from LN_2, and place onto a paper towel on the bench top. Pour out any liquid LN_2 from both chambers, and place the ball into the large chamber. Pour sample one into the large half of chamber one; lightly hit the bottom of the overturned tube to dislodge any stuck buffer or cell pellet. Close the chamber, and place back into LN_2 to recool. Repeat this process with chamber two and sample two. Keep labeled sample tubes and caps nearby.
4. Turn on the mixer mill. Set the frequency to 15 Hz and breaking rounds to 3 min. Pull up knobs on mixer mill holders and twist to keep them open. Make sure that holders are fully loosened with the side wheels.
5. Remove chambers from LN_2 with tongs, and loosen one quarter of a turn. Place chambers into the holders, and use the side wheels to tighten them into place. Once mostly tightened, close the knobs on the holders, and tighten for 1–2 more clicks. Be sure that the chamber is properly sealed within the holder, and do not overtighten the wheels.
6. Press start, and record which side of the mill each chamber was in as well as the orientation of the chambers. When the round is complete, lift knobs on the chamber holder and twist to hold open, loosen wheels, remove chambers, tighten chambers, and return them to LN_2 to recool.
7. Repeat **steps 5** and **6** two more times such that each sample has had three total rounds of breaking, alternating the side and orientation that each chamber is on each round. During the last round of breakage, precool two spatulas (one per sample).
8. While chambers are recooling, place the empty sample tube from sample one uncapped into LN_2 to recool. Dump out any LN_2 within the tube prior to collecting sample.
9. Remove chamber one from LN_2, and tap the chamber firmly on the top and the sides with metal tongs; this should dislodge most of the cell powder to the bottom of the large chamber. Place the chamber onto a paper towel on the bench top, and unscrew the small half of chamber and remove vertically.
10. Use a cooled spatula to scoop cell powder from the top chamber into the open conical. When most of the powder is gone from the top chamber, transfer the ball into this half, and proceed to scoop the rest of the powder from the larger half of the chamber. When complete, cap sample, and store indefinitely at −80 °C, move the used chambers aside, and repeat **steps 8–10** with sample two.

11. When you are done with both chambers, rinse them with hot water and then distilled water. Make sure to remove the silicone rings and wash the entire chamber thoroughly. Spray chambers with methanol and allow to air dry. If multiple rounds of milling are to be done using the same chambers, be sure to let them dry completely in between use. Move forward to Subheading 3.4.

3.4 Extract Preparation (See Note 22)

1. Label two non-stick 1.5 mL tubes and six screw cap nuclease-free 2 mL tubes per sample, and place tubes on ice. Preheat a water bath to 30 °C, and cool both a large centrifuge with 50 mL adaptors and a microfuge for 1.5 mL tubes to 4 °C.
2. Hold tubes of cell powder in a 30 °C water bath, and swirl until thawed. Quickly move the tubes to ice, and proceed immediately to **step 3**.
3. Spin samples for 5 min at 3000 rcf at 4 °C.
4. Move supernatant to 1.5 mL tubes on ice (~2 mL total for each sample), and spin for 10 min at 20,000 rcf at 4 °C.
5. Take the supernatant into a screw cap 2 mL tube (combine duplicate tubes), avoiding unwanted material (*see* **Note 22**). Briefly vortex each sample, and dispense into 200 μL aliquots in the screw cap tubes on ice. Mix 5 μL of each leftover sample with 495 μL of 10 mM Tris pH 7, to make a 1:100 dilution. Freeze sample aliquots in LN_2 and store at −80 °C, or keep two aliquots per sample on ice if proceeding directly to Subheading 3.5.
6. Measure the 260 nm absorbance of the diluted samples with a nanodrop. Record the A_{260} units, and calculate the total A_{260} units present in each undiluted aliquot. If a diluted sample has a A_{260} of greater than 5 such that there are more than 100 A_{260} units in the total aliquot of 200 μL, do not use the whole aliquot for digestion. Split the aliquot to keep the A_{260} units within the range of 20 and 100 A_{260} units per digestion (*see* **Note 22** for sample calculations).
7. Store extract aliquots at −80 °C indefinitely, or move forward to Subheading 3.5.

3.5 Footprint Isolation by Sucrose Gradient (See Note 23)

Note that the directions below apply to the BioComp Gradient Master and BioRad Economonitor and may differ for other gradient collection setups.

1. Cool the SW 41 Ti rotor, paired buckets, and ultracentrifuge to 4 °C. Thaw two extract aliquots per sample, and keep on ice (one for digestion and footprint isolation and one for "mock" digestion to ensure the quality of the extract).

2. Make 50 mL of polysome gradient buffer by mixing 1 mL 1 M Tris pH 8.0, 3.5 mL 2 M KCL, 250 μL 1 M $MgCl_2$, 100 μL 50 mg/mL CHX, 25 μL 1 M DTT, 50 μL SUPERase·In, and 45.08 mL of water (enough for four gradients).
3. Make 30 mL of 10% sucrose. Mix 3 g of ultra-pure sucrose with 27.8 mL of polysome gradient buffer in a 50 mL tube. Make 30 mL of 50% sucrose by mixing 14.9 g of ultra-pure sucrose with 20.8 mL of polysome gradient buffer in a 50 mL tube. Shake vigorously until fully dissolved (usually ~15–30 min, using a lab shaker).
4. Calculate the amount of RNAse I (100 U/μL) to add to each sample for digestion. Add 10 U RNAse I per A_{260} unit of extract. For example, if a sample contains 30 A_{260} units total in the 200 μL aliquot, add 3 μL of RNAse I.
5. Add RNAse I to "cut" sample tubes and mix by gently flicking tubes. Set up "mock" digested samples by adding an equal volume of SUPERase·In, in place of RNAse I. Incubate for 1 h at room temperature with slight mixing. When the hour is complete, place tubes onto ice.
6. Take out a 12 mL ultracentrifuge tube for each "cut" or "mock" digested sample, and clean tubes with compressed air to remove any debris.
7. Place a tube in the marker block, and mark a line along the upper edge of the block; repeat for all tubes. Collect sucrose solutions from shaker, and allow all bubbles to rise to the surface before use.
8. Fill a 10 mL serological pipette with ~6 mL of 10% sucrose, and slowly dispense sucrose into the bottom of an ultracentrifuge tube until it reaches the marked line. Pull ~6.3 mL of 50% sucrose into a 10 mL syringe, wipe the outside of the metal syringe tip, and place the tip of the syringe at the bottom of the tube. Slowly inject 50% sucrose until it reaches the marked line. As the interface between the layers rises, move the tip of the syringe such that it is always just below the interface, and remove carefully to avoid disruption of the interface. Repeat for all tubes.
9. Place a rubber cap onto each tube, and lower the cap at an angle with the side of the cap with the hole being the last to lower. Take note of where each cap hole is by marking the side of the tube.
10. Turn on the gradient station, and make 7–47% sucrose gradients for SW 41 Ti rotor buckets for tubes with short caps (81.5° tilt, rotating at 16 rpm, for 2 min). Ensure the platform is leveled, and place tubes in holder. Run the program.

11. Retrieve the cooled SW 41 Ti buckets, and uncap them while the program is running.
12. Carefully remove tubes from the gradient station. Remove caps by raising the side of the cap with the hole (near the marking) first at an angle. Wipe the outside of each tube carefully to remove any sucrose. It is very important that sucrose does not get into the rotor buckets or threading along the bucket lids. Place each tube into a bucket, loosely cap the buckets, and recool for 15 min at 4 °C. Soak rubber caps in water until you are ready to wash them.
13. Pulse spin down "cut" and "mock" digested samples, and return to ice. Retrieve the cooled gradients, and remove the caps to the rotor buckets. Make a list of which samples will go into each bucket. Load each sample onto the top of the respective gradient by slowly pipetting the sample against the wall of the tube, just over the gradient, until the entire sample is floating across the top. Repeat with all samples.
14. Balance opposing buckets (1 and 4, 2 and 5, 3 and 6) to within 10 mg of each other. Use 10% sucrose to carefully add weight to the top of the gradient of the lighter of the paired buckets, and include the caps when balancing each bucket.
15. Screw lids onto buckets tightly, and avoid contacting sucrose with lids. Hang buckets on their respective positions of the rotor, and carefully place the rotor into the ultracentrifuge. Always hang empty buckets (without ultracentrifuge tubes within them) on the rotor when spinning less than six gradients, and always make sure to balance the rotor properly.
16. Set up a run that is cooled to 4 °C, lasts for 3 h, and spins at 35,000 rpm (151,000 rcf).
17. Start ultracentrifugation, and ensure the centrifuge reaches proper speed and temperature. It will take 5–10 min for the vacuum to fully engage and for the temperature and speed to adjust. While the spin is running, rinse all equipment used to set up the gradients including the gradient caps and syringe tips, and wipe the gradient station with a wet paper towel. Also label two 2 mL screw cap tubes for each "cut" sample to use for monosome collection.
18. After the spin is completed, remove the rotor and place buckets into holders, enter the spin into the ultracentrifuge log, and turn off the ultracentrifuge.
19. Turn on the gradient station, UV lamp, and computer. Place a collection tip onto the gradient piston, and position output tubing into a waste container. Set the voltage rate on the UV lamp to between −2.0 and +2.0 to achieve maximum detection range.

20. Carefully remove the first "mock" digested gradient from the bucket using tweezers, snap tube into the top piece of the gradient holder, and ensure it is properly inserted and can spin freely before removing support from the bottom of the tube. Place the tube and top into the cylindrical gradient holder, and spin to lock into place. Place the holder onto the platform, and spin to lock into place with the window facing you.
21. On the gradient panel, select "fractionate," and lower the piston slowly until it comes just into contact with the top of the fraction. Choose "Rset" to mark the position of the top of the gradient, and select "Singl." Set the following parameters on the gradient station panel, 0.2 mm/s "Speed," between 75 and 80 mm "Dist," and 1 fraction for "Numb."
22. Press "start" on the gradient station, and start recording on the UV monitor. You can either collect all polysome fractions for downstream analysis or discard them into a waste container if not needed.
23. Watch profile on the computer screen. You should first see a large spike of material at the top of the tube representing free mRNA and other cellular material that absorbs at 260 nm. Once that material has been cleared, you should see much smaller peaks representing the small (40S) ribosome subunit and the large (60S) ribosome subunit. If your resolution is good, you may see a peak in between the two, which might represent small ribosome scanning species. Following the 60S peak, you will see a very large peak corresponding to the monosome, followed by periodic peaks representing sequentially more ribosomes occupying a single mRNA transcript (disome, trisome, etc.). At the end of the gradient, you will see a large accumulation of material that is not periodically separated into peaks and may represent cellular material other than polysomes. *See* Fig. 2a for an example uncut polysome trace from high-quality extract.
24. Once data for the uncut sample is recorded, save the trace. Return the piston to the top position. Remove the finished tube, and replace with a new ultracentrifuge tube full of Milli-Q water. Rinse the tubing by collecting water through the piston, raise the piston again, and drive the piston through the empty tube to move air through the tubing.
25. Repeat **steps 20–24** with your cut samples. You should still expect to see peaks representing the 40S and 60S ribosome subunits, although they may be less distinct than in your "mock" sample. The monosome peak should be much larger than in the "mock" sample, and it often smears into the 60S subunit peak in the "cut" sample. Be careful to collect only

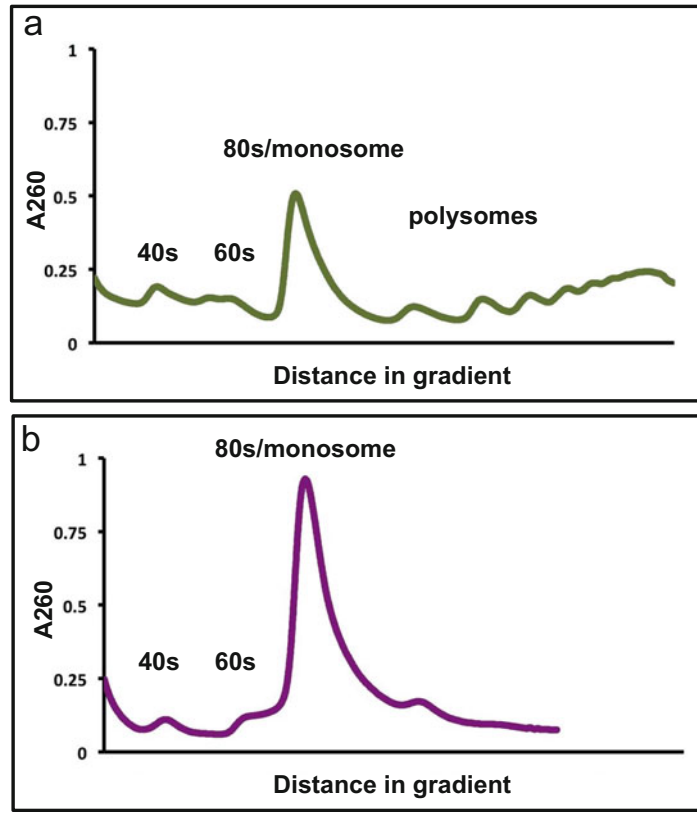

Fig. 2 The A_{260} traces of non-digested and digested cell extract sucrose gradients. (**a**) The A_{260} polysome profile of non-digested meiotic cell extract with high quality. The top (and leftmost) spike of free mRNAs and other cell material is cropped out for clarity. Major ribosome species peaks are labeled. It is possible to see a small peak in between the 40S and 60S peak that we believe may represent scanning 40S subunits with additional initiation factors. (**b**) The A_{260} profile of digested cell extract; major ribosome species peaks are labeled. The polysome peaks in a digested sample should ideally be non-existent. Here there is a small disome peak suggesting a low level of incomplete polysome digestion, which is typical. The monosome peak is very large and runs close to the 60S peak in the digested samples as shown. When collecting the monosome fraction for FP isolation, exclude any 60S or disome peak

material from the monosome and not the 60S peak. *See* Fig. 2b for a "cut" polysome example trace. Collect the monosome fraction into the labeled screw cap 2 mL tubes, and place on ice. Repeat collection for remaining "cut" samples.

26. Flash freeze collected monosome fractions in screw cap tubes, and be sure to not overfill the tubes as the sucrose will expand while frozen and may break the tubes.
27. Be sure to properly clean the gradient station after use. This includes rinsing the tubing with water, drying it by running air

through, and wiping down the entire gradient station surface with a wet paper towel. Any rack, bucket, lid, or ultracentrifuge tube holder used should be thoroughly rinsed with water and air-dried. Turn off all equipment. Move forward to Subheading 3.6 or store monosome fractions indefinitely at −80 °C.

3.6 RNA Extraction from Monosome Fraction (See Note 24)

All steps described are to be carried out at room temperature unless otherwise noted.

1. Heat one thermomixer to 65 °C, and keep one thermomixer at room temperature.
2. Prepare the following sets of labeled tubes using non-stick 1.5 mL tubes: two tubes/sample with 750 μL of acid phenol pre-warmed to 65 °C, two tubes/sample with 40 μL of 20% SDS on ice, two tubes/sample with 700 μL of acid phenol at room temperature, two tubes/sample with 600 μL of chloroform at room temperature, and two tubes/sample with 40 μL of 3 M NaOAc.
3. Thaw monosome fractions on ice, and vortex well before use. Dispense 1.4 mL of each sample into two tubes (700 μL to each), each containing 40 μL of 20% SDS on ice. Heat tubes at 65 °C until the SDS dissolves. Next, add the SDS/sample mixtures to 750 μL of preheated (65 °C) acid phenol. Incubate for 5 min at 65 °C with vigorous mixing. To prevent caps from popping open during incubation, place a flat plastic object (e.g., the lid of a box of micropipette tips) over the sample lids and tape on firmly.
4. Chill samples for 5 min on ice, and then spin for 2 min at 20,000 rcf.
5. Recover the aqueous layer (~600 μL), and add to new tubes containing 700 μL of acid phenol at room temperature. Mix well and incubate for 5 min with vigorous mixing at room temperature. Spin the tubes for 2 min at 20,000 rcf.
6. Recover the aqueous layer (~475 μL), and add to the tube with 600 μL of chloroform; mix well. Incubate for 30 s at room temperature with vigorous mixing. Spin for 2 min at 20,000 rcf.
7. Move the aqueous layer (~360 μL) to a new tube containing 40 μL of 3 M NaOAc, add 1 volume of isopropanol and 2.5 μL of GlycoBlue, and vortex.
8. Chill at −20 °C for at least 30 min, or longer/overnight if desired.
9. Cool a microfuge to 4 °C. Spin samples for 30 min at 20,000 rcf, at 4 °C. After spin, remove the supernatant, and wash the RNA pellet in 750 μL of 80% ice-cold ethanol.

10. Pulse spin samples at 20,000 rcf at 4 °C, and discard the supernatant. Pulse spin to collect any residual ethanol and remove all liquid.
11. Air dry the pellet, and resuspend in 5 μL of 10 mM Tris pH 7. Pipette up and down to fully resuspend the RNA, pulse spin, and combine samples from duplicate tubes. Move forward to Subheading 3.7.

3.7 FP Size Selection

1. Prepare a 15% TBE-Urea gel in 1× TBE, and pre-run gel at 200 V for 15 min (*see* **Notes 5** and **16**). Rinse wells with 1× TBE to remove urea.
2. Add 10 μL of 2× urea sample buffer to each RNA sample, and mix well.
3. Prepare 2 lanes worth of both the 28-mer and 31-mer RNA oligos in 1× urea sample buffer by mixing 2 μL of each 10 μM stock with 20 μL of 2× urea sample buffer and 18 μL of water. Mix 1 μL of the 10 bp ladder with 9 μL of water and 10 μL of 2× urea sample buffer to make 1 lane of ladder (optional).
4. Denature all samples, control oligos, and ladder in 1× urea sample buffer by incubating at 80 °C for 2 min and then placing directly on ice.
5. Pulse spin the samples and load the gel. Always flank each FP RNA sample with a 28-mer and a 31-mer oligo control lane on either side. Do not run actual samples in adjacent lanes as cross-contamination may occur.
6. Run the gel for 65 min at 200 V (*see* **Note 5**). While the gel is running, prepare a 0.5 mL tube for each sample by piercing a hole in the bottom with a 20-gauge needle and nesting this tube into a non-stick 1.5 mL tube. Preheat a thermomixer to 70 °C. Image the gel as described in Subheading 3.8. After imaging, cut out the size-selected footprint samples as described below in **step 7** (*see* **Note 25**).
7. For each lane of ribosome footprint sample, cut out all the material that falls within the size range of ~28–31 nt. Make cuts just below the 28-mer oligo and just over the 31-mer oligo to ensure the full range of FP sizes are collected. Place the excised gel pieces into the labeled 0.5 mL tubes with holes, nested within the 1.5 mL tubes. Excise and extract one of the 28-mer RNA oligo lanes; this can be used as a sizing control later in the protocol. *See* Fig. 3 for an example FP size selection gel. Move forward to Subheading 3.9 and use RNA-specific instructions. Resuspend samples in 10 μL of water and the 28-mer control oligo in 14.5 μL of 10 mM Tris–HCl. After extraction, proceed to Subheading 3.10.

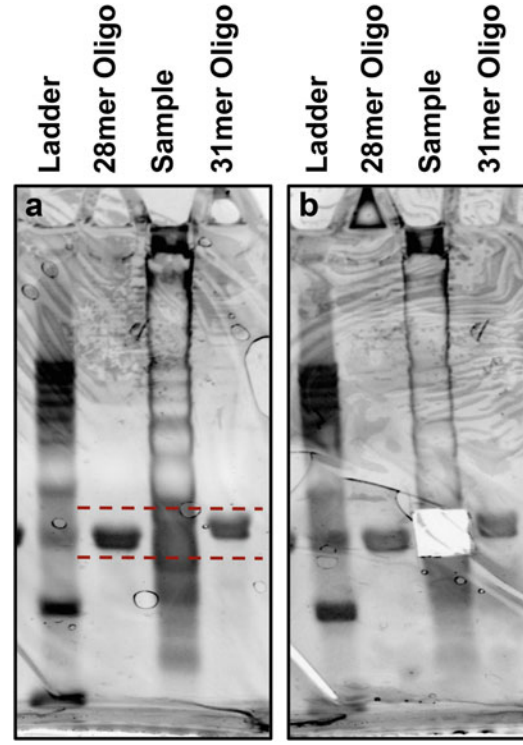

Fig. 3 An example image of a size selection gel for collecting FPs. (**a**) The gel prior to the excision of FP samples is shown. The size range to be collected is marked and encompasses the range between and including the 28-mer and 31-mer RNA oligo controls. Diffuse banding seen along the sample lane represents digested rRNA fragments extracted from the monosome sample. A 10 bp DNA ladder is shown for clarity here, but is unnecessary for FP sizing. (**b**) The size selection gel post-excision of FP sample is shown to highlight the selected sizes

3.8 Gel Staining and Imaging (See Notes 16 and 25)

1. Mix 6 μL of SYBR Gold with 60 mL of 1× TBE (you can reuse the TBE directly from the gel box).
2. Carefully remove the gel from its plastic casing and place into the SYBR Gold and 1× TBE mixture, and incubate with low shaking for 5 min.
3. Place the gel onto clean Saran wrap and image. Carry the gel on the Saran wrap to the light box, and proceed to gel cutting protocol for each specific procedure.

3.9 Gel Extraction (See Note 26)

1. Place excised gel pieces into the 0.5 mL tubes, and spin the nested tubes for 3 min at 20,000 rcf to force the gel through the hole. Repeat spin if gel did not break through on the first attempt. When most of the gel has been broken through the hole, collect any unbroken material in the 0.5 mL tube, and move it to the 1.5 mL tube.

2. Add 700 μL of water to the gel pieces, and incubate for 10 min at 70 °C with vigorous shaking.
3. Use a cut p1000 tip to transfer gel mixture to a centrifuge tube filter column.
4. Centrifuge for 3 min at 20,000 rcf to recover the elution (free of gel debris). Transfer this eluate to a new non-stick tube.
5. Add 2 μL of GlycoBlue, 78 μL of 3 M NaOAc for RNA, or 78 μL of 3 M NaCl for DNA, for precipitation.
6. Add 780 μL of isopropanol. Vortex and precipitate for at least 30 min (overnight works well) at −20 °C.
7. Cool a microfuge to 4 °C, and spin samples at 20,000 rcf for 30 min at 4 °C to pellet the DNA or RNA.
8. Remove the supernatant, and add 1 mL of ice-cold 80% ethanol. Invert several times to wash pellet, and pulse spin at 4 °C and 20,000 rcf.
9. Carefully remove the supernatant, and pulse spin to collect the residual ethanol and remove the remaining liquid. Air dry pellet for about 10 min or for as long as it takes for ethanol to evaporate off.
10. Resuspend in either 10 mM Tris–HCl pH 8 for DNA, or pH 7 for RNA, or water if proceeding to Subheading 3.10, making sure to use the correct volume for the next procedure.

3.10 rRNA Subtraction (See Notes 9, 14, and 28)

Perform rRNA subtraction on all FP samples. It is not necessary to perform rRNA subtraction on the 28-mer oligo control sample, and to minimize reagent costs, we typically do not. Instead, you may place this aside to use as a control in the dephosphorylation and linker ligation steps.

1. Prepare a 1:10 dilution of asDNA2b and asDNA3b by mixing 1.5 μL of each 20 μM stock oligo with 13.5 μL of water (makes 15 μL enough for 2 subtractions). Make all of the MyOne C1 Dynabead solutions described in Subheading 2.4.
2. Add the following to the tubes with 10 μL of RNA samples: 5 μL of asDNA1b (20 pmoles/μL), 6 μL of 1:10 asDNA2b (2 pmoles/μL), 5 μL of 1:10 asDNA3b (2 pmoles/μL), and 3 μL of 20× SSC.
3. Mix well, and incubate samples at 80 °C for 2 min; then place directly back on ice.
4. Add 1 μL of SUPERase·In to each tube and mix.
5. Incubate for 15 min at 37 °C with low shaking, and place back on ice for at least 5 min.
6. While samples are shaking, prepare MyOne Streptavidin C1 Dynabeads by doing the following: vortex the MyOne C1 Streptavidin Dynabeads to resuspend and take 150 μL of

beads/sample into a non-stick tube; wash beads three times in 150 μL of 1× Dynabead B&W buffer using inversion to mix and magnetic rack placement to pellet; wash beads two times in 150 μL of Dynabead Solution A; and wash beads two times in 150 μL of Dynabead Solution B. Finally resuspend beads in 30 μL of 2× Dynabead B&W buffer/sample, and dispense into aliquots of 30 μL in each tube (1 per sample).

7. Pulse spin samples to collect at bottom of tubes, and add all (30 μL) to the Dynabeads. Incubate at room temperature for 15 min with low shaking.
8. Place tubes on the magnet rack, and collect the supernatant (60 μL) to a new 1.5 mL non-stick tube on ice. Add 468 μL of water, 70 μL of 3 M NaOAc, and 2 μL of GlycoBlue. Vortex to mix and add 600 μL of isopropanol; vortex again.
9. Incubate samples at −20 °C for at least 30 min.
10. Pellet RNA by spinning for 20,000 rcf for 30 min at 4 °C, remove supernatant, and wash pellets by adding 1 mL of ice-cold 80% ethanol and by inverting the tubes several times. Pulse spin samples at 4 °C and 20,000 rcf.
11. Remove supernatant carefully, pulse spin samples to remove any remaining ethanol, and air dry pellet until ethanol has evaporated ~10 min. Resuspend pellet in 14.5 μL of 10 mM Tris–HCl pH 7, and proceed to Subheading 3.11.

3.11 Dephosphorylation and Linker Ligation (See Note 27)

1. Resuspend rRNA subtracted RNA FPs and the gel-extracted 28-mer RNA oligo control in 14.5 μL of Tris–HCl pH 7. Move 14 μL of sample to a new non-stick tube for dephosphorylation reactions.
2. Set up a dephosphorylation reaction for each sample and control by adding the following, 2 μL of 10× T4 PNK buffer, 2 μL of (10 U/μL) T4 PNK, and 2 μL of SUPERase·In. Pipette up and down to mix well, and incubate for 1 h at 37 °C, with mixing.
3. Prepare the linker ligation reactions in the same tubes by adding 14 μL of 50% w/v PEG-8000, 2 μL of 10× T4 RNA ligase buffer, 1 μL of 20 μM pre-adenylated linker oligo, 2 μL of T4 Rnl2(tr) K227Q (200 U/μL), and 1 μL of water. Mix well by pipetting up and down, and incubate for 3 h at 22 °C, with mixing.
4. Purify ligations on an Oligo Clean & Concentrator column as described (*see* **Note 8**). Add 10 μL of water to bring sample volume up to 50 μL. Add 100 μL of Oligo Binding Buffer, and mix well. Add 400 μL of 100% ethanol, mix well, and load onto a Zymo-Spin Column in a collection tube.

5. Spin the column at 12,000 rcf for 30 s, and discard the flow-through.
6. Add 750 μL of DNA wash buffer to the column, spin at 12,000 rcf for 30 s, and discard the flow-through.
7. Spin the column for 1 min at 20,000 rcf.
8. Place the column in a new 1.5 mL non-stick tube, and add 10.5 μL of water directly to the column matrix. Spin at 12,000 rcf for 30 s to elute the RNA.
9. Add 10 μL of 2× urea sample buffer to each sample and control, and mix well.
10. Denature all samples and control lanes by incubating at 80 °C for 2 min, and then place back on ice.
11. Prepare a 10% TBE-Urea gel by placing in 1× TBE and pre-running the gel for 15 min at 200 V (*see* **Notes 5** and **16**).
12. Rinse urea out of the wells with 1× TBE, and load the gel, 20 μL/lane. Do not place different samples in adjacent lanes; use a control sizing oligo or an empty lane between them. Run the gel at 200 V for 50 min (*see* **Note 5**). While the gel is running, prepare a 0.5 mL tube for each sample by piercing a hole in the bottom with a 20-gauge needle and placing this tube into a non-stick 1.5 mL tube. Preheat a thermomixer to 70 °C. Image the gel as described in Subheading 3.8. After imaging, cut out the linker-ligated footprint samples as described in **step 13**.
13. Excise the larger band representing samples that now have a linker. There will be a large lower band with un-ligated linker that should be avoided. Use the linker-ligated 28-mer RNA oligo as a sizing guide if the samples are of low concentration and hard to visualize on the gel. Cut out and gel-extract the linker-ligated 28-mer control oligo, and use as a positive control and sizing guide for Subheading 3.12.
14. Proceed to Subheading 3.9 and follow directions for RNA samples. Resuspend the pellets in 10.5 μL of 10 mM Tris–HCl pH 7, and proceed to Subheading 3.12.

3.12 Reverse Transcription (See Note 14)

Use the linker-ligated 28-mer control as a positive control and sizing guide for the reverse transcription steps.

1. Take 10 μL of precipitated RNA samples and controls in 10 mM Tris–HCl pH 7, and move to a new 1.5 mL non-stick tube.
2. Add the following to each tube, 3.28 μL of 5× FS buffer, 0.82 μL of 10 mM dNTPs, and 0.5 μL of the 20 μM RT primer, and mix well.
3. Denature samples for 2 min at 80 °C, and place directly on ice.

4. Add 0.5 μL of SUPERase·In and 0.82 μL of 0.1 M DTT to each sample, and mix well.
5. Add 0.82 μL of Superscript III to each tube. Incubate at 48 °C for 30 min with low mixing.
6. Add 1.8 μL of 1 M NaOH to each tube, and mix well to hydrolyze any remaining RNA template.
7. Incubate at 98 °C for 20 min with low mixing. Use a thermomixer with a thermo top to prevent samples from evaporating out of the tube. GlycoBlue will turn pink.
8. Add 1.8 μL of 1 M HCl, and mix well to neutralize samples.
9. Place tubes on ice, and add 20 μL of 2× urea sample buffer.
10. Denature samples at 95 °C for 3 min prior to loading, and then place on ice. Note that each sample will be split between two gel lanes (20 μL each lane).
11. Pre-run a 10% TBE-Urea gel at 200 V for 15 min (*see* **Note 5**). Rinse the wells with 1× TBE before loading to remove urea.
12. Load the gel with 20 μL/lane. Do not place different samples in adjacent wells, leave an empty well, or run a sizing oligo between them.
13. Run samples for 65 min at 200 V, and proceed to stain and visualize gel as described in Subheading 3.8 (*see* **Note 5**). While the gel is running, prepare a 0.5 mL tube for each sample by piercing a hole in the bottom with a 20-gauge needle and placing this tube into a non-stick 1.5 mL tube. Preheat a thermomixer to 70 °C.
14. Use the 28-mer RNA oligo that has been linker-ligated and carried through the RT reaction as a sizing guide. Cut out and gel-extract the band corresponding to RT elongated product, and avoid any lower bands corresponding to leftover RT primer or non-ligated linker. The footprint bands from each sample can run slightly higher than the control, as footprints include a mixture of sizes, with 28 nt as the lowest size selected. *See* Fig. 4 for an example of a gel showing samples after reverse transcription and indicating which material to collect from the gel.
15. Proceed to Subheading 3.9 and use directions for precipitating DNA. Following extraction, resuspend pellets in 15.5 μL of 10 mM Tris–HCl pH 8, and proceed to Subheading 3.13.

3.13 Circularization (See *Note 14*)

1. After resuspending pellets from RT gel extraction in Tris–HCl pH 8, move 15 μL of each sample to PCR tubes.
2. Prepare circularization reactions by adding 2 μL of CircLigase II 10× reaction buffer, 1 μL of 1 mM ATP, and 1 μL of 50 mM $MnCl_2$ to each sample, and mix well.

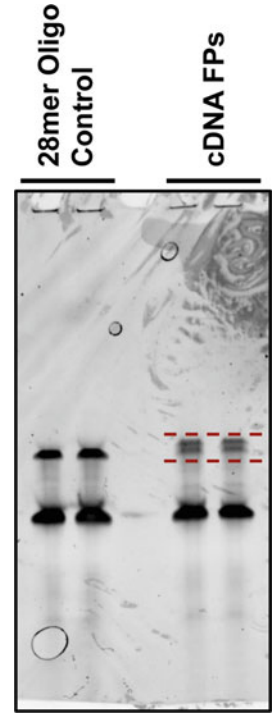

Fig. 4 An example image of a size selection gel to collect linker-ligated and reverse transcribed FP samples. The 28-mer oligo lanes represent 28-mer oligo that was dephosphorylated, ligated to the linker, and reverse transcribed as a positive control and a sizing guide. FP samples can run slightly larger than the 28-mer control and occasionally show banding. The size range that should be excised from the gel is marked and includes the larger sets of bands in each lane and roughly matches the size of the 28-mer control

3. Add 1 μL of CircLigase II to each sample, and mix well.
4. Incubate at 60 °C for 1 h and then 80 °C for 10 min to heat inactivate the enzyme. Put tube on ice if proceeding directly to Subheading 3.14 or store at −20 °C indefinitely.

3.14 PCR Amplification (See Notes 14, 29 and 30)

1. Prepare PCR reactions for each sample by mixing the following in a PCR tube: 3.34 μL of 5× HF Phusion buffer, 0.34 μL of 10 mM dNTPs, 0.8 μL of 10 μM PCR F primer, 0.8 μL of 10 μM barcoding primer (*Specific to each sample*), 10.4 μL of water, and 0.16 μL of Phusion polymerase.
2. Pipette 1 μL of each circularized sample into the corresponding PCR tube, and mix well.
3. Perform PCR with the following steps: (a) 98 °C for 30s, (b) 98 °C for 10s, (c) 60 °C for 10s, (d) 72 °C for 10s, and repeat **steps b–d** for the desired number of cycles for each sample, and (e) hold at 4 °C indefinitely.

4. Add 3.4 μL of 6× DNA loading dye to each PCR tube, and mix well.
5. Prepare a 8% TBE gel in 1× TBE buffer, and pre-run the gel for 15 min at 180 V (*see* **Note 5**).
6. Load samples onto the gel (20 μL/sample), and run the gel for 55 min at 180 V (*see* **Note 5**). While the gel is running, prepare a 0.5 mL tube for each sample by piercing a hole in the bottom with a 22-gauge needle and placing this tube into a non-stick 1.5 mL tube. Preheat a thermomixer to 70 °C.
7. Proceed to Subheading 3.8. Then excise the amplified sequencing libraries as described below in **step 8**.
8. Excise the PCR amplified sequencing library bands as described. *See* Fig. 5 and reference notes for directions on how to select the proper number of PCR cycles to amplify each sample library. Excise the discrete bands that correspond to the size of the major product in the 28-mer control lane. Leave behind any large smeary bands and bands corresponding to excess primers or empty vector that has been carried through the library preparation and shows up in smaller bands (*see* Fig. 5).
9. Proceed to Subheading 3.9 and follow directions for a DNA sample; resuspend pellets in 10 μL of 10 mM Tris–HCl pH 8. Proceed to Subheading 3.15.

3.15 TapeStation Analysis for Quality Control Assessment (See Note 10)

1. Use an Agilent Bioanalyzer or TapeStation to assess the quality of the sequencing libraries prior to submission. We use the Agilent High Sensitivity D1000 ScreenTape and TapeStation to quantify and observe the sizes of our sequencing libraries as described below.
2. Allow all reagents to equilibrate at room temperature for 30 min, and vortex buffer and ladder before use.
3. Make a 1:2 dilution of each sample by mixing 1.5 μL of sample with 1.5 μL of 10 mM Tris–HCl pH 8.
4. Mix 2 μL of the High Sensitivity D1000 Sample Buffer with 2 μL of the High Sensitivity D1000 Ladder. Prepare samples by mixing 2 μL of the 1:2 dilution with 2 μL of the High Sensitivity D1000 Sample Buffer.
5. Spin down samples and ladder, and then vortex at 2000 rpm for 1 min.
6. Pulse spin to collect sample at the bottom of the tube.
7. Load samples into the Agilent 4200 TapeStation instrument, and place the ladder in position A1.
8. Select the required sample positions on the 4200 TapeStation Controller Software.

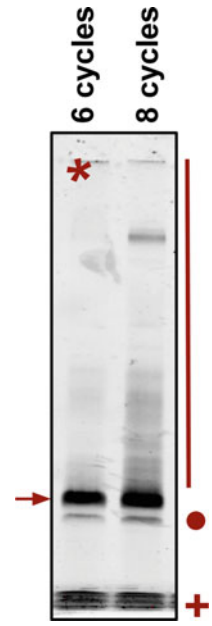

Fig. 5 An example image of a size selection gel to isolate PCR amplified and multiplexed samples. Two lanes are shown with the same sample amplified for different numbers of cycles. The desired product is the brightest band and is marked with an arrow. For this sample, six cycles of amplification provided robust linear amplification with low levels of undesirable background products, as shown by the higher amounts of smear (marked with a red line) in the lane with eight cycles of amplification. Take care to also exclude any material representing amplified empty vector (shown with a red circle) or excess primer from the amplification step (shown with a red + sign)

9. Specify the position of the ladder, click Start, and specify a filename with which to save the results.
10. If samples look as expected, submit to your sequencing core following their submission guidelines. Sequence samples using a 50 nt single read run on an HiSeq4000. Use data for your preferred downstream analysis (*see* **Note 17**).

4 Notes

1. Only use cycloheximide that is certified for treatment of live cells. We use the Biotechnology Performance Certified cycloheximide from Sigma.
2. Choose a glass filtration apparatus that will yield rapid filtration of the entire sample collected and the proper size of corresponding filter membranes. We use a 90 mm glass filtration unit from Fisher Scientific and 90 mm cellulose nitrate filter membranes from Whatman.

3. Our protocol uses the SW 41 Ti ultracentrifuge rotor and the Gradient Master gradient station from BioComp. We find the Seton polyclear ultracentrifuge tubes the most reliable in this system; however, other tubes may be used as long as they are compatible with the SW 41 Ti rotor and the gradient station used. We use the BioRad EM-1 Economonitor as the UV monitor at our gradient station; other monitors may be suitable if they are compatible with the gradient station used.
4. Use of a coprecipitant is optional, but can be necessary if sample quantity is low. We prefer to use GlycoBlue as a coprecipitant, as it not only improves precipitation efficiency but also allows for easy visualization of small nucleic acid RNA pellets.
5. We use 8 ×8 cm pre-cast gels that fit in the Life Technologies mini gel tank. The recommended voltage and length of run listed in our directions reflects conditions that yield good separation in this context. If an alternate size of gel is used, the voltage or length of run needed to achieve good separation may vary.
6. Other high sensitivity nucleic acid stains could work; we prefer SYBR Gold because it is highly sensitive and can be used to rapidly stain gels after electrophoresis.
7. For most enzymes needed in this protocol, we specified the exact enzyme and provider we typically use. Other suppliers of the same or similar enzymes may work in some cases, but enzyme preparations vary substantially between companies, and we have not verified any alternatives to those listed here.
8. We use the Oligo Clean & Concentrator kit from Zymo Research to purify and concentrate linker-ligated mRNA footprints prior to gel electrophoresis size selection. Other kits or methods that will reliably purify and elute small RNA fragments may be suitable for this purpose. If an alternate kit is used, follow the corresponding manufacturer's protocol.
9. Magnetic streptavidin beads are used in this protocol to remove rRNA fragments by pulling down biotinylated antisense rRNA oligos. We use the MyOne Streptavidin C1 Dynabeads (and corresponding recommended buffers) as they are optimized for use with nucleic acids and have a high binding capacity. Other types of magnetic streptavidin beads may be suitable for this purpose; however, we have not confirmed the efficiency of rRNA pull-downs using alternate beads and buffers.
10. We use an Agilent TapeStation for library quality control. If another means of quality control is used, substitute the listed tape station reagents with reagents suitable for your preferred quality control analysis and follow the manufacturer recommended instructions.

11. We use the Retsch MM400 to cryogenically mill ribosome profiling samples. Similar cryogenic mills should also work as long as they can fit similar volumes of sample. Directions here are for the Retsch MM400 and may not be applicable to other instruments.
12. Much of the power of ribosome profiling relies on its ability to distinguish whether changes to gene expression are occurring as a result of transcriptional or translational control. In order to determine this, one must perform an mRNA sequencing (RNAseq) experiment in parallel. We describe how to collect a matched sample for RNAseq in the cell harvesting step of this protocol. In our experience, extracting RNA directly from a frozen cell pellet (rather than from the extract used for RNAse treatment and ribosome profiling) minimizes the 3′ end bias that can occur when mRNA is PolyA-selected directly from cell extract. Our RNAseq protocol uses hot acid phenol extraction of RNA, followed by PolyA selection to enrich the sequencing sample for mRNAs, rather than rRNAs and tRNAs which are highly abundant. We next use alkaline-based mRNA fragmentation and select mRNA fragments that have a similar size distribution to FPs, by cutting a slice from a 15% TBE-Urea gel. These mRNA fragments are then carried through the same library preparation protocol as described in detail here for FP samples (dephosphorylation, linker ligation, reverse transcription, circularization, and PCR amplification/multiplexing). The only modification of these steps of the protocol to process mRNA rather than FP samples is the approach for enrichment of the fragments of choice (rRNA subtraction vs PolyA selection), and thus library prep can be done in parallel for RNAseq and FP samples. We note that PolyA selection is known to produce bias towards mRNAs with longer PolyA tails and also produces an enrichment for 3′ mRNA ends for samples, due to baseline degradation. It is possible (and often valuable) to perform total RNA sequencing in the absence of PolyA selection to avoid bias; however, this will result in a vast majority of sequencing reads (>90%) being dedicated to ribosomal and transfer RNAs, rather than mRNA. rRNA depletion kits are an alternative option, although these will also produce some biases in sequencing data yielded.
13. Ribosome profiling relies on taking bulk measurements of the translation occurring within a large population of cells at a given time. During a meiotic experiment, it is important to validate that the population at the time of sample collection largely represents the cellular state you wish to measure. Part of this consideration includes confirming that the population is well synchronized during the experiment. Lack of synchrony during a meiotic experiment can lead to measurements of gene

expression that are inaccurate or misleading, as the total expression levels will represent an average of each gene's expression over the spread of cellular states present in the population, rather than the measurement of each gene's expression during one particular state. There are many ways to confirm the cellular state and assess synchrony during a meiotic experiment, and staging methods can vary from those that give only broad information about meiotic completion (e.g., by assessing the dynamics of tetrad formation by light microscopy) to specialized approaches for detailed assessments of the morphology of a specific cellular compartment. One very reliable method used to determine the synchrony and staging of a meiotic culture at discreet time points uses DAPI staining to visualize chromosomal DNA. This allows one to assess when cells undergo the two chromosomal divisions. To use this method, take samples at defined intervals during the meiotic experiment. Fix the samples (most fixation methods will work; we use formaldehyde fixation), and then permeabilize the membranes (most permeabilization methods will work; we use an alcohol- or detergent-based method). Add DAPI stain and visualize on a fluorescent microscope. As the cells progress through meiosis, they will undergo two nuclear divisions. You can determine what percentage of the population is within each stage of meiosis as determined by how many nuclear masses are visualized and the shape of those masses. To gain information on the stages of meiosis prior to the nuclear divisions, or more detailed analysis of the stages during the nuclear divisions, fix samples in accordance with your favorite immunofluorescence microscopy protocol and stain for beta-tubulin and DAPI. The meiotic spindle goes through discreet morphological changes prior to the nuclear first division occurring, and this approach provides also more detailed assessment of nuclear division stages (e.g., anaphase I, metaphase II, etc.).

14. Many of the steps described include adding small volumes of multiple components to the same tube at the same time. It can be helpful to make a master mix in some of these cases, but if you do, you will want to make master mix for an extra sample so that you do not run out due to loss of material on pipette tips during transfer.
15. As discussed in the introduction to this method, pre-treatment of cells with cycloheximide is optional. Collect cells with no cycloheximide pre-treatment if, for example, quantitative codon-specific positional information within gene bodies is critical to downstream analyses.
16. Preventing RNAse degradation of samples (post RNAse I digestion) is of extreme importance. Before running or imaging any gels with RNA samples, we suggest rinsing the gel chamber with an RNAse decontaminant (e.g., RNAse AWAY).

17. For more instructions on how to process and analyze the resulting sequencing data from your experiment, *see* Ingolia et al. (2012) [13].
18. Synchronous sporulation requires a shift to post-diauxic growth prior to sporulation induction; it is best that overnight cultures grow to saturation (>10 OD_{600} in YPD, and >5 OD_{600} in BYTA). Sporulation requires respiration, and sporulation media includes a non-fermentable carbon source while lacking nitrogen and fermentable carbon sources. It is important that there is no carryover of nutrients from the BYTA culture, so the yeast must be washed well with sterile water prior to resuspension in SPO. Thaw yeast from frozen stocks on glycerol plates to select for and confirm cells are competent for respiration prior to use. It is also important to ensure the yeast are properly aerated prior to and during sporulation. To do this, grow all cultures in flasks that can hold at least 10× the volume of the actual culture, and confirm shaking is at least 280 rpm for small orbital and 260 rpm for large orbital shakers. For meiosis experiments, it is useful to confirm the culture sporulated homogeneously using light microscopy. To check the staging at time of sample collection, fixed samples can be analyzed for DAPI staining or immunofluorescence. For vegetative exponential phase growth cultures, it is important that the yeast have time to recover from their saturated state after overnight growth. Allow cultures to double at least three times on day of collection to ensure cells have fully transitioned back to exponential growth.
19. Note that exponentially growing cells in rich media contain the most ribosomes and overall translation levels of any conditions known in yeast. This results in polysome profiles that contain more disomes, trisomes, etc. in "mock" (non-RNAse-digested) samples than seen in other conditions. Meiotic cells have a translation rate of ~50% that of exponentially growing cells (Brar lab, unpublished data), and thus polysome levels are low compared to monosome levels. It is important to have an idea of the "normal" polysome traces expected for "mock" samples under a given condition to assess quality of samples to be used for ribosome profiling. Also note that samples with lower translation levels will result in fewer ribosome footprints than exponentially growing cells, and thus bands on gels for these samples may be faint and amounts of oligos and enzymes added to reactions may need to be adjusted if FP quantities differ dramatically from the types of samples described here. It is also possible to collect more cells in conditions with lower translation to address the issue of low FP quantity.
20. It is important to work efficiently and quickly after CHX addition such that each sample has the same length of CHX

treatment. Minimizing the time spent collecting cells will also help to avoid any gene expression changes that are secondary effects from CHX treatment, or from cells being removed from the culture media. Be careful not to rip the membrane as you are collecting cells following filtration as this will substantially slow your ability to scrape the yeast off. We recommend rapidly scraping most of the yeast (90%) into a single pile, rapidly scraping that pile into the LN_2 with the large spatula, and then collecting the remaining patch (10%) into the 2 mL tube with another spatula, followed by immediate flash freezing. Note that polysome lysis buffer must be transferred in drops to samples ahead of time, such that all LN_2 has evaporated off before mixer milling but that samples containing polysome lysis buffer pellets can be stored indefinitely at −80 °C. We recommend using 2 mL of polysome lysis buffer for cell collections of the quantities described here. However, you may adjust the volume of lysis buffer and the number of cells collected to make the concentration of the resulting extract higher or lower, based on analysis of extract from a pilot experiment.

21. Cryomilling requires a lot of LN_2, so it is best to start with a full styrofoam box as well as a backup dewar in case you need to refill the box during breaking. It is important that samples do not thaw during breaking. To prevent this, everything that will touch the samples is precooled in LN_2. Wear two sets of latex or nitrile gloves underneath a pair of cryo-gloves to protect your hands from the caustic temperature of LN_2 and exposed materials. Closed chambers should not be outside of LN_2 for more than 5 min at a time, and open chambers should not be out of LN_2 for more than 3 min. When collecting powder, it is important to collect as much as you can as fast as you can. Do not continue to collect small amounts of extract if the chamber has been outside of LN_2 for more than 3 min – directions here include material in excess, such that gathering most of the cell powder will provide a sufficient amount of extract for multiple aliquots. If it is difficult to get enough cell powder out of the chambers using only 2 mL of polysome lysis buffer, increase the lysis buffer and/or cell number as necessary to increase material or use smaller (10 mL stainless steel cannisters). Small spatulas with curved ends work well for collecting cell powder. Chambers can become very tight from rotation during breaking cycles; to prevent chambers from locking shut, loosen and re-tighten the chamber each time it is removed or placed into LN_2. It is also important to keep the chamber halves in their matched pairs. If only one sample is to be lysed, place an empty chamber (no ball) on the opposing side to balance the mill.

22. Keep thawed extract on ice as much as possible, and minimize the overall time the extract is thawed until it is flash frozen. After the final spin, there will be a pellet, a hazy layer just above the pellet, and a top layer of unwanted material floating on the top of the supernatant. Be careful to collect only the middle layer of supernatant representing cleared cell extract, and minimize collection of other materials. Pipetting slowly and using multiple rounds of smaller volume collections works best. We collect ~75% of the volume of our input, and leave behind material to avoid collecting non-desired material. To calculate the number of A_{260} units present in each sample aliquot, first multiply the diluted measurement by 100. This represents the A_{260} units that would be present for 1 mL of undiluted sample. Each aliquot is 200 μL, so multiply the A_{260} units present in 1 mL of sample by 0.2. An example calculation is shown below. Sample 1 (1:100) = 1.3 A_{260} units/mL.

    Sample 1 = 1.3 A_{260} * 100 = 130 A_{260} units/mL.

    Sample 1 (aliquot) = (130 A_{260} units/mL) * 0.200 mL = 26 A_{260} units.

23. While preparing the ultracentrifuge tubes, check each tube for and discard any tubes with cracks. Once the sucrose gradients are made, take great care to minimize the amount of movement they are subjected to, and keep them in a stable rack. While fractionating, it is important that the collection tip of the gradient station holds a full seal and is not mixing the gradient as it descends. It is helpful to remove the collection tip, rinse it, and press it firmly against a smooth flat surface to extend the sides prior to collecting each sample. The total volume of the digested monosome peak will be larger than needed for making a ribosome profiling library and may be more than 2 mL; it is useful however to collect the entire fraction and keep it frozen as backup material.

24. The volume of the digested monosome peak is too large to extract in a single 1.5 mL tube; the samples should be very well mixed and can be aliquoted into smaller-scale RNA extractions (two extractions of 700 μL monosome fraction are usually more than enough material). Leftover monosome fraction can be re-frozen and kept as backup material in case something fails downstream. Hot acid phenol extractions release gas and pressure will build in the tubes during mixing. Opening tubes periodically to release pressure can prevent tubes from popping open during the extraction. Work in the hood, with goggles, and change gloves frequently to protect yourself from hot acid phenol. Wear all recommended personal protective equipment, and follow all regulations for use of hot phenol and disposal of phenol-chloroform mixtures. The aqueous layer containing RNA should always be on top, using this extraction protocol.

25. The same gel staining and imaging process is used throughout this protocol four times. For all gels, limit the amount of time that passes after running the gel, until cutting out the desired bands, as diffusion can occur. Place gels on clean Saran wrap for ease of imaging. For any gels prior to PCR amplification (footprint size selection, post-linker ligation, and post-RT), do not run different samples adjacent to each other. Always run an empty lane or a control oligo in between. This in combination with using new single-use scalpels to cut out each sample will help to prevent cross-contamination. Post-PCR amplification, when unique barcodes are included for each sample, samples can be run adjacent to each other and cut with the same scalpel. It is useful to excise and gel-extract one lane of the 28-mer control RNA oligo from the FP size selection gel. This oligo can be carried through the subsequent enzymatic library preparation steps as a control sample and can additionally serve as a sizing guide during each subsequent gel extraction. Always record a new image of the gel post-excision, or make a marking on the original image to show the material that was taken through to the next steps of the protocol.
26. This procedure is used four times in the protocol with only minor modifications. The first variation is the size of the hole used to break the gel: use a 20-gauge needle for the 15% and 10% TBE-Urea gels and a 22-gauge needle for the 8% gels. The second is the salt used to precipitate the extracted materials. For any step prior to reverse transcription, follow directions for RNA. For any step after reverse transcription, follow directions for DNA. The third is the solution used to resuspend the final pellet: for DNA pellets (post-RT), use 10 mM Tris–HCl pH 8; for RNA pellets (prior to RT), use 10 mM Tris–HCl pH 7.
27. It is possible to use either gel-extracted or fresh 28-mer RNA oligo as a control for dephosphorylation and linker ligation. To use fresh oligo, mix 1 μL of 10 μM oligo with 13 μL of water. To use gel-extracted oligo, treat as the rest of the samples. Do not add ATP to the reaction. T4 PNK will dephosphorylate ATP and phosphorylate itself instead of dephosphorylating the FPs if ATP is added.
28. This protocol is optimized to remove three rRNA fragments found in great abundance in ribosomal footprint samples in the original ribosome profiling paper [1]:

    RDN25 734-760 (Here referred to as "rRNA#1", 5′AAGAG GTGCACAATCGACCGATCCTG3′).

    RDN25 2502-2528 (Here referred to as "rRNA#2", 5′TAG TTTCTTTACTTATTCAATGAAGCGG3′).

    RDN25 3167-3193 (Here referred to as "rRNA#3", 5′AATA TAGATGGATACGAATAAGGCGTC3′).

rRNA subtraction will greatly lower the amount of rRNA present in the final footprint samples for sequencing; however, it is common to still have between 30 and 50% of the total "FP" sequencing reads align to rRNA. This pull-down protocol was optimized on untailed, radiolabeled RNA to maximize rRNA pull-down and minimize non-specific loss of material. Modifications such as increasing the concentration of biotinylated antisense oligos, variation in hybridization buffer conditions, variations in hybridization temperature, increased ratio of Dynabeads to biotinylated antisense oligos, decreased volume of hybridization/Dynabead incubations, and the absence of Tween in the Dynabead buffers did not yield any increases to rRNA subtraction in our hands.

29. Mix circularized samples thoroughly prior to use. Choose a unique indexing primer for each sample; this will allow you to sequence all the samples on the same sequencing lane. Keep detailed notes of which sample gets which index. Estimate the number of amplification cycles you will use for each sample based on the amount of material seen in the RT gel. If samples were bright on the reverse transcription gel, they are likely concentrated and will not need many cycles of PCR (try six or eight cycles). If the samples are dim on the gel, try 10 or 12 cycles. It is important to try at least two amplification cycle choices per sample to ensure that product is amplifying linearly. After imaging the gel, re-amplify any samples that have large non-desired products with fewer cycles. Re-amplify any samples in which the main product band is not bright using more cycles. *See* Fig. 5 for example images.
30. If samples are outside the range of quantification with a 1:2 dilution, try a higher dilution and reanalyze. Footprint samples should have a single major peak corresponding to the size of the full length of any FPs plus the extensions of the linker, RT primer, and PCR primers. If the size is not correct, re-amplify and extract again. Do not submit samples for sequencing if they do not pass the quality control analysis at this step.

Acknowledgments

We thank Nick Ingolia, the Ingolia lab, and Ze Cheng for technical assistance and helpful conversations. We thank Calvin Jan for sharing validated linker and oligo sequences.

References

1. Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324:218–223. https://doi.org/10.1126/science.1168978
2. Ingolia NT, Hussmann JA, Weissman JS (2019) Ribosome profiling: global views of translation. Cold Spring Harb Perspect Biol 11:a032698. https://doi.org/10.1101/cshperspect.a032698
3. Brar GA, Weissman JS (2015) Ribosome profiling reveals the what, when, where and how of protein synthesis. Nat Rev Mol Cell Biol 16:651–664. https://doi.org/10.1038/nrm4069
4. Brar GA, Yassour M, Friedman N et al (2012) High-resolution view of the yeast meiotic program revealed by ribosome profiling. Science 335:552–557. https://doi.org/10.1126/science.1215110
5. Cheng Z, Otto GM, Powers EN et al (2018) Pervasive, coordinated protein-level changes driven by transcript isoform switching during meiosis. Cell 172:910–923.e16. https://doi.org/10.1016/j.cell.2018.01.035
6. Mitchell AP (1994) Control of meiotic gene expression in Saccharomyces cerevisiae. Microbiol Rev 58:56–70
7. Neiman AM (2011) Sporulation in the budding yeast Saccharomyces cerevisiae. Genetics 189:737–765. https://doi.org/10.1534/genetics.111.127126
8. Berchowitz LE, Gajadhar AS, van Werven FJ et al (2013) A developmentally regulated translational control pathway establishes the meiotic chromosome segregation pattern. Genes Dev 27:2147–2163. https://doi.org/10.1101/gad.224253.113
9. Carlile TM, Amon A (2008) Meiosis I is established through division-specific translational control of a cyclin. Cell 133:280–291. https://doi.org/10.1016/j.cell.2008.02.032
10. Berchowitz LE, Kabachinski G, Walker MR et al (2015) Regulated formation of an amyloid-like translational repressor governs gametogenesis. Cell 163:406–418
11. Jin L, Zhang K, Xu Y et al (2015) Sequestration of mRNAs modulates the timing of translation during meiosis in budding yeast. Mol Cell Biol 35:3448–3458. https://doi.org/10.1128/mcb.00189-15
12. Hollerer I, Higdon A, Brar GA (2018) Strategies and challenges in identifying function for thousands of sORF-encoded peptides in meiosis. Proteomics 18:1700274. https://doi.org/10.1002/pmic.201700274
13. Ingolia NT, Brar GA, Rouskin S et al (2012) The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. Nat Protoc 7:1534–1550. https://doi.org/10.1038/nprot.2012.086
14. Ingolia NT (2016) Ribosome footprint profiling of translation throughout the genome. Cell 165:22–33
15. Santos DA, Shi L, Tu BP, Weissman JS (2019) Cycloheximide can distort measurements of mRNA levels and translation efficiency. Nucleic Acids Res 47:4974–4985. https://doi.org/10.1093/nar/gkz205
16. Duncan CDS, Mata J (2017) Effects of cycloheximide on the interpretation of ribosome profiling experiments in Schizosaccharomyces pombe /631/337 /631/337/574 /38/91 / 38/39 article. Sci Rep 7:1. https://doi.org/10.1038/s41598-017-10650-1
17. Cheng Z, Brar GA (2019) Global translation inhibition yields condition-dependent de-repression of ribosome biogenesis mRNAs. Nucleic Acids Res 47:5061–5073. https://doi.org/10.1093/nar/gkz231
18. Gerashchenko MV, Gladyshev VN (2014) Translation inhibitors cause abnormalities in ribosome profiling experiments. Nucleic Acids Res 42:e134. https://doi.org/10.1093/nar/gku671
19. Hussmann JA, Patchett S, Johnson A et al (2015) Understanding biases in ribosome profiling experiments reveals signatures of translation dynamics in yeast. PLoS Genet 11: e1005732. https://doi.org/10.1371/journal.pgen.1005732
20. Requião RD, de Souza HJA, Rossetto S et al (2016) Increased ribosome density associated to positively charged residues is evident in ribosome profiling experiments performed in the absence of translation inhibitors. RNA Biol 13:561–568. https://doi.org/10.1080/15476286.2016.1172755
21. Lareau LF, Hite DH, Hogan GJ, Brown PO (2014) Distinct stages of the translation elongation cycle revealed by sequencing ribosome-protected mRNA fragments. elife 2014: e01257. https://doi.org/10.7554/eLife.01257
22. Lecanda A, Nilges BS, Sharma P et al (2016) Dual randomization of oligonucleotides to reduce the bias in ribosome-profiling libraries.

Methods 107:89–97. https://doi.org/10.1016/j.ymeth.2016.07.011

23. Weinberg DE, Shah P, Eichhorn SW et al (2016) Improved ribosome-footprint and mRNA measurements provide insights into dynamics and regulation of yeast translation. Cell Rep 14:1787–1799. https://doi.org/10.1016/j.celrep.2016.01.043
24. Kearse MG, Goldman DH, Choi J et al (2019) Ribosome queuing enables non-AUG translation to be resistant to multiple protein synthesis inhibitors. Genes Dev 33:871–885. https://doi.org/10.1101/gad.324715.119
25. McGlincy NJ, Ingolia NT (2017) Transcriptome-wide measurement of translation by ribosome profiling. Methods 126:112–129. https://doi.org/10.1016/j.ymeth.2017.05.028

Chapter 5

Quantitative Comparisons of Translation Activity by Ribosome Profiling with Internal Standards

Yinuo J. Wang and Wendy V. Gilbert

Abstract

Ribosome profiling is a genome-wide approach to map the positions of ribosomes on messenger RNAs. The abundance of ribosome-protected fragments can be used within condition to compare relative translation activities between different transcripts and between distinct conditions for the same transcript. A unified and routine method is currently lacking, however, to normalize between conditions for differences in global translation levels. Here we describe experimental and computational methods to use an orthogonal species spike-in, or internal standard, to enable absolute comparisons of translation activity between conditions. This simple modification of standard ribosome profiling provides a robust approach for accurately interpreting the effects of diverse genetic, chemical, and environmental perturbations of translation.

Key words Ribosome profiling, Translation, Spike-in normalization, Next-generation sequencing

1 Introduction

Translation is an essential node in the regulation of protein levels. Rapid remodeling of translation in response to environmental cues is a well-known phenomenon in microbes, both in terms of bulk activity and gene-specific adaptation [1–3]. In metazoans, growth and nutrient signals converge on translation by modulating the activity of key translation initiation factors. For example, the target of rapamycin complex, which phosphorylates 4E-binding protein to control the availability of the cap-binding initiation factor eIF4E, globally increases cap-dependent translation in the presence of abundant nutrients [4, 5]. Nutrient availability also affects functional levels of initiator tRNA in the ternary complex through the general control pathway [6, 7]. Moreover, hyperactive translation is a hallmark of most cancer cells as is overexpression of the majority of translation initiation factors [8]. Thus, a full understanding of the translation program between different biological conditions must account for differences in global translation activity.

Vyacheslav M. Labunskyy (ed.), *Ribosome Profiling: Methods and Protocols*, Methods in Molecular Biology, vol. 2252, https://doi.org/10.1007/978-1-0716-1150-0_5,

There are a number of tools to quantify bulk translation. Classic examples include sucrose gradient sedimentation of translationally active polyribosomes followed by quantitation of polysome versus monosome areas under the curve [9], metabolic labeling with radioisotope-containing amino acids [10], and puromycin labeling followed by western blotting with an anti-puromycin antibody [11]. Recent techniques are also based on detection of amino acid analogs incorporated during elongation but offer insight beyond traditional ensemble measurements (reviewed in [12]). Such bulk measurements reveal varying degrees of translational inhibition in response to distinct cellular stresses including heat shock, nutrient deprivation, viral infection, and disruption of protein folding in the endoplasmic reticulum. Likewise, differences in the magnitude of bulk translational repression following genetic depletion or pharmacological inactivation stratify translation factors, with general functions attributed to factors whose loss leads to large reductions. Nevertheless, it is well established that specific mRNAs respond differently to depletion of "general" translation factors [5, 6, 13, 14]. Furthermore, adaptive cellular responses to stress require mRNA-specific translational activation under conditions of bulk translational repression [1–3].

Ribosome profiling is a powerful technique to illuminate changes in mRNA-specific translation activity by providing a genome-wide snapshot of the locations of ribosomes on mRNAs with single-nucleotide resolution [15]. By combining ribosome profiling with total mRNA abundance measurements by RNA-seq it is possible to measure the contribution of translational control to gene expression changes and to estimate protein output per mRNA [16, 17]. However, most ribosome profiling studies are limited to comparing relative footprint abundances between conditions, without an inherent feature that captures global translation levels. This limitation leads to nonintuitive results such as apparent translational activation of hundreds of mRNAs following inhibition of general translation initiation factors. For example, depletion of eIF4G1 [18] or eIF4A1 [19] in yeast would be expected to reduce translation of most ORFs yet leads to an apparent increase in ribosome density on gene bodies downstream of short transcript leaders. Similarly, inhibition of the mTOR1-eIF4E-eIF4G axis by Torin also causes an apparent increase in ribosome density on a broad set of mRNAs, despite a strong expectation of a net inhibitory effect on ribosome loading [5].

Several reports have used normalization techniques to address this issue (reviewed in [20]). Han et al. [21] published the first use of an internal standard in a study of co-translational folding, in which they combined ribosome profiling with epitope capture of the emerging nascent chain. Following RNase digestion, total monosomes or samples enriched for nascent epitopes were RNA-extracted and mixed with a single 28-nucleotide spike-in RNA. This approach in principle allows absolute comparisons of

footprint mRNA levels between samples, although relying on a single sequence precludes analysis of how reproducibly the standard itself was captured across libraries. Popa et al. [22] addressed this issue by fragmenting a 4.5 kilobase T7 transcription product to 30 nucleotides. They added this standard in a fixed ratio to the mass of total RNA isolated from pelleted monosomes. In this case, the standards benchmark the total ribosome content of each partially processed sample but do not account for technical variability from sedimentation and RNA extraction. To quantify absolute translation changes induced by rocaglamide A and other initiation inhibitors, Iwasaki et al. [23] normalized cytosolic ribosome footprints to total mitochondrial ribosome footprints that co-purified within each sample. The key advantage of this method is that primary and spike-in RNAs are co-processed from the earliest possible stage, i.e., cell lysis. However, it is unclear what component of standard ribosome profiling buffers biochemically stabilizes the association between mRNAs and mitochondrial ribosomes, which are insensitive to cycloheximide [24]. Moreover, the method is not generalizable to comparisons between growth conditions or treatment regimens that perturb mitochondrial translation. In summary, each of these normalization strategies deviates significantly from the ideal procedure, which should be robust to variability in library capture, independent from the biological perturbation, and maximizes the number of steps during which the primary sample is co-processed with standards.

Here we present a technique for using a complex lysate of human footprints as an internal standard for yeast ribosome profiling libraries (Fig. 1) [25]. The spike-in is added at a fixed percentage relative to the total number of ribosomes per sample, thereby enabling direct comparisons of the overall fraction of ribosome-engaged footprints between samples. We show how to accurately quantify ribosome content by gel electrophoresis of ribosomal RNA (rRNA) extracted from a defined volume of lysate. The standard is pooled with primary sample immediately prior to 80S isolation on a gradient, which maximizes co-processing while keeping the standards separate from the biological perturbation in question. The method is readily adapted to any pair of eukaryotic species with sufficient genome divergence to uniquely assign reads of 15–34 nucleotides.

2 Materials

2.1 Preparation of HeLa Lysate

1. HeLa cells, ATCC CCL-2.
2. Eagle's Minimum Essential Medium (EMEM) + 10% fetal bovine serum (FBS).
3. Cell culture incubator with 5% CO_2 set to 37 °C.
4. 15 cm dishes, tissue-cultured treated.

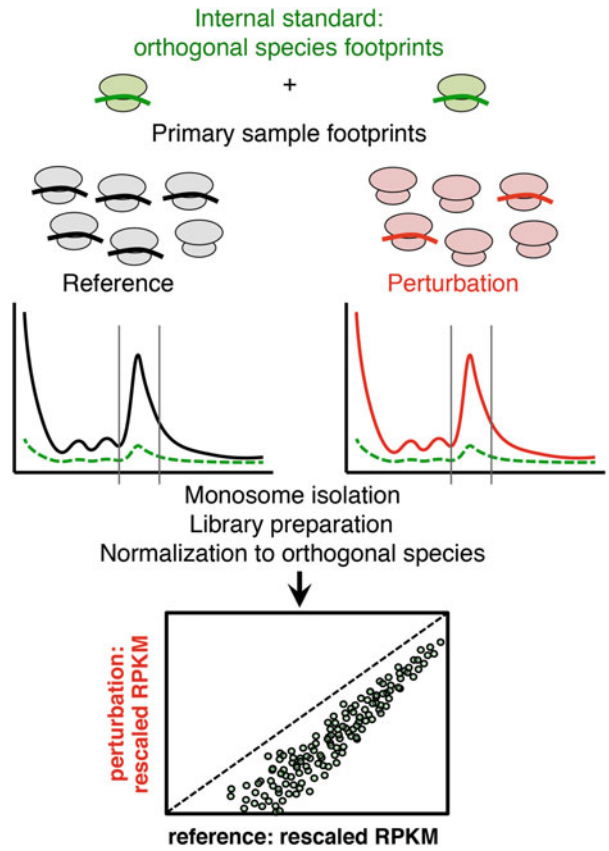

Fig. 1 Use of orthogonal species footprints as internal standards allows quantitative comparisons of translation activity by ribosome profiling. The internal standard is added at a fixed ratio to the total number of ribosomes in each primary sample. Following normalization to the orthogonal reads, the primary reads then reflect the total fraction of translating ribosomes

5. Phosphate-buffered saline (PBS).
6. Cell scrapers.
7. Sterile 1 mL syringes and 26 gauge needles.
8. HeLa lysis buffer: 20 mM Tris–HCl pH 7.5, 10 mM $MgCl_2$, 0.2 M KCl, 1% Triton X-100, 0.2 mg/mL cycloheximide (add fresh), 4 mM DTT (add fresh), 2× EDTA-free protease inhibitors (add fresh).
9. HeLa gradient buffer: 20 mM HEPES-KOH pH 7.4, 5 mM $MgCl_2$, 0.1 M KCl, 2 mM DTT (add fresh), 0.1 mg/mL cycloheximide (add fresh).
10. Automated gradient mixer.
11. SW41 tubes.
12. Beckman SW41 rotor and ultracentrifuge.
13. Automated gradient fractionator.

2.2 Preparation of Saccharomyces cerevisiae Lysate

1. Whatman cellulose nitrate 0.45 μm membrane filter.
2. Footprint lysis buffer: 20 mM Tris–HCl pH 8.0, 140 mM KCl, 5 mM $MgCl_2$, 1% Triton X-100, 0.1 mg/mL cycloheximide (add fresh).
3. Yeast gradient buffer: 20 mM Tris–HCl pH 8.0, 150 mM KCl, 5 mM $MgCl_2$, 0.5 mM DTT (add fresh), 0.1 mg/mL cycloheximide (add fresh).
4. Isopropanol.
5. Liquid nitrogen and dewar.
6. Cryogenic grinder with liquid nitrogen connection.
7. Vacuum pump with glassware to collect filtrate and a reservoir.
8. NanoDrop or microvolume UV-Vis spectrophotometer.
9. Automated gradient mixer.
10. SW41 tubes.
11. Beckman SW41 rotor and ultracentrifuge.
12. Automated gradient fractionator.

2.3 Quantification of Ribosomal RNA

1. AES buffer: 50 mM sodium acetate pH 5.2, 10 mM EDTA, 1% SDS.
2. Acid phenol (pH 6.0).
3. Phenol/chloroform/isoamyl alcohol (25:24:1, pH 6.7).
4. Chloroform.
5. 2 mL gel phase lock tubes.
6. 3 M sodium acetate pH 5.2.
7. Isopropanol.
8. Glycoblue.
9. TE buffer: 10 mM Tris–HCl pH 8.0, 1 mM EDTA.
10. Synergel.
11. Molecular biology grade agarose.
12. 0.5× TBE buffer: 50 mM Tris–HCl, 50 mM boric acid, 1 mM EDTA.
13. Ethidium bromide.
14. Formamide.
15. 10× RNA loading buffer without dye: 50% glycerol, 5 mM EDTA (pH 8.0).
16. NanoDrop or microvolume UV-Vis spectrophotometer.
17. Quantitative gel imager, such as the Typhoon FLA 9500.
18. Densitometry analysis software, such as ImageJ.

2.4 RNase Digestion, Addition of Internal Standard, and Footprint Isolation from Sucrose Gradients

1. RNase I (100 U/μL) (*see* **Note 1**).
2. Yeast gradient buffer: 20 mM Tris–HCl pH 8.0, 150 mM KCl, 5 mM $MgCl_2$, 0.5 mM DTT (add fresh), 0.1 mg/mL cycloheximide (add fresh).
3. 20% SDS.
4. Acid phenol (pH 6.0).
5. Phenol/chloroform/isoamyl alcohol (25:24:1, pH 6.7).
6. Chloroform.
7. 3 M sodium acetate (pH 5.2).
8. Isopropanol.
9. Glycoblue.
10. Automated gradient mixer.
11. SW41 tubes.
12. Beckman SW41 rotor and ultracentrifuge.
13. Automated gradient fractionator.

2.5 Small RNA Enrichment, Dephosphorylation, and Footprint Size Selection

1. Zymo-Spin V column.
2. Guanidine hydrochloride (Gu HCl) buffer: 8 M guanidine hydrochloride, 20 mM MES hydrate, 20 mM EDTA.
3. T4 polynucleotide kinase (PNK) (10 U/μL).
4. RNasin Plus.
5. 15mer RNA size marker: 5′-AUGUACACGGAGUCG-3′.
6. 28mer RNA size marker: 5′- AUGUACACGGAGUC GACCCGCAACGCGA-3′.
7. 34mer RNA size marker: 5′- AUGUACACGGAGUCGAG CACCCGCAACGCGAAUG-3′.
8. 2× formamide-bromophenol blue (FB) loading buffer: 95% formamide, 20 mM EDTA (pH 8.0), 0.05% (w/v) bromophenol blue, 0.05% (w/v) xylene cyanol.
9. Commercial or freshly prepared 15% polyacrylamide/8.0 M urea/0.5× TBE gel.
10. SYBR Gold stain.
11. RNA elution buffer: 0.3 M sodium acetate pH 5.5, 1 mM EDTA, 100 U/mL RNasin Plus (add fresh).
12. Isopropanol.
13. Glycoblue.
14. Spin-X centrifuge filter.

2.6 3′ Adapter Ligation and Ribosomal RNA Depletion

1. 100 μM pre-adenylated DNA adapter, 5′-rAppTGGAATTCT CGGGTGCCAAGG/3ddC/−3′.
2. T4 RNA ligase I (10 U/μL).
3. 10× T4 RNA Ligase I buffer.
4. PEG-8000.
5. Small RNA ladder.
6. 2× FB loading buffer: 95% formamide, 20 mM EDTA (pH 8.0), 0.05% (w/v) bromophenol blue, 0.05% (w/v) xylene cyanol.
7. Commercial or freshly prepared 10% polyacrylamide/8.0 M urea/0.5× TBE gel.
8. SYBR Gold stain.
9. Ribo-Zero Gold rRNA removal kit (yeast) (*see* **Note 2**).
10. 3 M sodium acetate (pH 5.2).
11. Isopropanol.
12. Glycoblue.
13. Optional: Qubit RNA High Sensitivity assay kit.
14. NanoDrop or microvolume UV-Vis spectrophotometer.

2.7 cDNA Synthesis and Circularization

1. 10× RT buffer: 0.5 M Tris–HCl pH 8.0, 0.6 M NaCl, 0.1 M DTT. Store at −20 °C.
2. 5 μM RT primer with Unique Molecular Identifier (UMI): 5′-/5Phos/N_{10} GATCGTCGGACTGTAGAACTCTGAACCT GTCGGTGGTCGCCGTATCATT /iSp18/CACTCA/iSp18/ GCCTTGGCACCCGAGAATTCCA.
3. 0.24 M $MgCl_2$.
4. 25 mM dNTPs.
5. RNasin Plus.
6. AMV reverse transcriptase (10 U/μL).
7. 1.0 M NaOH.
8. 1.0 M HCl.
9. 2× FB loading buffer: 95% formamide, 20 mM EDTA (pH 8.0), 0.05% (w/v) bromophenol blue, 0.05% (w/v) xylene cyanol.
10. Commercial or freshly prepared 10% polyacrylamide/8.0 M urea/0.5× TBE gel.
11. SYBR Gold stain.
12. DNA elution buffer: 0.3 M NaCl, 1 mM EDTA, 10 mM Tris–HCl pH 8.0.
13. Low molecular weight DNA ladder.
14. CircLigase ssDNA ligase.

2.8 PCR and Preparation for Sequencing

1. 10 μM forward PCR primer: 5′-AATGATACGGCGACCACCGA.
2. Q5 High-Fidelity 2× Master Mix.
3. 10 μM TruSeq Small RNA PCR Index Primer: 5′-CAAGCAGAAGACGGCATACGAGAT [6 bases] GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA, where [6 bases] are Illumina RPI1-48 [26].
4. 50 bp DNA ladder.
5. Freshly prepared 8% polyacrylamide/0.5× TBE gel (no urea).
6. SYBR Gold stain.
7. DNA elution buffer: 0.3 M NaCl, 1 mM EDTA, 10 mM Tris–HCl pH 8.0.

3 Methods

3.1 Preparation of HeLa Lysate

1. Grow early passage HeLa cells in a cell culture incubator with 5% CO_2 at 37 °C. Use EMEM +10% FBS as growth medium.
2. The day before harvest, plate cells in 15 cm dishes such that they attain 50-70% confluency the following day. Set up at least one 15 cm dish per 5 primary libraries, plus one additional 15 cm plate for an analytical polysome profile.
3. Aspirate media and wash cells with 37 °C PBS.
4. Add 1.0 mL ice-cold HeLa lysis buffer to the plate. Rapidly and vigorously scrape cells into the buffer. Pipet lysate over the plate to wash and then tilt the plate to collect lysate. Transfer to a 1.5 mL tube and keep on ice (*see* **Note 3**).
5. Pass lysate ten times through a 26-gauge needle to shear genomic DNA.
6. Spin lysate at 20,000 × *g* for 10 min at 4 °C. Pool all supernatant into a single tube and mix gently.
7. Set aside an aliquot of ~5 A260 units for an analytical polysome profile and ~1 A260 unit for extraction and NanoDrop of total RNA (*see* Subheading 3.3 **steps 1–6**). Aliquot remaining lysate into aliquots of ~3 A260 units, flash freeze in liquid nitrogen, and store at −80 °C. Excess lysates prepared in a large batch can be stored in this way for use in multiple experiments.
8. Verify polysome capture by fractionating ~5 A260 units of lysate through a sucrose gradient (**steps 9** and **10** below). Poor polysome capture can result in an unreliable representation of human footprint reads in the final libraries.

9. Prepare linear 10–50% sucrose gradients with HeLa gradient buffer in SW41 tubes. Load lysate and spin in a Beckman SW41 rotor at 35,000 rpm (151,263 × *g*) for 3 h at 4 °C.
10. Fractionate gradients and visualize the ribosomal species, including polysomes, by continual monitoring of absorbance at 260 nm.

3.2 Preparation of Saccharomyces cerevisiae Lysate

1. Grow *S. cerevisiae* cultures to desired state for profiling.
2. Collect cells by rapid filtration as follows to minimize perturbations to physiological translation (*see* **Note 4**). Pierce holes in the cap of a 50 mL tube, and submerge the open tube into liquid nitrogen. Also pre-chill metal spatulas in liquid nitrogen. Add 10 mL isopropanol to the vacuum pump reservoir to avoid foaming. Start the vacuum line, and pour <500 mL of cell culture onto the Whatman filter. It will take ~50 s for the media to elute. When the cells retained on the filter have a creamy consistency, rapidly scrape using the pre-chilled spatulas into the prepared 50 mL tube with liquid nitrogen.
3. Into the same tube, drip 1.0 mL footprint lysis buffer for every 350 OD_{600} units of cells. Store samples at −80 °C until cryogrinding.
4. Prepare extract by cryogrinding frozen cells with a ball mill under continuous liquid nitrogen flow. Run ten cycles of the following program: 60 s of grinding at 10 Hz, 60 s of cooling.
5. Thaw extracts on ice.
6. Centrifuge lysate at 3000 × *g* for 5 min at 4 °C to remove debris.
7. Take supernatant to a new tube. Centrifuge at 21,000 × *g* for 10 min at 4 °C.
8. Take supernatant to a new tube. Check the yield by NanoDrop of a 1:10 dilution.
9. Divide supernatant into new tubes: approximately 10 A260 units for an analytical polysome profile, 50 A260 units for library preparation, 10 A260 units for quantification of total rRNA (*see* Subheading 3.3). Any lysates not processed immediately can be flash frozen in liquid nitrogen and stored at −80 °C.
10. Verify polysome capture by fractionating ~10 A260 units of lysate through a sucrose gradient (**steps 11** and **12** below). If polysome to monosome ratios are variable between replicates or different than expected, growth and harvest (**steps 1–8**) may need to be repeated.

11. Prepare linear 10–50% sucrose gradients with yeast gradient buffer in SW41 tubes. Load lysate, and spin in a Beckman SW41 rotor at 35,000 rpm (151,263 × *g*) for 3 h at 4 °C.
12. Fractionate gradients, and visualize the ribosomal species, including polysomes, by continual monitoring of absorbance at 260 nm.

3.3 Quantification of Ribosomal RNA

1. Thaw primary and spike-in lysate reserved for RNA quantification (from Subheading 3.2, **step 9** and Subheading 3.1, **step 7**, respectively). Divide into three samples of equal volume for three technical replicates of total RNA extraction.
2. Add AES to a final volume of 500 μL for each sample. Add 500 μL of acid phenol and vortex thoroughly (*see* **Note 5**).
3. Heat samples at 65 °C with vortexing or vigorous shaking for 20 min. Place samples on ice for 5 min.
4. Transfer contents to a gel phase lock tube, and perform the following extractions: 1× chloroform, 2× phenol/chloroform/isoamyl alcohol, and 1× chloroform. Be sure to transfer the aqueous phase completely at each step.
5. Transfer the aqueous phase to a new tube, and precipitate the RNA (*see* **Note 6**). Resuspend RNA pellets in TE.
6. NanoDrop the purified RNA to determine yield. Use this value to calculate the amount of total RNA per unit volume of lysate.
7. The following steps apply only to the primary samples. Prepare standards by mixing purified RNA from all samples and then creating a dilution series of at least five points. Alternatively, a dilution series of total RNA from exponential yeast can also serve as standards (*see* **Note 7**).
8. Using the NanoDrop readings as guides, prepare for gel loading a quantity of total RNA that falls within the range of the standard curve: add an equal volume of formamide, and then add 10× RNA loading buffer to a final concentration of 1×. No heat denaturation is required before loading.
9. Run standards and samples on a 0.7% agarose/0.9% Synergel/0.5× TBE/1× ethidium bromide gel until the 25S and 18S rRNAs are clearly resolved (*see* **Note 8**) (Fig. 2a). Image the gel, and quantify the 25S and 18S intensities of each standard and sample by densitometry. Ensure that the raw sample intensities fall within the linear range of the standard curve (Fig. 2b). Use the sum of 25S and 18S intensities to calculate the amount of rRNA (in arbitrary units) per unit volume of lysate (Fig. 2c).

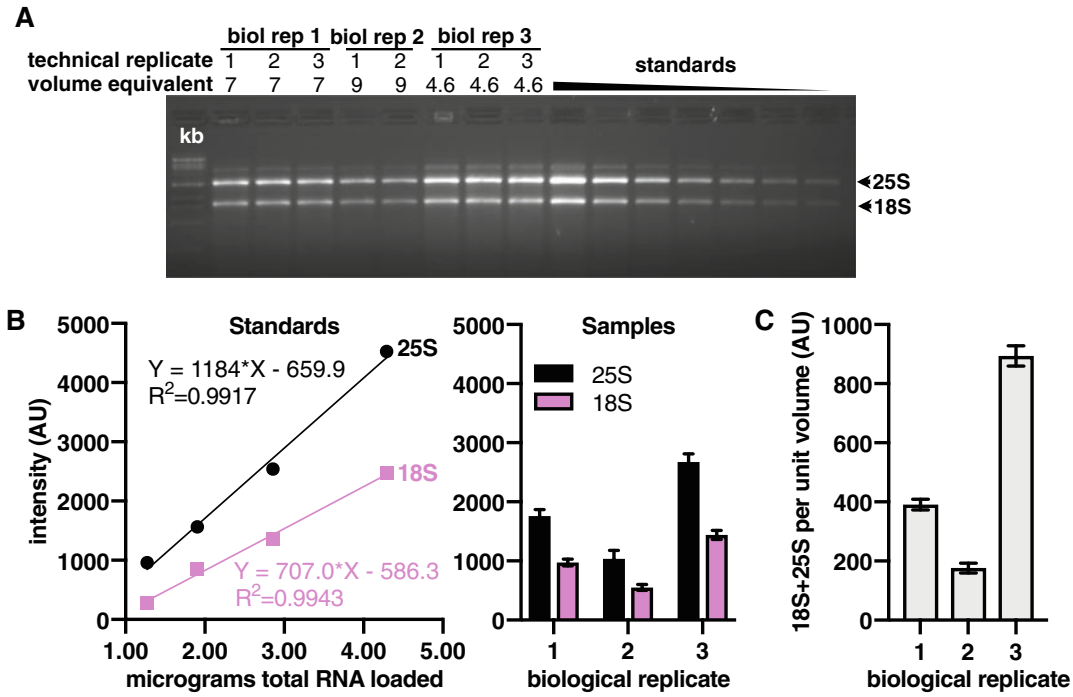

Fig. 2 Quantification of total rRNA content in primary samples. (**a**) Total RNA was extracted from defined volumes of lysate and separated on a 0.7% agarose/0.9% Synergel/0.5× TBE/1× ethidium bromide gel. Biological replicates refer to yeast cultures that were grown and harvested independently. Technical replicates are aliquots of lysate taken from the same biological replicate and subjected to total RNA extraction in parallel. A 1.5-fold dilution series of total RNA from exponential yeast was run in parallel as standards. (**b**) (Left) Quantification of 25S and 18S band intensities in the standards. (Right) Verification that sample intensities fell within the linear range of quantification. Mean ± S.D. (**c**) Calculations of rRNA per unit volume of lysate used to determine the amount of spike-in addition. Mean ± propagated S.D.

3.4 RNase Digestion, Addition of Internal Standard, and Footprint Isolation from Sucrose Gradients

1. Calculate how much HeLa lysate is required for each sample by starting with the primary sample that is most translationally active. Using the total RNA measurements from Subheading 3.3, **step 6**, the HeLa lysate should be spiked in at ≥2% of the total RNA in this sample. The amount of HeLa lysate required for every other primary sample is then adjusted according to the respective rRNA content, such that the ratio of HeLa lysate/primary rRNA is identical across all samples.
2. Set up yeast RNase I digestions with 15 U per each A260 unit of lysate (*see* **Note 9**). Mix in thermomixer at 400 rpm, 25 °C for 60 min. Place on ice.
3. Transfer the necessary amount of HeLa lysate plus at least 10% surplus to a new tube (*see* **Note 10**). Add 0.5 U of RNase I per microliter of lysate. Mix in thermomixer at 400 rpm, 25 °C for 30 min. Place on ice.

4. Add digested HeLa lysate to each primary sample as calculated in **step 1**. Mix well by pipetting.
5. Prepare linear 10–50% sucrose gradients with yeast gradient buffer in SW41 tubes. Load samples, and spin in a Beckman SW41 rotor at 35,000 rpm (151,263 × *g*) for 3 h at 4 °C.
6. Fractionate gradients with continual monitoring of A260. Pool the 80S fractions.
7. Add SDS to a final concentration of 1.1%. Mix with an equal volume of acid phenol, and incubate at 65 °C with vigorous vortexing for 20 min. Place samples on ice for 5 min.
8. Perform the following extractions: 1× chloroform, 2× phenol/chloroform/isoamyl alcohol, and 1× chloroform.
9. Transfer the aqueous phase to a new tube, and precipitate the RNA (*see* **Note 6**). Half the sample can be stored in precipitation solution at −20 °C as backup. Precipitate the other half, and resuspend RNA pellets in 10 μL 10 mM Tris–HCl pH 7.0 to prepare for footprint size selection.

3.5 Small RNA Enrichment, Dephosphorylation, and Footprint Size Selection

1. **Steps 1–7** enrich for small RNAs by taking the eluate of a first column and capturing it on a second column. Add 250 μL of Gu HCl buffer to the resuspended RNA. Add 125 μL of 100% ethanol and mix (*see* **Note 11**).
2. Add sample to a Zymo-Spin V column with 100 μg capacity. Spin for 1 min at room temperature at 12,000 × *g*.
3. Discard the column. Add 460 μL of 100% ethanol to the flow-through and mix.
4. Add sample to a new Zymo-Spin V column. Spin for 1 min at room temperature at 12,000 × *g*.
5. Add 800 μL of room temperature 80% ethanol to the column. Spin for 1 min. Discard the flow-through. Repeat this wash step.
6. Spin for 2 min to remove residual ethanol.
7. Transfer the column to a new 1.5 mL tube. Add 400 μL of H_2O to the column, and let stand for 5 min. Spin for 1 min to elute the RNA.
8. Precipitate the RNA (*see* **Note 6**), and resuspend in 10 μL 10 mM Tris–HCl pH 7.0 for dephosphorylation.
9. Transfer 9.5 μL RNA to a PCR tube. Add 0.5 μL of RNasin Plus, 1.25 μL of 10× PNK buffer, and 1.25 μL of PNK.
10. Incubate for 1 h at 37 °C.
11. Prepare a 15% polyacrylamide/8.0 M urea/0.5× TBE gel. Pre-run at 200 V for 20 min.

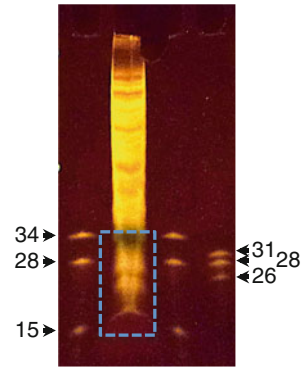

Fig. 3 Size selection of footprint RNA from 15–34 nt on a 15% polyacrylamide/8 M urea gel. Various RNA size markers are indicated. The region boxed in blue was excised for library construction

12. Prepare size selection markers: For each well, combine the 15, 28, and 34mer RNA oligos at 0.5 μM each in 10 μL total volume (*see* **Note 12**). Add 10 μL of 2× FB loading buffer.
13. Add 12.5 μL of 2× FB loading buffer to each dephosphorylation reaction.
14. Denature all samples at 75 °C for 2 min. Place on ice.
15. Load libraries flanked by size selection markers on each side. Run the gel for 65 min at 200 V, or until the bromophenol blue reaches the bottom.
16. Submerge gels in 0.5× TBE/1× SYBR Gold. Stain for 5 min with gentle orbital shaking.
17. Visualize gels using Epi Blue illumination (*see* **Note 13**). Excise libraries from the bottom of the 15mer marker to the top of the 34mer marker (Fig. 3).
18. Pierce the bottom of a 0.5 mL tube using an 18 gauge needle, and nest within a 1.5 mL tube. Transfer the gel slice to the 0.5 mL tube. Centrifuge at max speed for 3 min to force the gel through the needle hole, and crush the gel slice (*see* **Note 14**).
19. Remove the 0.5 mL tube. Add 0.4 mL of RNA elution buffer to the 1.5 mL tube.
20. Place in thermomixer set to 70 °C, and incubate with shaking at 1500 rpm for 20 min.
21. Cut the end of a P1000 pipet tip to a wide slant, and transfer the gel solution to a Spin-X column. Spin at 16,000 × *g* for 3 min.
22. Transfer flow-through to a new 1.5 mL tube and precipitate (*see* **Note 6**). Resuspend in 7 μL H_2O for 3′ adapter ligation.

3.6 3′ Adapter Ligation and Ribosomal RNA Depletion

1. Take 1.0 μL of RNA to a new tube. Dilute and measure the concentration by NanoDrop or Qubit.
2. Transfer 5–20 picomoles of RNA to a PCR tube. Add water to a final volume of 6.5 μL.
3. Add the following components to each reaction: 1.0 μL RNasin Plus, 0.5 μL of 100 μM pre-adenylated adapter, 1.0 μL T4 RNA ligase buffer without ATP, 3.0 μL PEG-8000, and 1.0 μL T4 RNA ligase I (*see* **Notes 15** and **16**).
4. Incubate at 22 °C for 5 h.
5. **Steps 5–7** will purify ligated from unligated footprints. Mix the reaction with 2× FB loading buffer and load on a pre-run 10% polyacrylamide/8.0 M urea/0.5× TBE gel. Also load 1.0 μL of unligated RNA for comparison and the Abnova small RNA ladder. Run at 200 V until bromophenol blue is at the bottom.
6. Submerge gels in 0.5× TBE/1× SYBR Gold. Stain for 5 min with gentle orbital shaking. Excise the ligated RNA, and purify according to Subheading 3.5, **steps 18–21** (*see* **Note 17**).
7. Transfer filtered eluate to a new 1.5 mL tube and precipitate (*see* **Note 6**). Resuspend in 14 μL of H_2O for ribosomal RNA depletion.
8. Wash rRNA depletion magnetic beads in bulk by transferring 112.5 μL of beads per library to a 1.5 mL nonstick tube. Let beads stand on magnetic rack for at least 1 min.
9. Remove the supernatant. Remove tube from rack, and wash with nuclease-free water in a volume equal to the volume of slurry in **step 8**. Pipet beads up and down to resuspend.
10. Take off the water and repeat the wash. Remove the water from the second wash.
11. Add 30 μL of the Magnetic Bead Resuspension Solution for every 112.5 μL of the original slurry in **step 8**. Mix by pipetting up and down until the clumps are dispersed.
12. Aliquot 32.5 μL of resuspended slurry to a new 1.5 mL nonstick tube for each library.
13. Add 0.5 μL of RiboGuard RNase Inhibitor to each tube and vortex briefly. Keep at room temperature.
14. Add the following components individually to a new tube in the following order: 2 μL Reaction Buffer, 13 μL RNA, 2 μL rRNA Removal Solution. Pipet to mix and incubate at 68 °C for 10 min.
15. Incubate at room temperature for 5 min.
16. Add mixture to the washed beads from **step 13**. Pipet up and down ten times immediately. Vortex for 10 s on medium speed, taking care to avoid splashing beads onto the lid.

17. Incubate at room temperature for 5 min.
18. Vortex for 10 s on medium speed, and then incubate at 50 °C for 5 min with a heated lid.
19. Immediately place the tubes on a magnetic rack. Let stand for at least 1 min.
20. Transfer the supernatant to a new tube. Precipitate the RNA (*see* **Note 6**), and resuspend in 7.0 μL of 10 mM Tris–HCl pH 8.0 for cDNA synthesis.

3.7 cDNA Synthesis and Circularization

1. Take 6.2 μL of RNA to a PCR tube. Add 1.0 μL of 5.0 μM RT primer and 0.8 μL of 10× RT buffer. In addition, set up a no-template control reaction. Anneal the primer using the following program: 65 °C 2 min, 55 °C 2 min, 45 °C 2 min, 42 °C 2 min. Place on ice.
2. Add extension reagents to each tube: 0.6 μL 10× RT buffer, 2.24 μL 25 mM dNTPs, 1.16 μL 0.24 M $MgCl_2$, 1.0 μL RNasin Plus. Add 1.0 μL AMV RT individually to each tube.
3. Incubate at 42 °C for 60 min.
4. Place tubes on ice. Add 1.5 μL of 1.0 M NaOH to hydrolyze RNA.
5. Incubate at 98 °C for 15 min.
6. Place on ice and add 1.5 μL of 1.0 M HCl to neutralize. Add 16 μL of 2× FB loading buffer to each reaction to prepare for gel size selection of RT extension products.
7. Mix 0.5 μL DNA ladder with 9.5 μL H_2O and 10 μL 2× FB loading buffer for each lane. Load a pre-run 10% polyacrylamide/8 M urea/0.5× TBE gel with each library sample divided between two lanes and the DNA ladder flanking each pair of library lanes. Run at 200 V until xylene cyanol is at the bottom.
8. Submerge gels in 0.5× TBE/1× SYBR Gold. Stain for 5 min with gentle orbital shaking. The no-template control reactions should show the RT primer running around 90–100 base pairs. Excise from just above the primer band to the bottom of the 150 base pair marker. Purify according to Subheading 3.5, **steps 18–21**, but using DNA elution buffer.
9. Transfer the eluate to a new 1.5 mL tube and precipitate (*see* **Note 6**). Resuspend in 16 μL 10 mM Tris–HCl pH 8.0 for circularization.
10. Transfer 15 μL cDNA to a PCR tube. Add 2 μL 10× CircLigase buffer, 1 μL 1 mM ATP, 1 μL 50 mM $MnCl_2$, and 1 μL 0.5× CircLigase (0.5 μL CircLigase diluted in 0.5 μL 1× CircLigase buffer) (*see* **Note 18**). Incubate at 60 °C for 6 h followed by 80 °C for 10 min. Store at −20 °C until ready for PCR.

3.8 PCR and Preparation for Sequencing

1. Prepare small-scale PCR reactions to identify the optimal cycle number. These typically range from 5 to 14 cycles. For two test reactions, add 8.35 μL 2× Q5 master mix, 0.84 μL PCR forward primer, 0.84 μL indexing reverse primer, 5.67 μL H_2O, and 1.0 μL cDNA template to a new tube. Mix and dispense 8.0 μL into each destination tube.
2. Run the following program: 98 °C 30 s, [98 °C 10 s, 60 °C 20 s, 72 °C 20 s]. Repeat steps in brackets for desired cycle number.
3. Add DNA loading dye to each sample. Load samples and 500 ng of 50 bp ladder on an 8% polyacrylamide/0.5× TBE gel (no urea). Run at 200 V for 55 min or until the xylene cyanol is 2–3 cm from the bottom.
4. Submerge gels in 0.5× TBE/1× SYBR Gold. Stain for 5 min with gentle orbital shaking. The 100 bp marker should be at the bottom of the gel. No-insert cassettes are 122 bp, while the desired library products should be ~137–156 bp. Identify the earliest cycle number at which PCR products are visible.
5. Prepare large-scale PCR reactions: 25 μL 2× Q5 master mix, 2.52 μL 10 μM forward PCR primer, 2.52 μL 10 μM reverse indexing primer, 16.96 μL H_2O, 3 μL cDNA. Run thermocycler program as in **step 2**. Repeat gel running as in **steps 3** and **4**. Each PCR reaction will need to be divided between two lanes. Maintain an empty well between pairs of sample lanes to prevent cross-contamination.
6. Excise above the 122 bp no-insert cassette up to ~160 bp (Fig. 4). Elute gel slice overnight at room temperature in 450 uL of DNA elution buffer or 750 uL if the gel slice is large.
7. Precipitate the elution buffer (*see* **Note 6**), and resuspend DNA pellet in 15 μL nuclease-free H_2O.

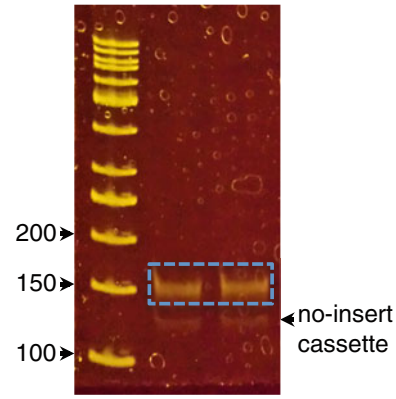

Fig. 4 Size selection of PCR library amplicons on an 8% polyacrylamide native gel. The region boxed in blue was excised and purified for sequencing

8. Quantify library concentrations by Fragment Analyzer or qPCR. Pool at the appropriate concentration and sequence on an Illumina Sequencer (*see* **Note 19**).

3.9 Data Analysis

Libraries were analyzed using custom Bash and Python 2.7 scripts, all of which are available upon request.

1. Following library-level demultiplexing into individual fastq files, collapse PCR duplicates using the FASTX-Toolkit command `fastx_collapser -Q33 -o`. Remove the 3′ adapter and 5′ decamer UMI sequences using `cutadapt --adapter TGGAATTCTCGGGTGCCAAGG --cut 10 --overlap 3 --minimum-length 15`.
2. Use STAR `--runMode genomeGenerate` to create a joint reference of the *S. cerevisiae* and human genomes. Note that chromosome names cannot be shared between the two species. Map processed reads from **step 1** using `STAR --runMode alignReads --alignEndsType EndToEnd --outFilterMismatchNoverLmax 0.05 --outSAMunmapped Within --outFilterMultimapNmax 1 --outFilterMultimapScoreRange 1 --outSAMattributes All --outStd BAM_Unsorted --outSAMtype BAM Unsorted --outFilterType BySJout --outFilterScoreMin 15 --outReadsUnmapped Fastx --outSAMattrRGline ID:foo`. Only uniquely mapped reads will be retained.
3. Sort and index the output BAM file (*see* **Note 20**), and then split into yeast- and human-only BAM files using `samtools view -b -L chromosomes.bed`, where chromosomes.bed is a BED file with the chromosome names belonging to each species.
4. Convert the yeast BAM alignments into genome coverage vectors. We used Python 2.7 HTSeq functions to write a custom script for the following functions. Iterate through the BAM file, and record the genome positions of read 5′ and 3′ ends, as well as each read length. Use read 5′ ends that span −12 of the start codon to −15 of the termination codon to tally the raw counts per gene. Only genes with ≥64 reads should be retained for further analysis. Convert read counts to reads per kilobase per million reads, excluding reads from the first eight codons of each gene.
5. Perform basic quality control of the yeast libraries. The vast majority of reads should map to CDS regions when binned by transcriptome feature (i.e., 5′ transcript leader, CDS, 3′ UTR). A histogram of the read length distribution should show an enrichment of ~28mers for *S. cerevisiae* (different read lengths predominate in other organisms). Severe elongation defects that generate an empty A site [27] may also show a second enrichment of ~22mers. A meta-read plot centered at the

annotated translation start site should show a peak of read 5′ ends at −12 nucleotides from the A of AUG, as previously described [28, 29]. Similarly, centering the meta-read plot at the termination codon should show a peak of read 5′ ends at −15 nucleotides from the first U. Three-nucleotide periodicity should also be evident from these same plots.

6. Determine counts per gene from the human spike-in sample. Run the htseq-count package with command line options `htseq-count human.bam annotations.gff3 --format=bam --stranded=no --type=CDS --idattr=gene_id --mode=union --samout=out.sam > counts.out`, where `human.bam` is the BAM file of human-only alignments from **step 3**, `annotations.gff3` is an annotations file of the human genome, and `out.sam > counts.out` generates a text file of the read counts mapping to each gene. This command will only track reads mapping to CDS regions.
7. **Steps 7** and **8** describe two different approaches to generate global scaling factors based on the representation of spike-in counts. Normalize the translation activity between libraries by dividing the total number of yeast CDS footprints (calculated in **step 4**) by the total number of human CDS footprints (generated in **step 6**). Qualitatively, samples with the highest yeast polysome to monosome ratio should also have the highest ratio of yeast to human CDS counts, because a higher fraction of the total yeast ribosomes were mRNA-engaged.
8. To more accurately compare the translation activity on each gene between libraries, a scaling factor can be derived by linear regression of human RPKM expression values between two libraries (Fig. 5a). The library with the highest human read coverage is fixed as the calibrator library against which all other libraries are rescaled. This analysis is more accurate when filtered for the top percentile of human gene expression. For example, in Wang et al. [25], only the top 5 percentile of human genes were included. Human RPKMs from the calibrator library are then plotted against human RPKMs from each library to be rescaled. The slope of the linear regression is a scaling factor for all yeast RPKM values from the rescaled library (Fig. 5b).
9. To determine statistically significant changes in relative gene expression, unscaled counts can be used with DESeq2 for ribosome profiling alone [30] or Xtail for coupled RNA-seq and ribosome profiling libraries [31]. To test for significant changes based on globally scaled counts, Reference [32] provides a statistical framework that explicitly encompasses absolute changes in gene-specific mRNA levels between conditions, which was validated on biological samples collected at different growth rates and libraries with artificially varied levels of spike-in standards.

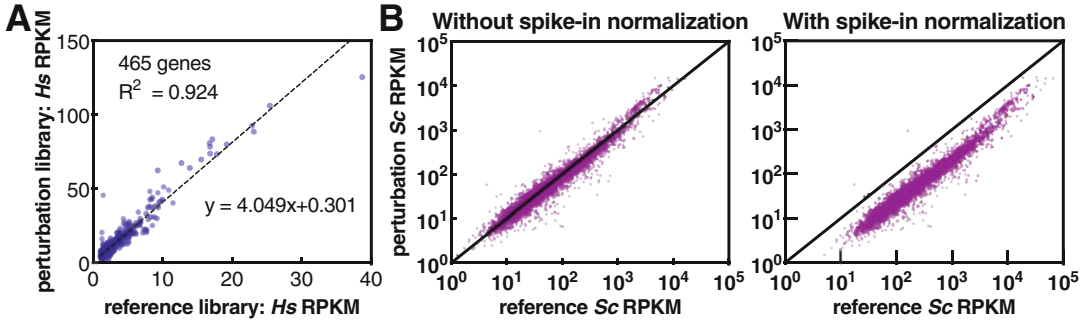

Fig. 5 (**a**) Linear regression of spike-in species RPKM values between two libraries. The slope is the scaling factor of overall translation activity between the libraries. Data are from [25]. (**b**) The scaling factor calculated in (**a**) was used to transform yeast RPKM values from the library on the x-axis in (**a**)

4 Notes

1. This protocol uses RNase I (Ambion catalog no. AM2294). The source of RNase I is critical for this step and has been optimized for this protocol. We found that RNase I enzymes from different sources may affect the quality and reproducibility of the method.
2. This product has been discontinued. An alternative for yeast rRNA depletion is the RiboMinus Transcriptome Isolation Kit (ThermoFisher K155003). A comprehensive evaluation of alternatives for human rRNA removal is provided in [33].
3. The lysate preparations described here omit in vivo cycloheximide incubation, which is essential within the primary sample to avoid artifactual distortions of ribosome position [17, 34]. In practice, the spike-in sample can be harvested with in vivo cycloheximide to improve yield and simplify downstream processing. However, the spike-in reads should subsequently be interpreted only on the level of counts per gene rather than for precise ribosome positioning.
4. To avoid stressing the cells, the entire harvest procedure until submersion in liquid nitrogen should take <1 min. The cells should not run completely dry on the Whatman filter. If desired, a small pellet of cells can be scraped from the filter into a separate 1.5 mL tube and frozen in liquid nitrogen for preparing matched RNA-seq libraries.
5. In our experience the contribution of rRNA to the overall A260 signal depends on cell state, particularly following extreme nutrient depletion that causes ribosome turnover. In such cases the rRNA/A260 can be variable between biological replicates. Therefore we have devised an accurate quantitation technique that depends on near-complete recovery of extracted

RNA. We have found that hot phenol extraction combined with gel phase lock tubes is more reproducible than column purification from commercial kits. If phase lock tubes are unavailable, a constant sub-volume of the aqueous layer (for example, 400 out of 500 μL) can be transferred at each extraction step.

6. We use the following standard precipitation protocol throughout: if the initial volume of aqueous solution is low, additional water or buffer can be added as an optional first step. Add 1/9th volume of sodium acetate pH 5.2 and 2.0 μL of Glycoblue. For gel elution in RNA or DNA buffer, the sodium acetate should be omitted. Mix by pipetting, and then add an equal volume of 100% isopropanol. Mix by inversion. Chill at −20 °C for at least 30 min. Spin at 21,000 × *g* for 30 min at 4 °C. Wash the pellet with 80% ethanol for footprint (15–60 mer) RNA and 70% ethanol for longer RNA or DNA. Spin at 21,000 × *g* for 10 min at 4 °C. Remove the supernatant completely and air-dry the pellet for 5 min before resuspending.
7. With our gel system and imager, a 1.5-fold dilution series from 5.0 to 0.5 μg total RNA gave a linear signal with $R^2 > 0.99$. The exact quantity of standards and samples loaded on the gel may need to be optimized.
8. The Synergel maintains crisp banding while clearly resolving tRNAs and all 4 rRNAs. Ethidium bromide must be mixed into the gel during casting. Post-run staining does not yield sufficient intensity to accurately quantify standards. Wide-toothed combs give more reliable quantification. Because ethidium bromide will migrate toward the anode, all samples should preferably be loaded on the same row of wells. If this is not possible, prepare a master batch of total RNA standards to include on each row of each gel.
9. The amount of RNase I is determined by overall A260 because variable amounts of tRNA in the lysates affect the amount of RNase needed to digest the ribosome-mRNA complexes. The total RNA concentrations determined in Subheading 3.3, **step 6** could also be used to normalize RNase I addition to each sample.
10. The yeast and human extracts are digested separately because they have different sensitivities to RNase I.
11. The 33% ethanol for the first column application retains large RNAs on the column, while small RNAs are eluted. The small RNAs are then captured on and eluted from a second column per standard methods. This step replaces the original Ingolia et al. [15] method of eluting small RNAs through a Microcon YM-100 microconcentrator, which has been discontinued.

12. Fifteen to 34 mer footprints have been shown to capture quality control intermediates and all intermediates of normal elongation [27, 35, 36]. Disomes can be size selected from 40–60 nucleotides to enrich for ribosome queuing events [35].
13. UV transillumination damages nucleic acids and compromises library construction.
14. The gel slice can alternatively be left intact, placed in a 1.5 mL tube with 0.4 mL elution buffer, and eluted overnight with rotation at 4 °C. For DNA extractions, elute overnight with rotation at room temperature.
15. The pre-adenylated adapter was synthesized by reacting adenosine 5′-phosphorimidazolide with a 5′-phosphorylated DNA oligo by the method of [37]. The Universal miRNA Cloning Linker (NEB S1315S) is a commercially available alternative, in which case the region of the RT primer that anneals to the adapter must be modified for this sequence.
16. The first three reaction components can be assembled and dispensed as a master mix. PEG-8000 and T4 RNA ligase I should be added individually in the listed order.
17. Depleting the unligated adapter is essential because libraries with inserts of ~15 nt are minimally resolved from no-insert libraries on the final PCR gels.
18. We have found these conditions, which reduce enzyme cost, to circularize cDNA as efficiently as the manufacturer's protocol requiring 1.0 μL CircLigase.
19. The libraries described here were sequenced on an Illumina HiSeq 2500 with 75 base pair single-end reads using High Output Mode V4. Ensure that all adapter sequences are compatible with the Illumina platforms and sequencing chemistries available to the end user. If not, it may be necessary to modify the reverse transcription and PCR primer sequences provided here.
20. These BAM files will not contain quality strings because `fastx_collapser` cannot assign a quality string to a collapsed sequence. To run `htseq-count` downstream, insert a placeholder quality string (with length equivalent to the read length) to each line of the BAM file.

Acknowledgments

We thank Drs. Rachel Green, Colin Wu, Boris Zinshteyn, and Mary Thompson for helpful discussions, Maria Rojas-Duran for assistance with library preparation, and Drs. Joshua Arribere and Boris Zinshteyn for critical reading of the manuscript. This work was funded by NIH R01GM094303 and R01GM132358 to WVG and an NSF Graduate Research Fellowship to YJW.

References

1. Arribere JA, Doudna JA, Gilbert WV (2011) Reconsidering movement of eukaryotic mRNAs between polysomes and P bodies. Mol Cell 44(5):745–758. https://doi.org/10.1016/j.molcel.2011.09.019
2. Vaidyanathan PP, Zinshteyn B, Thompson MK, Gilbert WV (2014) Protein kinase A regulates gene-specific translational adaptation in differentiating yeast. RNA 20(6):912–922. https://doi.org/10.1261/rna.044552.114
3. Zhang Y, Burkhardt DH, Rouskin S, Li GW, Weissman JS, Gross CA (2018) A stress response that monitors and regulates mRNA structure is central to cold shock adaptation. Mol Cell 70(2):274–286 e277. https://doi.org/10.1016/j.molcel.2018.02.035
4. Kim DH, Sarbassov DD, Ali SM, King JE, Latek RR, Erdjument-Bromage H, Tempst P, Sabatini DM (2002) mTOR interacts with raptor to form a nutrient-sensitive complex that signals to the cell growth machinery. Cell 110 (2):163–175. https://doi.org/10.1016/s0092-8674(02)00808-5
5. Thoreen CC, Chantranupong L, Keys HR, Wang T, Gray NS, Sabatini DM (2012) A unifying model for mTORC1-mediated regulation of mRNA translation. Nature 485 (7396):109–113. https://doi.org/10.1038/nature11083
6. Dever TE, Feng L, Wek RC, Cigan AM, Donahue TF, Hinnebusch AG (1992) Phosphorylation of initiation factor 2 alpha by protein kinase GCN2 mediates gene-specific translational control of GCN4 in yeast. Cell 68 (3):585–596. https://doi.org/10.1016/0092-8674(92)90193-g
7. Hinnebusch AG, Ivanov IP, Sonenberg N (2016) Translational control by 5-′-untranslated regions of eukaryotic mRNAs. Science 352(6292):1413–1416. https://doi.org/10.1126/science.aad9868
8. Bhat M, Robichaud N, Hulea L, Sonenberg N, Pelletier J, Topisirovic I (2015) Targeting the translation machinery in cancer. Nat Rev Drug Discov 14(4):261–278. https://doi.org/10.1038/nrd4505
9. Pospisek M, Valasek L (2013) Polysome profile analysis--yeast. Methods Enzymol 530:173–181. https://doi.org/10.1016/B978-0-12-420037-1.00009-9
10. Lodish HF (1971) Alpha and beta globin messenger ribonucleic acid. Different amounts and rates of initiation of translation. J Biol Chem 246(23):7131–7138
11. Khoutorsky A, Bonin RP, Sorge RE, Gkogkas CG, Pawlowski SA, Jafarnejad SM, Pitcher MH, Alain T, Perez-Sanchez J, Salter EW, Martin L, Ribeiro-da-Silva A, De Koninck Y, Cervero F, Mogil JS, Sonenberg N (2015) Translational control of nociception via 4E-binding protein 1. elife 4:e12002. https://doi.org/10.7554/eLife.12002
12. Iwasaki S, Ingolia NT (2017) The growing toolbox for protein synthesis studies. Trends Biochem Sci 42(8):612–624. https://doi.org/10.1016/j.tibs.2017.05.004
13. Park E-H, Zhang F, Warringer J, Sunnerhagen P, Hinnebusch AG (2011) Depletion of eIF4G from yeast cells narrows the range of translational efficiencies genome-wide. BMC Genomics 12(1):1–8. https://doi.org/10.1186/1471-2164-12-68
14. Rubio CA, Weisburd B, Holderfield M, Arias C, Fang E, DeRisi JL, Fanidi A (2014) Transcriptome-wide characterization of the eIF4A signature highlights plasticity in translation regulation. Genome Biol 15(10):476. https://doi.org/10.1186/s13059-014-0476-1
15. Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324 (5924):218–223. https://doi.org/10.1126/science.1168978
16. Li GW, Burkhardt D, Gross C, Weissman JS (2014) Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. Cell 157(3):624–635. https://doi.org/10.1016/j.cell.2014.02.033
17. Weinberg DE, Shah P, Eichhorn SW, Hussmann JA, Plotkin JB, Bartel DP (2016) Improved ribosome-footprint and mRNA measurements provide insights into dynamics and regulation of yeast translation. Cell Rep 14 (7):1787–1799. https://doi.org/10.1016/j.celrep.2016.01.043
18. Zinshteyn B, Rojas-Duran MF, Gilbert WV (2017) Translation initiation factor eIF4G1 preferentially binds yeast transcript leaders containing conserved oligo-uridine motifs. RNA 23(9):1365–1375. https://doi.org/10.1261/rna.062059.117
19. Sen ND, Zhou F, Ingolia NT, Hinnebusch AG (2015) Genome-wide analysis of translational efficiency reveals distinct but overlapping functions of yeast DEAD-box RNA helicases Ded1 and eIF4A. Genome Res 25(8):1196–1205. https://doi.org/10.1101/gr.191601.115
20. McGlincy NJ, Ingolia NT (2017) Transcriptome-wide measurement of translation by ribosome profiling. Methods 126:112–129. https://doi.org/10.1016/j.ymeth.2017.05.028

21. Han Y, David A, Liu B, Magadan JG, Bennink JR, Yewdell JW, Qian SB (2012) Monitoring cotranslational protein folding in mammalian cells at codon resolution. Proc Natl Acad Sci U S A 109(31):12467–12472. https://doi.org/10.1073/pnas.1208138109
22. Popa A, Lebrigand K, Barbry P, Waldmann R (2016) Pateamine A-sensitive ribosome profiling reveals the scope of translation in mouse embryonic stem cells. BMC Genomics 17(1):52. https://doi.org/10.1186/s12864-016-2384-0
23. Iwasaki S, Floor SN, Ingolia NT (2016) Rocaglates convert DEAD-box protein eIF4A into a sequence-selective translational repressor. Nature 534(7608):558–561. https://doi.org/10.1038/nature17978
24. Thompson MK, Rojas-Duran MF, Gangaramani P, Gilbert WV (2016) The ribosomal protein Asc1/RACK1 is required for efficient translation of short mRNAs. eLife 5:e11154. https://doi.org/10.7554/eLife.11154
25. Wang YJ, Vaidyanathan PP, Rojas-Duran MF, Udeshi ND, Bartoli KM, Carr SA, Gilbert WV (2018) Lso2 is a conserved ribosome-bound protein required for translational recovery in yeast. PLoS Biol 16(9):e2005903. https://doi.org/10.1371/journal.pbio.2005903
26. Illumina (2019) Illumina Adapter Sequences. https://support.illumina.com/downloads/illumina-adapter-sequences-document-1000000002694.html
27. Wu CC, Zinshteyn B, Wehner KA, Green R (2019) High-resolution ribosome profiling defines discrete ribosome elongation states and translational regulation during cellular stress. Mol Cell 73(5):959–970 e955. https://doi.org/10.1016/j.molcel.2018.12.009
28. Schuller AP, Wu CC-C, Dever TE, Buskirk AR, Green R (2017) eIF5A functions globally in translation elongation and termination. Mol Cell 66(2):194–205. https://doi.org/10.1016/j.molcel.2017.03.003
29. Zinshteyn B, Gilbert WV (2013) Loss of a conserved tRNA anticodon modification perturbs cellular signaling. PLoS Genet 9(8):e1003675. https://doi.org/10.1371/journal.pgen.1003675
30. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15(12):550. https://doi.org/10.1186/s13059-014-0550-8
31. Xiao Z, Zou Q, Liu Y, Yang X (2016) Genome-wide assessment of differential translations with ribosome profiling data. Nat Commun 7:11194. https://doi.org/10.1038/ncomms11194
32. Athanasiadou R, Neymotin B, Brandt N, Wang W, Christiaen L, Gresham D, Tranchina D (2019) A complete statistical model for calibration of RNA-seq counts using external spike-ins and maximum likelihood theory. PLoS Comput Biol 15(3):e1006794. https://doi.org/10.1371/journal.pcbi.1006794
33. Herbert ZT, Kershner JP, Butty VL, Thimmapuram J, Choudhari S, Alekseyev YO, Fan J, Podnar JW, Wilcox E, Gipson J, Gillaspy A, Jepsen K, BonDurant SS, Morris K, Berkeley M, LeClerc A, Simpson SD, Sommerville G, Grimmett L, Adams M, Levine SS (2018) Cross-site comparison of ribosomal depletion kits for Illumina RNAseq library construction. BMC Genomics 19(1):199. https://doi.org/10.1186/s12864-018-4585-1
34. Gerashchenko MV, Gladyshev VN (2014) Translation inhibitors cause abnormalities in ribosome profiling experiments. Nucleic Acids Res 42(17):e134–e134. https://doi.org/10.1093/nar/gku671
35. Guydosh NR, Green R (2014) Dom34 rescues ribosomes in 3′ untranslated regions. Cell 156(5):950–962. https://doi.org/10.1016/j.cell.2014.02.006
36. Guydosh NR, Green R (2017) Translation of poly(A) tails leads to precise mRNA cleavage. RNA 23(5):749–761. https://doi.org/10.1261/rna.060418.116
37. Lau NC, Lim LP, Weinstein EG, Bartel DP (2001) An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science 294(5543):858–862. https://doi.org/10.1126/science.1065062

Chapter 6

Genome-Wide Analysis of Translation in Replicatively Aged Yeast

Hanna Barlit, Manish K. Rai, Sara I. Shoushtari, Carine Beaupere, and Vyacheslav M. Labunskyy

Abstract

Protein synthesis is an essential process that affects major cellular functions including growth, energy production, cell signaling, and enzymatic reactions. However, how it is impacted by aging and how the translation of specific proteins is changed during the aging process remain understudied. Although yeast is a widely used model for studying eukaryotic aging, analysis of age-related translational changes using ribosome profiling in this organism has been challenging due to the need for isolating large quantities of old cells. Here, we provide a detailed protocol for genome-wide analysis of protein synthesis using ribosome profiling in replicatively aged yeast. By combining genetic enrichment of old cells with the biotin affinity purification step, this method allows large-scale isolation of aged cells sufficient for generating ribosome profiling libraries. We also describe a strategy for normalization of samples using a spike-in with worm lysates that permits quantitative comparison of absolute translation levels between young and old cells.

Key words Ribo-Seq, Mother enrichment program, Protein translation, Aging, *Saccharomyces cerevisiae*, Spike-in, Yeast

1 Introduction

Many age-related human diseases are associated with the loss of protein homeostasis including type 2 diabetes, cancer, neurodegeneration, and cardiovascular disease [1–3]. Despite the fundamental importance of protein translation in the aging process, relatively few studies have investigated translational changes in protein synthesis accompanying aging. Yeast *Saccharomyces cerevisiae* is one of the most established model systems to study eukaryotic aging given its short lifespan and amenability to genetic manipulations [4, 5]. However, quantitative analysis of protein synthesis in replicatively aged yeast has been difficult due to the challenges of purifying large quantities of old cells [6].

Hanna Barlit and Manish K. Rai contributed equally to this work.

Vyacheslav M. Labunskyy (ed.), *Ribosome Profiling: Methods and Protocols*, Methods in Molecular Biology, vol. 2252, https://doi.org/10.1007/978-1-0716-1150-0_6,

Ribosome profiling (Ribo-Seq) has recently emerged as a powerful tool to monitor protein translation at the genome-wide level [7]. This method is based on deep sequencing of ribosome protected ~28 nt mRNA fragments, which allows genome-wide quantification of translation at nucleotide resolution [8, 9]. In addition to global analysis of the actively translated regions of the transcriptome, combining ribosome profiling with total mRNA abundance measurements by RNA-Seq can also be used to estimate the translation efficiency of each transcript and contribution of both transcriptional and translational regulation to changes in gene expression between experimental conditions [10, 11]. Although Ribo-Seq has been used to study translational changes and mechanisms of translational regulation in different physiological states as well as in response to physiological stress conditions (including nutrient limitation, heat shock, oxidative stress, endoplasmic reticulum stress), the standardized protocol for analyzing translational changes with aging is currently lacking. Here we describe a detailed ribosome profiling protocol used in our lab for analyzing changes in protein translation in replicatively aged yeast cells (Fig. 1). To isolate large quantities of aged cells sufficient for generating ribosome profiling libraries, we utilize the mother enrichment program (MEP) [12]. This method permits efficient isolation of the old mother cells by using an estradiol-inducible system that leads to cell cycle arrest specifically in the daughter cells. By combining the MEP with affinity purification of biotin-labeled mother cells [13], we are able to routinely isolate more than 1×10^8 cells, which provides sufficient material to perform Ribo-Seq. Importantly, we compare changes in protein synthesis in replicatively aged cells with genetically identical young cells that were labeled with biotin and cultured under the same conditions, allowing direct comparison between young and old cells.

Previous studies have also shown that the bulk protein synthesis is significantly reduced during aging in a range of organisms and different cell types [14] as well as in replicatively aged yeast cells [15]. In conditions when the global translation is decreased, normalization of raw sequencing reads to account for differences in overall rates of protein synthesis is required for absolute quantification of translation changes between samples. To overcome this limitation, we use a spike-in with the lysate prepared from *Caenorhabditis elegans*. For this, the yeast cellular extracts used for mRNA and ribosome footprint isolation are spiked with equal amounts of the worm lysate proportional to the quantity of total RNA in each sample. Adding this important spike-in control from the evolutionarily distant species allows us to quantify differences in overall changes of protein translation in replicatively aged yeast cells. This strategy can be directly applied for normalization of ribosome profiling experiments in other eukaryotic species or different cellular states to enable accurate quantification of in vivo translation.

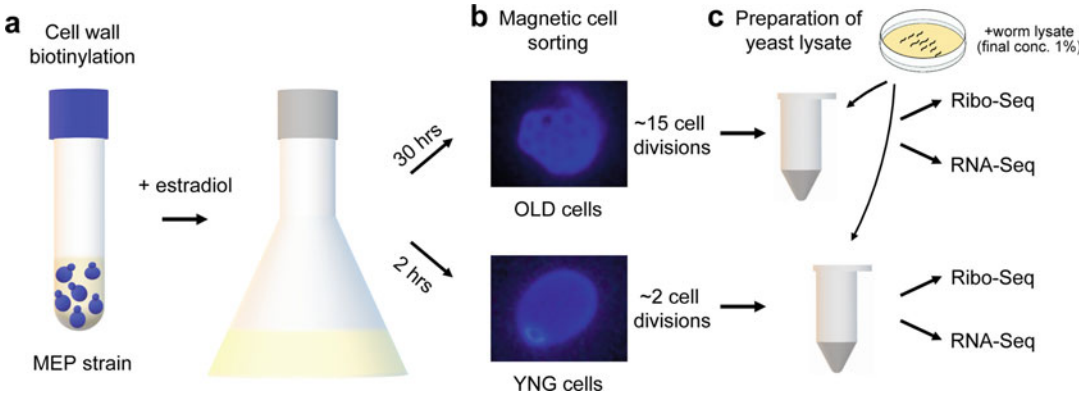

Fig. 1 Overview of the Ribo-Seq protocol for analysis of translation in replicatively aged yeast cells. (**a**) Yeast cells are biotinylated and are grown in the presence of estradiol for 2 h (YNG) or 30 h (OLD), which induces the mother enrichment program (MEP) preventing division of daughter cells. (**b**) Biotinylated mother cells are then separated using magnetic cell sorting enabling isolation of large quantities of old cells sufficient for generating ribosome profiling libraries. The age of yeast cells that have been isolated using the MEP and magnetic sorting is determined by counting the number of "bud scars" stained with calcofluor dye. The population of OLD cells obtained after sorting on average contains ~11–13 more bud scars per cell compared to YNG cells. (**c**) An equal amount of the worm lysate spike-in control (1%) is added into each sample proportional to the number of A_{260} units in yeast lysates allowing normalization of translation changes in aged and young cells

2 Materials

2.1 Preparation of Worm Lysate

1. *C. elegans* N2 strain.
2. *E. coli* OP50 strain.
3. 60 mm Petri plates.
4. Nematode Growth Medium (NGM): 1.7% (w/v) agar, 50 mM NaCl, 0.25% (w/v) peptone, 1 mM $CaCl_2$, 5 μg/mL cholesterol, 25 mM KPO_4, 1 mM $MgSO_4$.
5. M9 buffer: 22 mM KH_2PO_4, 42 mM Na_2HPO_4, 86 mM NaCl, 1 mM $MgSO_4$.
6. Lysis buffer: 20 mM Tris–HCl pH 8.0, 140 mM KCl, 5 mM $MgCl_2$, 100 μg/mL cycloheximide, 0.5 mM DTT, 1% Triton X-100.
7. Liquid nitrogen.
8. 50 mL conical tubes.
9. Screw-cap 2 mL tubes.
10. Nonstick 1.5 mL RNase-free tubes.
11. BioSpec cryomill.
12. Chrome-steel beads, 3.2 mm.
13. Stainless steel microvials, 1.8 mL.
14. Spectrophotometer.

2.2 Isolation of Replicatively Aged Yeast Cells

1. Yeast MEP strain (UCC8773, MATa his3Δ1 leu2Δ0 ura3Δ0 lys2Δ0 hoΔ::Pscw11-cre-EDB78-NatMX loxP-CDC20-intron-loxP-HphMX loxP-UBC9-loxP-LEU2) [16].
2. YPD medium: 1% yeast extract, 2% peptone, 2% glucose.
3. 100 mg/mL nourseothricin (Nat): prepare 100 mg/mL stock in sterile water, and store at −20 °C.
4. 50 mg/mL hygromycin B.
5. Refrigerated centrifuge (with a rotor for 50 mL conical tubes).
6. Refrigerated microcentrifuge.
7. Phosphate-buffered saline (PBS).
8. EZ-Link Sulfo-NHS-LC-Biotin.
9. 0.1 M glycine-PBS: prepare 0.1 M glycine solution in PBS from 2.5 M glycine stock just before use.
10. PBS + BE: 1 mg/mL BSA, 2 mM EDTA in PBS.
11. 1 mM 17β-estradiol: prepare a 1 mM stock in ethanol, and store at −20 °C.
12. Dynabeads Biotin Binder.
13. DynaMag magnet.
14. Resuspension buffer: 20 mM Tris–HCl pH 8.0, 140 mM KCl, 5 mM $MgCl_2$.
15. Liquid nitrogen.
16. 50 mL conical tubes.
17. 15 mL conical tubes.
18. Screw-cap 2 mL tubes.
19. Nonstick 1.5 mL RNase-free tubes.
20. Automated cell counter.

2.3 Counting Bud Scars

1. 4% formaldehyde in PBS.
2. Calcofluor (fluorescent brightener 28).
3. Fluorescent mounting medium.
4. Glass slides with coverslips.
5. Fluorescence microscope with DAPI emission filter.

2.4 Preparation of Yeast Lysate

1. Lysis buffer: 20 mM Tris–HCl pH 8.0, 140 mM KCl, 5 mM $MgCl_2$, 100 μg/mL cycloheximide, 0.5 mM DTT, 1% Triton X-100.
2. Liquid nitrogen.
3. 50 mL conical tubes.
4. Screw-cap 2 mL tubes.
5. Nonstick 1.5 mL RNase-free tubes.

6. BioSpec cryomill.
7. Chrome-steel beads, 3.2 mm.
8. Stainless steel microvials, 1.8 mL.
9. Spectrophotometer.

2.5 Footprint Extraction

1. Ultracentrifuge with SW-41 Ti rotor.
2. Thin-wall polyallomer tubes, 13.2 mL.
3. Gradient buffer: 20 mM Tris–HCl pH 8.0, 140 mM KCl, 5 mM $MgCl_2$, 100 μg/mL cycloheximide, 0.5 mM DTT.
4. 10% sucrose: prepare 10% sucrose solution in gradient buffer just before use.
5. 50% sucrose: prepare 50% sucrose solution in gradient buffer just before use.
6. Chase solution: 20 mM Tris–HCl pH 8.0, 140 mM KCl, 5 mM $MgCl_2$, 60% sucrose.
7. BioComp Gradient Master.
8. RNase I, 100 U/μL.
9. Head-over-heels rotator.
10. Gradient fractionation system including tube piercer stand.
11. UV monitor.
12. Syringe pump.
13. 0.5 mL centrifugal filters (100 kDa MWCO).
14. Release buffer: 20 mM Tris–HCl pH 7.0, 2 mM EDTA, 40 U/mL Superase-In.

2.6 Footprint Fragment Purification

1. 20% SDS.
2. Acid-phenol/chloroform, pH 4.5 (with IAA, 125:24:1).
3. Nonstick 1.5 mL RNase-free tubes.
4. 3 M NaOAc, pH 5.5.
5. 10 mg/mL glycogen.
6. Ethanol.

2.7 Poly(A) mRNA Extraction

1. 20% SDS.
2. Acid-phenol/chloroform, pH 4.5 (with IAA, 125:24:1).
3. Nonstick 1.5 mL RNase-free tubes.
4. 3 M NaOAc, pH 5.5.
5. 10 mg/mL glycogen.
6. Ethanol.
7. Dynabeads Oligo $(dT)_{25}$.
8. DynaMag magnet.

9. Binding buffer: 100 mM Tris–HCl pH 7.5, 500 mM LiCl, 10 mM EDTA, 1% LiDS, 5 mM DTT.
10. Washing buffer A: 10 mM Tris–HCl pH 7.5, 150 mM LiCl, 1 mM EDTA, 0.1% LiDS.
11. Washing buffer B: 10 mM Tris–HCl pH 7.5, 150 mM LiCl, 1 mM EDTA.
12. 10 mM Tris–HCl pH 8.0.
13. 10× RNA fragmentation buffer: 100 mM $ZnCl_2$, 100 mM Tris–HCl pH 7.0.
14. RNase-free water.

2.8 Dephosphorylation

1. T4 polynucleotide kinase (PNK), 10,000 U/mL.
2. Superase-In RNase inhibitor, 20 U/μL.
3. 10× TBE buffer: 1 M Tris–HCl pH 8.3, 0.9 M boric acid, and 0.01 M EDTA.
4. 2× TBE-urea sample buffer.
5. 15% polyacrylamide TBE-urea gels.
6. SYBR Gold nucleic acid gel stain.
7. Blue light transilluminator.
8. RNase-free disposable pellet pestles.
9. Nonstick 1.5 mL RNase-free tubes.
10. RNA elution buffer: 20 mM Tris–HCl, pH 7.0, 2.0 mM EDTA.
11. 3 M NaOAc, pH 5.5.
12. 10 mg/mL glycogen.
13. Ethanol.

2.9 3′-Adapter Ligation

1. T4 RNA ligase 2 truncated KQ, 200,000 U/mL.
2. 100 ng/μL preadenylated 3′-adapter /5rApp/AGATCGGAAGAGCACACGTCT/3ddC/.
3. 5′ deadenylase, 10 U/μL.
4. Rec J exonuclease, 10 U/μL.

2.10 Reverse Transcription

1. Superscript III reverse transcriptase, 200 U/μL.
2. 10 mM deoxynucleotide solution mix (dNTPs).
3. 8 μM reverse transcription (RT) primer
   5'-pGATCGTCGGACTGTAGAACTCTGAACGTGTAGATCTCGGTGGTCGCCGTATCATT/iSP18/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3′.
4. 2 M NaOH.
5. 2 M HCl.

6. RNase-free water.
7. 3 M NaOAc, pH 5.5.
8. 10 mg/mL glycogen.
9. Ethanol.
10. 10× TBE buffer: 1 M Tris–HCl pH 8.3, 0.9 M boric acid, and 0.01 M EDTA.
11. 2× TBE-urea sample buffer.
12. 10% polyacrylamide TBE-urea gels.
13. SYBR Gold nucleic acid gel stain.
14. Blue light transilluminator.
15. RNase-free disposable pellet pestles.
16. Nonstick 1.5 mL RNase-free tubes.
17. 20 mM Tris–HCl pH 7.0.

2.11 Circularization and PCR Library Amplification

1. RNase-free water.
2. CircLigase II single-stranded DNA (ssDNA) ligase, 100 U/μL.
3. 20 μM forward PCR primer
   5'-AATGATACGGCGACCACCGAGATCTACACGTTCAG
   AGTTCTACAGTCCGACG-3′.
4. 20 μM indexed reverse PCR primers (the index specific for each primer is underlined)
   Index primer 1
   CAAGCAGAAGACGGCATACGAGAT<u>CGTGAT</u>GTGACTG
   GAGTTCAGACGTGTGCTCTTCCGATCT.
   Index primer 2
   CAAGCAGAAGACGGCATACGAGAT<u>ACATCG</u>GTGACTG
   GAGTTCAGACGTGTGCTCTTCCGATCT.
   Index primer 3
   CAAGCAGAAGACGGCATACGAGAT<u>GCCTAA</u>GTGACTG
   GAGTTCAGACGTGTGCTCTTCCGATCT.
   Index primer 4
   CAAGCAGAAGACGGCATACGAGAT<u>TGGTCA</u>GTGACTG
   GAGTTCAGACGTGTGCTCTTCCGATCT.
   Index primer 5
   CAAGCAGAAGACGGCATACGAGAT<u>CACTGT</u>GTGACTG
   GAGTTCAGACGTGTGCTCTTCCGATCT.
   Index primer 6
   CAAGCAGAAGACGGCATACGAGAT<u>ATTGGC</u>GTGACTG
   GAGTTCAGACGTGTGCTCTTCCGATCT.
   Index primer 7

CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTG
GAGTTCAGACGTGTGCTCTTCCGATCT.
Index primer 8
CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTG
GAGTTCAGACGTGTGCTCTTCCGATCT.
Index primer 9
CAAGCAGAAGACGGCATACGAGATCTGATCGTGACTG
GAGTTCAGACGTGTGCTCTTCCGATCT.
Index primer 10
CAAGCAGAAGACGGCATACGAGATAAGCTAGTGACTG
GAGTTCAGACGTGTGCTCTTCCGATCT.
Index primer 11
CAAGCAGAAGACGGCATACGAGATGTAGCCGTGACTG
GAGTTCAGACGTGTGCTCTTCCGATCT.
Index primer 12
CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTG
GAGTTCAGACGTGTGCTCTTCCGATCT.

5. Phusion high-fidelity DNA polymerase, 2,000 units/mL.
6. 10 mM dNTPs.
7. 5× DNA loading dye.
8. 10× TBE buffer: 1 M Tris–HCl pH 8.3, 0.9 M boric acid, and 0.01 M EDTA.
9. Non-denaturing 8% polyacrylamide TBE gels.
10. SYBR Gold nucleic acid gel stain.
11. Blue light transilluminator.
12. RNase-free disposable pellet pestles.
13. Nonstick 1.5 mL RNase-free tubes.
14. 10 bp DNA ladder, 1 μg/μL.

3 Methods

3.1 Preparation of Worm Lysate

1. Culture 2,000–4,000 synchronized worms on lawns of *E. coli* OP50 on each NGM agar plate (3–5 plates) at 20 °C to young adulthood (24 h after L4).
2. Harvest worms by washing the plates with 2 mL M9 buffer.
3. Transfer the mixture into 1.5 mL tubes. Wait for 10 min to allow young adult worms to settle down at the bottom of the tube. Remove the supernatant containing eggs and bacteria.
4. Resuspend the worms by adding 1 mL M9 buffer to each tube. Wait for 10 min to allow young adult worms to settle down at

the bottom of the tube, and remove the supernatant. Repeat the wash two more times.

5. Transfer washed worms into new 1.5 mL tubes (~500 μL worm pellet/tube). Remove the supernatant and resuspend in 1 mL lysis buffer. Slowly dispense the mixture into the 50 mL tube with liquid nitrogen. Transfer the flash-frozen droplets into 2 mL screw-cap tubes and store them at −80 °C.
6. Pre-chill a 1.8 mL stainless steel tube with three chrome-steel beads in liquid nitrogen. Add 0.35–0.4 g frozen worm pellets, and cover with a silicone rubber cap.
7. Homogenize frozen worm pellets by cryogrinding for 10 s at 4,200 rpm, and immediately chill the tube in liquid nitrogen. Repeat ten times.
8. Add 1 mL of lysis buffer, and mix well by pipetting. Transfer into a new 1.5 mL tube.
9. Remove debris by centrifugation at 20,000 × *g* for 5 min at 4 °C, and transfer the supernatant into a new 1.5 mL tube.
10. Determine OD_{260} and dilute the worm lysate to 50 OD_{260} units/mL with lysis buffer. Transfer 50 μL aliquots of the worm lysate into new tubes (2.5 OD_{260} units/tube). Flash freeze in liquid nitrogen, and store at −80 °C. Prepare a large batch that can be used for multiple experiments.

3.2 Isolation of Replicatively Aged Yeast Cells

1. Starting from a single colony, grow the MEP strain in 5 mL of YPD supplemented with 100 μg/mL Nat and 300 μg/mL hygromycin B (to maintain selection) at 30 °C overnight.
2. Measure the OD_{600} and dilute the overnight culture to OD_{600} ~ 0.2 in 40 mL YPD supplemented with 100 μg/mL Nat and 300 μg/mL hygromycin B. Divide the diluted culture equally into two 50 mL conical tubes.
3. Culture cells at 30 °C with shaking until OD_{600} reaches 0.6 (log phase).
4. Harvest cells in 50 mL tubes in a refrigerated centrifuge at 3,000 × *g*, 4 °C for 3 min.
5. Wash cells two times in 10 mL of sterile PBS.
6. Resuspend cells in 1 mL PBS.
7. Dilute 1 μL of the cell suspension to 50 μL by adding 49 μL of H_2O (50-fold dilution). Determine the cell density using an automated cell counter or a hemocytometer.
8. Transfer 3×10^8 cells to a new 1.5 mL tube for labeling with biotin, centrifuge at 2,000 × *g*, 4 °C for 3 min, and remove the supernatant.
9. Warm-up EZ-Link Biotin to room temperature before opening. Prepare 10 mg/mL EZ-Link Biotin stock solution in PBS

and add 300 μL to cells from the previous step (100 μL per 1×10^8 cells).

10. Rotate the tube for 30 min at room temperature.
11. Wash biotin-labeled cells two times in 1 mL 0.1 M glycine-PBS to quench and remove free biotin.
12. Centrifuge the sample at 2,000 rpm ($400 \times g$) at 4 °C for 2 min (*see* **Note 1**).
13. Resuspend the cells in 1 mL PBS and divide them into two 1.5 mL tubes (1.5×10^8 cells/tube). Keep on ice until inoculation.
14. Inoculate 1.5×10^8 labeled cells into 20 mL of cold YPD media containing 1 μM 17β-estradiol and grow at 30 °C for 2 h. These cells will be used to isolate the young cells' sample (YNG) (Fig. 1; *see* **Note 2**).
15. For isolation of old cells, inoculate 1.5×10^8 labeled cells from **step 13** to 700 mL of cold YPD media and add 17β-estradiol to 1 μM final concentration to induce the MEP. Grow cells at 30 °C for 20 h, then add 300 mL of fresh YPD media containing 1 μM 17β-estradiol, and continue growing at 30 °C for an additional 10 h (30-h total incubation time). These cells will be used to isolate the old cells' sample (OLD) (Fig. 1; *see* **Note 3**).
16. After the incubation is completed, harvest cells by centrifugation in 50 mL tubes in a refrigerated centrifuge at $3{,}000 \times g$, 4 °C for 3 min. Remove the supernatant.
17. Wash cells twice in 30 mL of cold PBS. Resuspend washed cells in 1 mL (YNG) or 10 mL (OLD) of PBS + BE, and transfer to a 1.5 mL tube (YNG) or a 15 mL conical tube (OLD), respectively (*see* **Note 4**).
18. Resuspend Dynabeads Biotin Binder before use. Transfer 300 μL of Dynabeads Biotin Binder (1.2×10^8 beads) into a 1.5 mL tube for YNG cells and 1 mL (4×10^8 beads) for OLD cells. Remove the buffer by placing on the magnet for 30 s and discard the supernatant.
19. Wash Dynabeads Biotin Binder in each tube two times in 1 mL PBS + BE buffer.
20. Add an aliquot of Dynabeads Biotin Binder to YNG and OLD cells from **step 16** and incubate at 4 °C for 1 h with rotation.
21. Place the tube with YNG cells on the magnet for 30 s, and carefully remove the supernatant without disturbing the beads/labeled cells attached to the magnet. Proceed to washing the beads (**step 22**). For OLD cells, transfer a 1.5 mL aliquot of cell suspension into a 1.5 mL tube. Place the tube on the magnet for 2 min and slowly remove unlabeled cells without disturbing the beads/old cells attached to the magnet.

Add another aliquot of cells to the tube. Repeat this process several times until all OLD cell suspension is processed.

22. Wash the beads in each tube three times with 1 mL of cold PBS. Transfer the beads/labeled cells into an RNase-free 1.5 mL tube.
23. Repeat **step 22** two more times.
24. Resuspend the cells in 1 mL YPD. Incubate the cells with shacking at 30 °C for 30 min to recover.
25. Dilute 1 μL of cells to 10 μL by adding 9 μL of H_2O (tenfold dilution). Determine the cell density using an automated cell counter or a hemocytometer. Calculate the yield obtained after cell sorting. For both YNG and OLD samples, save 10 μL of cells for counting bud scars (*see* **Note 5**).
26. Centrifuge cells 3,000 rpm (800 × *g*) for 3 min. Remove the supernatant. Resuspend cells in 150 μL of resuspension buffer, and slowly dispense the mixture into the 50 mL tube with liquid nitrogen. Transfer the flash-frozen droplets into 2 mL screw-cap tubes and store at −80 °C.

3.3 Counting Bud Scars

1. For counting bud scars, mix 10 μL of cells with 1 mL of 4% formaldehyde (in PBS), and rotate for 10 min at room temperature.
2. Centrifuge at 3,000 rpm (800 × *g*) for 3 min and remove the supernatant.
3. Wash cells twice with 1 mL PBS. Resuspend cells in 400 μL PBS containing 1 mg/mL calcofluor. Prepare 1 mg/mL calcofluor solution in PBS from 10 mg/mL stock just prior to use. Incubate cells for 15 min in the dark at room temperature.
4. Wash cells once with 1 mL PBS. Centrifuge at 3,000 rpm (800 × *g*) for 3 min and resuspend in 20 μL H_2O.
5. Mix 5 μL of cells with 5 μL of fluorescent mounting media, place on a glass slide, and cover with a coverslip. Image the stained bud scars on a fluorescence microscope using a DAPI emission filter. If cell sorting is successful, OLD cells should contain on average ~ 11–13 more bud scars per cell compared to YNG cells (Fig. 1).

3.4 Preparation of Yeast Lysate

1. Prepare several tubes containing flash-frozen droplets of lysis buffer (250 μL per tube) by slowly dispensing the buffer into the 50 mL tube with liquid nitrogen. Transfer the flash-frozen droplets into 2 mL screw-cap tubes and store them at −80 °C until cryogrinding.
2. Pre-chill a 1.8 mL stainless steel tube with three chrome-steel beads in liquid nitrogen. Add frozen yeast pellets (*see* Subheading 3.2, **step 26**) (*see* **Note 6**).

3. To the same tube, add 250 μL of flash-frozen lysis buffer droplets, and cover with a silicone rubber cap.
4. Homogenize yeast cells by cryogrinding for 10 s at 4,200 rpm, and immediately chill the tube in liquid nitrogen. Repeat ten times.
5. Add 500 μL of lysis buffer, and mix well by pipetting. Transfer into an RNase-free 1.5 mL tube.
6. Remove debris by centrifugation at 20,000 × *g* for 5 min at 4 °C, and transfer the supernatant into a new 1.5 mL tube.
7. Determine OD_{260} of the yeast lysate. Calculate the amount of the worm lysate spike-in control (*see* Subheading 3.1, **step 10**) required for each sample. Add 1 A_{260} unit of the worm lysate per 100 A_{260} units of each yeast lysate (to a final concentration 1%) for normalization (*see* **Note 7**).
8. Divide the lysate equally into two aliquots that will be used for the preparation of "footprint" and "mRNA" libraries. Flash freeze in liquid nitrogen, and store at −80 °C. The lysates can be stored at −80 °C indefinitely.

3.5 Footprint Extraction

1. Pre-chill the SW-41 Ti rotor with buckets to 4 °C.
2. Prepare linear sucrose gradients (10–50% sucrose) in gradient buffer in 13.2 mL thin-wall polyallomer tubes using BioComp Gradient Master following the manufacturer's instructions (*see* **Note 8**). Keep the gradients at 4 °C until use.
3. Thaw an aliquot of yeast lysate ("footprint" sample) on ice. Calculate the amount of RNase I (100 U/μL) needed for digestion for each sample. Add 20 U of RNase I per each A_{260} unit of yeast lysate and incubate for 1 h at room temperature with gentle rotation on a head-over-heels rotator.
4. Centrifuge lysate at 20,000 × *g* for 5 min at 4 °C to remove debris. Transfer the supernatant into an RNase-free 1.5 mL tube and put on ice.
5. Load lysate on top of a 10–50% sucrose gradient.
6. Centrifuge the gradients at 35,000 rpm (151,000 × *g*) at 4 °C for 3 h using the ultracentrifuge with the SW-41 Ti rotor (*see* **Note 9**).
7. Turn on the gradient fractionation system and UV monitor. Set the range on UV monitor to 2.0 for maximum detection limit.
8. Carefully transfer the tube with the sucrose gradient into the piercer stand. Fill the syringe with Chase Solution and connect it to the piercer stand tubing.
9. Pierce the tube and start pumping the Chase Solution at 1 mL/min using the syringe pump. By monitoring UV absorbance at

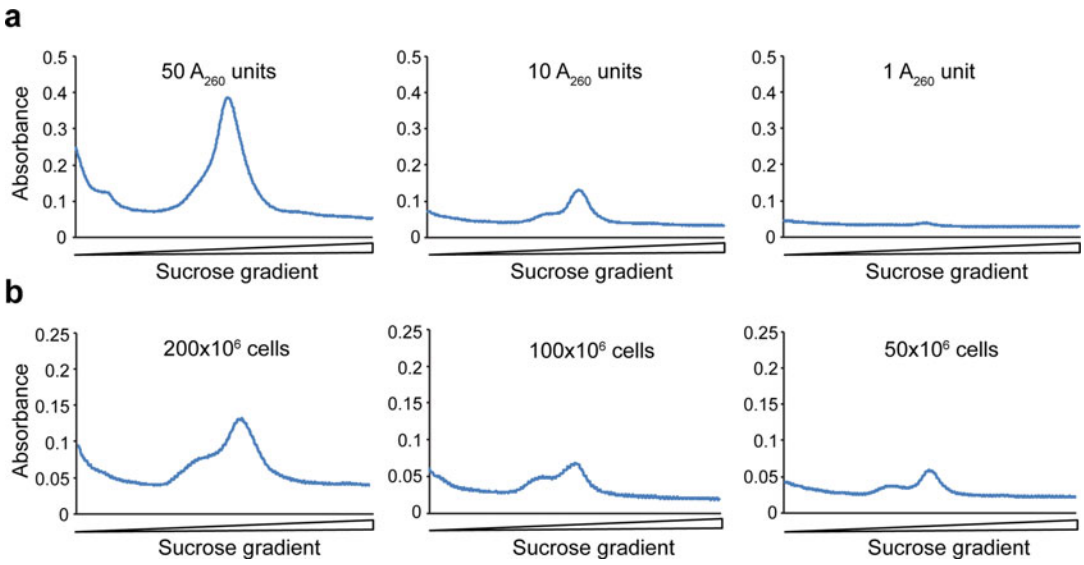

Fig. 2 Sucrose gradient fractionation of the yeast lysates prepared from varying quantities of cells. Cell lysates containing the indicated number of A_{260} units (**a**) or varying numbers of cells (**b**) were digested with RNase I and fractionated using 10–50% sucrose gradients. In our experience, we were able to collect the monosome peak and successfully prepare Ribo-Seq libraries from as low as 50×10^6 cells

254 nm, collect fractions corresponding to the monosome peak (Fig. 2), and put on ice.

10. Repeat fractionation for the rest of the samples (*see* **Note 10**).
11. Concentrate the monosome fractions using 0.5 mL centrifugal filters (100 kDa MWCO). For this, load the monosome fractions into the centrifugal filters, and centrifuge at 12,000 × *g* for 10 min at 4 °C. Discard the flow-through, and repeat until the volume reaches 100 μL.
12. Add 400 μL of release buffer. Pipette up and down to mix and incubate 10 min on ice. Transfer the filter unit into an RNase-free 1.5 mL collection tube.
13. Centrifuge at 12,000 × *g* for 10 min at 4 °C. Do not discard the flow-through.

3.6 Footprint Fragment Purification

1. Collect the flow-through (400 μL) containing footprint RNA fragments and transfer to an RNase-free 1.5 mL tube. Add 20 μL of 20% SDS (1% final concentration), and pipette up and down to mix.
2. Add 400 μL of acid-phenol/chloroform and vortex for 10 s (*see* **Note 11**).
3. Heat tubes at 65 °C for 5 min, and put on ice for 1 min. Centrifuge at 12,000 × *g* for 5 min at 4 °C, and transfer the aqueous layer (~350 μL) to an RNase-free 1.5 mL tube.

4. Add 35 μL of 3 M NaOAc, 3.5 μL of glycogen, and 2.5 volumes of 100% ethanol. Incubate at −20 °C for at least 1 h to precipitate RNA.

3.7 Poly(A) mRNA Extraction and mRNA Fragmentation

1. Thaw an aliquot of cell lysate ("mRNA" sample) on ice, and add 20 mM Tris–HCl pH 7.0 to bring the volume to 500 μL. Add 25 μL of 20% SDS (1% final concentration), and pipette up and down to mix.
2. Heat tubes at 65 °C briefly to dissolve SDS. Add 1 volume of acid-phenol/chloroform, and vortex for 10 s (*see* **Note 11**).
3. Heat tubes at 65 °C for 5 min, and put on ice for 1 min. Centrifuge at 12,000 × *g* for 5 min at 4 °C, and transfer the aqueous layer (~400 μL) to an RNase-free 1.5 mL tube.
4. Repeat phenol extraction. Add 1 volume of acid-phenol/chloroform, and vortex for 10 s. Heat tubes at 65 °C for 5 min, and put on ice for 1 min. Centrifuge at 12,000 × *g* for 5 min at 4 °C, and transfer the aqueous layer (~350 μL) to an RNase-free 1.5 mL tube.
5. Add 35 μL of 3 M NaOAc, 3.5 μL of glycogen, and 2.5 volumes of 100% ethanol. Incubate at −20 °C for at least 1 h to precipitate RNA.
6. Centrifuge the samples at 20,000 × *g* for 30 min at 4 °C to pellet the RNA, and remove the supernatant. Centrifuge at 20,000 × *g* for 30 s, remove the rest of the supernatant with a gel-loading tip, and air-dry the pellet for 5 min.
7. Dissolve the pellet in 300 μL of binding buffer, and pipette up and down to mix.
8. Resuspend Dynabeads Oligo$(dT)_{25}$ before use. Transfer 250 μL of beads to an RNase-free 1.5 mL tube and place the tube on a magnet. Wait for 30 s and remove the supernatant.
9. Take the tube out of the magnet and wash beads with 250 μL of fresh binding buffer. Remove the binding buffer by placing the tube on the magnet, wait for 30 s, and remove the supernatant.
10. Add 300 μL of the RNA sample (**step 7**), and mix by pipetting. Incubate for 5 min at room temperature with continuous mixing on a head-over-heels rotator.
11. Place the tube on the magnet, wait for 2 min, and remove the supernatant.
12. Wash beads twice with 600 μL of washing buffer A at room temperature. Place the tube on the magnet to separate beads from the supernatant between each wash.
13. Wash beads twice with 300 μL of washing buffer B at room temperature. Place the tube on the magnet to separate beads from the supernatant between each wash.

14. Remove the washing buffer. Add 20 μL of 10 mM Tris–HCl pH 8.0 and incubate at 65 °C for 2 min. Place the tube on the magnet, transfer the supernatant containing mRNA to a new 1.5 mL tube, and put on ice. Do not discard the beads.
15. Wash beads twice with 300 μL of washing buffer B.
16. Dilute the mRNA with four volumes of the binding buffer (e.g., if mRNA is eluted in 20 μL, add 80 μL of the binding buffer).
17. Place the tube on the magnet, wait for 2 min, and remove the supernatant.
18. Add the diluted mRNA and incubate with continuous mixing on a head-over-heels rotator for 5 min.
19. Repeat **steps 11–13**.
20. Remove the washing buffer. Add 20 μL of 10 mM Tris–HCl pH 8.0 and incubate at 65 °C for 2 minutes. Immediately place the tube on the magnet, and transfer the supernatant containing the mRNA to a new RNase-free tube.
21. Add 20 μL of RNase-free water, 4 μL of 3 M NaOAc, 1 μL of glycogen, and 2.5 volumes of 100% ethanol. Incubate at −20 °C for at least 1 h to precipitate RNA.
22. Centrifuge the samples at 20,000 × *g* for 30 min at 4 °C to pellet the RNA, and remove the supernatant. Centrifuge at 20,000 × *g* for 30 s, remove the rest of the supernatant with a gel-loading tip, and air-dry the pellet for 5 min.
23. Resuspend mRNA in 18 μL of RNase-free water. Add 2 μL of 10× RNA fragmentation buffer, incubate at 94 °C for 5 min, and put on ice.
24. Add 20 μL of RNase-free water, 4 μL of 3 M NaOAc, 1 μL of glycogen, and 2.5 volumes of 100% ethanol. Incubate at −20 °C for at least 1 h to precipitate RNA.

3.8 Dephosphorylation

1. Centrifuge "footprint" and fragmented "mRNA" samples at 20,000 × *g* for 30 min at 4 °C, and remove the supernatant. Centrifuge at 20,000 × *g* for 30 s, remove the rest of the supernatant with a gel-loading tip, and air-dry the pellet for 5 min.
2. Resuspend the pellet in 7.75 μL of RNase-free water. Add 1 μL of 10× T4 PNK buffer, 1 μL of T4 PNK, and 0.25 μL of Superase-In. Pipette up and down to mix and incubate at 37 °C for 1 h.
3. Rinse wells of the 15% TBE-urea gel from urea and pre-run the gel at 180 V for 15 min in 1× TBE buffer.
4. Add 10 μL of 2× TBE-urea sample buffer to 10 μL of each "footprint" and "mRNA" sample.

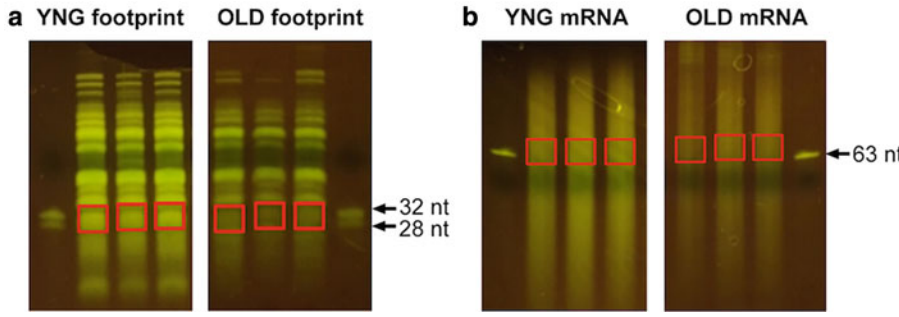

Fig. 3 Representative footprint and fragmented mRNA size selection gels after T4 polynucleotide kinase treatment. (**a**) Footprint fragments prepared from YNG and OLD yeast samples according to our protocol were separated on 15% polyacrylamide TBU-urea gel. 28 nt and 32 nt RNA oligonucleotides are used to guide the size of the gel slice that should be excised. (**b**) For fragmented mRNA samples, cut the gel slice around 50–70 nt. 63-mer RNA oligo is used as a control

5. Prepare an oligo sizing control containing 1 μL of 10 μM 32 nt RNA oligonucleotide, 1 μL of 10 μM 28 nt RNA oligonucleotide, 8 μL water, and 10 μL of 2× TBE-urea sample buffer for "footprint" samples. Prepare an oligo sizing control containing 1 μL of 10 μM 63 nt RNA oligonucleotide, 9 μL water, and 10 μL of 2× TBE-urea sample buffer for "mRNA" samples.
6. Heat the samples at 75 °C for 3 min, spin down at max speed for 10 s, and put on ice for 1 min. Load each sample into 2 wells of the 15% TBE-urea gel, and run the gel at 180 V for 1 h.
7. Dissolve 5 μL of SYBR Gold in 50 mL of RNase-free water. Stain the gel with SYBR Gold for 5 min, and protect from light.
8. Using a blue light transilluminator, cut the gel slices between 28 and 32 nt markers for "footprint" samples and ~50–70 nt for "mRNA" samples with a razor blade (Fig. 3). Freeze the polyacrylamide gel slices at −80 °C for 10 min.
9. Heat the gel slices at 70 °C for 2 min, and grind the gel using RNase-free disposable pellet pestles. Add 300 μL of RNA elution buffer, 1 μL Superase-In, and incubate at 37 °C for 3 h to extract RNA.
10. Centrifuge the samples at 12,000 × *g* for 5 min at 4 °C and transfer the supernatant to a new 1.5 mL tube. Add 30 μL of 3 M NaOAc, 3 μL of glycogen, and 2.5 volumes of 100% ethanol. Incubate at −20 °C for at least 1 h to precipitate RNA.

3.9 3′-Adapter Ligation

1. Centrifuge "footprint" and "mRNA" samples at 20,000 × *g* for 30 min at 4 °C, and remove the supernatant. Centrifuge at 20,000 × *g* for 30 s, remove the rest of the supernatant with a gel-loading tip, and air-dry the pellet for 10 min.
2. Resuspend the pellet in 4.75 μL of RNase-free water. Add 2 μL of 50% PEG-8000, 1 μL of 10× T4 RNA ligase buffer, 1 μL of

3′-adapter, 0.25 μL of Superase-In, 1 μL of T4 RNA ligase, and pipette up and down to mix. Incubate overnight at 16 °C.

3. Remove the excess of the adapter by adding 0.5 μL of 5′-deadenylase and 0.5 μL of Rec J exonuclease to the ligation reaction. Incubate at 30 °C for 30 min.
4. Add 30 μL of RNase-free water, 4 μL of 3 M NaOAc, 1 μL of glycogen, and 2.5 volumes of 100% ethanol. Incubate at −20 °C for at least 1 h to precipitate RNA.

3.10 Reverse Transcription

1. Centrifuge "footprint" and "mRNA" samples at 20,000 × *g* for 30 min at 4 °C, and remove the supernatant. Centrifuge at 20,000 × *g* for 30 s, remove the rest of the supernatant with a gel-loading tip, and air-dry the pellet for 10 min.
2. Resuspend the pellet in 11.5 μL of RNase-free water. Add 0.5 μL of 8 μM RT primer and 1 μL of 10 mM dNTPs. Incubate at 65 °C for 5 min, and put on ice.
3. Add 4 μL of 5× FS buffer (supplied with Superscript III reverse transcriptase), 2 μL of 0.1 M DTT, 0.5 μL of Superase-In, 0.5 μL of Superscript III reverse transcriptase. Incubate for 30 min at 48 °C, 1 min at 65 °C, 5 min at 80 °C.
4. Hydrolyze RNA, by adding 0.8 μL of 2 M NaOH, and incubate at 98 °C for 30 min. Add 0.8 μL 2 M HCl to neutralize the reaction.
5. Add 20 μL of RNase-free water, 4 μL of 3 M NaOAc, 1 μL of glycogen, and 2.5 volumes of 100% ethanol. Incubate at −20 °C for at least 1 h to precipitate DNA.
6. Centrifuge the samples at 20,000 × *g* for 30 min at 4 °C to pellet the DNA, and remove the supernatant. Centrifuge at 20,000 × *g* for 30 s, remove the rest of the supernatant with a gel-loading tip, and air-dry the pellet for 10 min.
7. Resuspend the pellet in 5 μL of RNase-free water. Add 5 μL of 2× TBE-urea sample buffer.
8. Prepare oligo sizing control containing 1 μL of 2.5 μM RT primer, 1 μL of 2.5 μM 128 nt marker oligonucleotide, 3 μL water, and 5 μL of 2× TBE-urea sample buffer.
9. Heat the samples at 75 °C for 3 min, spin down at max speed for 10 s, and put on ice for 1 min.
10. Rinse wells of the 10% TBE-urea gel from urea and pre-run the gel at 180 V for 15 min in 1× TBE buffer.
11. Load each sample into 1 well of the 10% TBE-urea gel, and run the gel at 180 V for 50 min.
12. Dissolve 5 μL of SYBR Gold in 50 mL of RNase-free water. Stain the gel with SYBR Gold for 5 min, and protect from light.

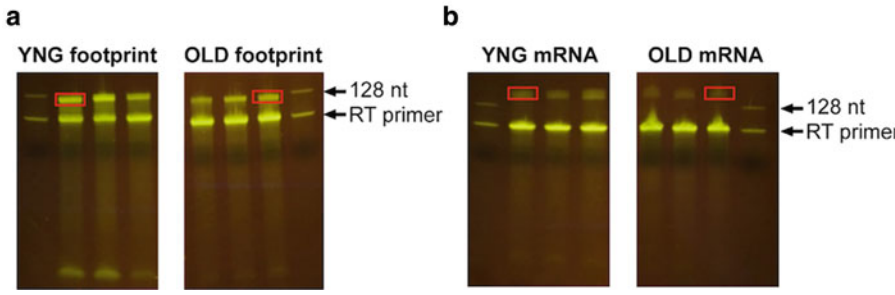

Fig. 4 Representative size selection gels used to isolate footprint and mRNA samples after reverse transcription. (**a**) Footprint samples obtained after reverse transcription were separated on 10% polyacrylamide TBU-urea gel. A mixture of RT primer and 128 nt marker oligonucleotide is used as a control. Cut the gel slice around 128 nt (upper band), which corresponds to the reverse transcription product. (**b**) For mRNA samples, cut just above the 128 nt marker. The size of the excised product should be around 150–170 nt

13. Using a blue light transilluminator, cut the gel slices around 128 nt and higher with a razor blade (Fig. 4). Freeze the polyacrylamide gel slices at −80 °C for 10 min.
14. Heat the gel slices at 70 °C for 2 min, and grind the gel using RNase-free disposable pellet pestles. Add 300 μL of 20 mM Tris–HCl pH 7.0, and incubate at 37 °C for 3 h to extract DNA.
15. Centrifuge the samples at 12,000 × *g* for 5 min at 4 °C and transfer the supernatant to a new 1.5 mL tube. Add 30 μL of 3 M NaOAc, 3 μL of glycogen, and 2.5 volumes of 100% ethanol. Incubate at −20 °C for at least 1 h to precipitate DNA.

3.11 Circularization

1. Centrifuge the samples at 20,000 × *g* for 30 min at 4 °C, and remove the supernatant. Centrifuge at 20,000 × *g* for 30 s, remove the rest of the supernatant with a gel-loading tip, and air-dry the pellet for 10 min.
2. Resuspend the pellet in 16.75 μL of RNase-free water. Add 2 μL of 10× ssDNA ligase buffer (supplied with enzyme), 1 μL of 50 mM $MnCl_2$, 0.25 μL of ssDNA ligase. Incubate at 60 °C for 1 h, and immediately heat at 80 °C for 10 min to inactivate the enzyme. Store the ssDNA ligation reaction product at −20 °C.

3.12 PCR Library Amplification

1. Set up PCR reactions on ice by mixing the following in a PCR tube: 146 μL of RNase-free water, 2 μL of 20 μM forward primer, 2 μL of 20 μM indexed reverse PCR primer, 40 μL of 5× Phusion high-fidelity DNA polymerase buffer, 4 μL of 10 mM dNTPs, 4 μL of ssDNA ligation reaction product, and 2 μL of Phusion high-fidelity DNA polymerase. Choose indexed reverse PCR primer specific for each sample. Pipette 50 μL of the PCR mixture into four different PCR tubes.

2. Perform the PCR amplification with varying number of cycles (6, 8, 10, 12) using the following settings:
   Initial denaturation: 1 min at 98 °C.

   Denaturation: 15 s at 94 °C.

   Annealing: 5 s at 55 °C.

   Elongation: 10 s at 65 °C.

   Final extension: 2 min at 65 °C

3. To each PCR product, add 5 μL of 3 M NaOAc, 1 μL of glycogen, and 2.5 volumes of 100% ethanol. Incubate at −20 °C for at least 1 h to precipitate DNA.
4. Centrifuge the samples at 20,000 × *g* for 30 min at 4 °C to pellet the DNA, and remove the supernatant. Centrifuge at 20,000 × *g* for 30 s, remove the rest of the supernatant with a gel-loading tip, and air-dry the pellet for 10 min.
5. Resuspend the pellet in 8 μL of RNase-free water, and add 2 μL of non-denaturing 5× DNA loading dye.
6. As a control, prepare a sample containing 0.5 μL of 10 bp DNA ladder, 7.5 μL water, and 2 μL of non-denaturing 5× DNA loading dye.
7. Pre-run the non-denaturing 8% TBE gel at 180 V for 15 min in 1× TBE buffer.
8. Load each sample into 1 well of the non-denaturing 8% TBE gel, and run the gel at 180 V for 35 min.
9. Dissolve 5 μL of SYBR Gold in 50 mL of RNase-free water. Stain the gel with SYBR Gold for 5 min, and protect from light.
10. Using a blue light transilluminator, cut the gel slices around 150 bp for "footprint" samples and around 180 bp for "mRNA" samples (Fig. 5) with a razor blade. Freeze the polyacrylamide gel slices at −80 °C for 10 min.
11. Heat the gel slices at 70 °C for 2 min, and grind the gel using RNase-free disposable pellet pestles. Add 300 μL of 20 mM Tris–HCl pH 7.0, and incubate at 37 °C for 3 h to extract DNA.
12. Centrifuge the samples at 12,000 × *g* for 5 min at 4 °C and transfer the supernatant to a new 1.5 mL tube. Add 30 μL of 3 M NaOAc, 3 μL of glycogen, and 2.5 volumes of 100% ethanol. Incubate at −20 °C for at least 1 h to precipitate DNA.
13. Centrifuge the samples at 20,000 × *g* for 30 min at 4 °C, and remove the supernatant. Centrifuge at 20,000 × *g* for 30 s, remove the rest of the supernatant with a gel-loading tip, and air-dry the pellet for 10 min.

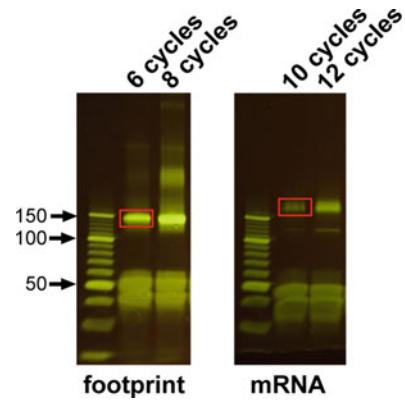

Fig. 5 Representative size selection gels used to isolate PCR amplified sequencing libraries. Following PCR amplification with varying number of cycles, the samples were separated on non-denaturing 8% polyacrylamide TBE gels. Select the desired number of cycles, in which the PCR products form a single, but bright enough band. Avoid lanes with high background. Cut the band ~150 bp for footprint libraries and 170–190 bp for mRNA libraries

14. Resuspend the amplified sequencing library in 20 μL of RNase-free water. Proceed to library quantification and high-throughput sequencing.

3.13 Library Quantification and High-Throughput Sequencing

1. The quantity and the size distribution of the amplified sequencing library can be determined using the Bioanalyzer High Sensitivity DNA assay. The expected size is 148–152 bp for "footprint" library and 170–190 bp for "mRNA" library [9].
2. We usually multiplex 12 samples labeled with individual barcodes in a single sequencing run. For multiplexing, perform accurate quantification of the libraries using a qPCR-based sequencing library quantification assay. Mix the libraries in equimolar ratios to achieve 10 nM final concentration of the pool. Store at −20 °C.
3. Sequence the pooled libraries using a single-end 50 bp run on an Illumina platform. Detailed instructions for downstream analysis of the Ribo-Seq data have been described previously [17]. For normalization of the Ribo-Seq sequencing data in young and old cells using an internal spike-in, *see* **Note 12**.

4 Notes

1. Handle biotin-labeled cells gently, and do not vortex.
2. From this step, adding antibiotics is not necessary. Extra care and aseptic techniques should be used to avoid contamination.

3. It is important that the OD_{600} of the cell culture does not exceed 1.0. We noticed that the efficiency of the mother enrichment program is decreased with time. Fresh estradiol needs to be added after 20-h incubation.
4. Sorting of cells should be performed on ice or in a cold room. Warming up the samples could cause the Dynabeads to lose biotin-binding capacity.
5. We typically recover ~50% of the starting number of cells. The population of old cells obtained after 30-h incubation on average contains ~11–13 more bud scars per cell compared to the young population.
6. It is important to always keep the sample frozen. Chill the tube in liquid nitrogen at least 10 s between each grinding cycle. If necessary, combine several replicates to achieve 1×10^8 cells per sample that will be divided equally between ribosomal footprint and mRNA libraries. We also found that adding flash-frozen droplets of lysis buffer to the sample makes grinding more efficient.
7. The amount of spike-in required for each sample is calculated based on OD_{260} values of the yeast lysates so that an equal amount of the worm lysate spike-in control (1%) is added into each sample proportional to the total RNA concentration. Alternatively, the same amount of the worm lysate spike-in could be added proportional to the cell number measurements (*see* Subheading 3.2, **step 25**) to allow normalization on a per-cell basis.
8. Other gradient makers or alternative methods for preparation of sucrose gradients can be used.
9. If fewer than six samples are being analyzed, attach all buckets, and arrange the filled tubes symmetrically in the rotor. Opposing tubes must be filled to the same level with the 10–50% sucrose gradient.
10. Once finished, clean the gradient fractionation system with RNase-free water. Thoroughly wash tubing and all removable components with warm water.
11. Caution: acid-phenol/chloroform is toxic; thus, avoid contact with skin and inhalation.
12. For normalization of the Ribo-Seq data in young and old cells based on overall changes in protein translation, we utilize internal controls using spike-in with worm lysates. Following demultiplexing and trimming of the 3′-adapter sequence AG ATCGGAAGAGCACACGTCT using Cutadapt software [18], sequencing reads are aligned to *S. cerevisiae* rRNA, tRNA, and sequences corresponding to *C. elegans* transcripts using Bowtie [19]. Reads that do not align are then mapped to the

S. cerevisiae genome, and the number of read counts per gene is determined by HTseq-count software [20]. To normalize the translation between the libraries, yeast RPKM (reads per kilobase per million mapped reads) values are adjusted using a global linear scaling factor calculated by linear regression of worm spike-in RPKM values between the libraries [21, 22].

Acknowledgments

This work was supported by the National Institutes of Health grants AG054566 and AG058713 to VML.

References

1. Koga H, Kaushik S, Cuervo AM (2011) Protein homeostasis and aging: the importance of exquisite quality control. Ageing Res Rev 10 (2):205–215. https://doi.org/10.1016/j.arr.2010.02.001
2. Yoshida H (2007) ER stress and diseases. FEBS J 274(3):630–658. https://doi.org/10.1111/j.1742-4658.2007.05639.x
3. Labbadia J, Morimoto RI (2015) The biology of proteostasis in aging and disease. Annu Rev Biochem 84:435–464. https://doi.org/10.1146/annurev-biochem-060614-033955
4. Kaeberlein M (2010) Lessons on longevity from budding yeast. Nature 464 (7288):513–519. https://doi.org/10.1038/nature08981
5. He C, Zhou C, Kennedy BK (2018) The yeast replicative aging model. Biochim Biophys Acta Mol basis Dis 1864(9 Pt A):2690–2696. https://doi.org/10.1016/j.bbadis.2018.02.023
6. Steinkraus KA, Kaeberlein M, Kennedy BK (2008) Replicative aging in yeast: the means to the end. Annu Rev Cell Dev Biol 24:29–54. https://doi.org/10.1146/annurev.cellbio.23.090506.123509
7. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324 (5924):218–223. https://doi.org/10.1126/science.1168978
8. Ingolia NT, Hussmann JA, Weissman JS (2019) Ribosome profiling: global views of translation. Cold Spring Harb Perspect Biol 11(5):a032698. https://doi.org/10.1101/cshperspect.a032698
9. Beaupere C, Chen RB, Pelosi W, Labunskyy VM (2017) Genome-wide quantification of translation in budding yeast by ribosome profiling. J Vis Exp (130):56820. https://doi.org/10.3791/56820
10. Li GW, Burkhardt D, Gross C, Weissman JS (2014) Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. Cell 157(3):624–635. https://doi.org/10.1016/j.cell.2014.02.033
11. Weinberg DE, Shah P, Eichhorn SW, Hussmann JA, Plotkin JB, Bartel DP (2016) Improved ribosome-footprint and mRNA measurements provide insights into dynamics and regulation of yeast translation. Cell Rep 14 (7):1787–1799. https://doi.org/10.1016/j.celrep.2016.01.043
12. Lindstrom DL, Gottschling DE (2009) The mother enrichment program: a genetic system for facile replicative life span analysis in Saccharomyces cerevisiae. Genetics 183(2):413–422, 411SI-413SI. https://doi.org/10.1534/genetics.109.106229
13. Smeal T, Claus J, Kennedy B, Cole F, Guarente L (1996) Loss of transcriptional silencing causes sterility in old mother cells of S. cerevisiae. Cell 84(4):633–642. https://doi.org/10.1016/s0092-8674(00)81038-7
14. Gonskikh Y, Polacek N (2017) Alterations of the translation apparatus during aging and stress response. Mech Ageing Dev 168:30–36. https://doi.org/10.1016/j.mad.2017.04.003
15. Hu Z, Xia B, Postnikoff SD, Shen ZJ, Tomoiaga AS, Harkness TA, Seol JH, Li W, Chen K, Tyler JK (2018) Ssd1 and Gcn2 suppress global translation efficiency in replicatively aged yeast while their activation extends lifespan. elife 7:e35551. https://doi.org/10.7554/eLife.35551
16. Henderson KA, Hughes AL, Gottschling DE (2014) Mother-daughter asymmetry of pH

underlies aging and rejuvenation in yeast. elife 3:e03504. https://doi.org/10.7554/eLife.03504

17. 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 (8):1534–1550. https://doi.org/10.1038/nprot.2012.086
18. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17(1):3. https://doi.org/10.14806/ej.17.1.200
19. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10(3):R25. https://doi.org/10.1186/gb-2009-10-3-r25
20. Anders S, Pyl PT, Huber W (2015) HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31(2):166–169. https://doi.org/10.1093/bioinformatics/btu638
21. Chen K, Hu Z, Xia Z, Zhao D, Li W, Tyler JK (2015) The overlooked fact: fundamental need for spike-in control for virtually all genome-wide analyses. Mol Cell Biol 36(5):662–667. https://doi.org/10.1128/MCB.00970-14
22. Wang YJ, Vaidyanathan PP, Rojas-Duran MF, Udeshi ND, Bartoli KM, Carr SA, Gilbert WV (2018) Lso2 is a conserved ribosome-bound protein required for translational recovery in yeast. PLoS Biol 16(9):e2005903. https://doi.org/10.1371/journal.pbio.2005903

Chapter 7

Tissue-Specific Ribosome Profiling in *Drosophila*

Xun Chen and Dion Dickman

Abstract

Robust mechanisms exist that serve to dynamically regulate the translation of mRNA into proteins across heterogeneous tissues. These processes ensure timely generation of proteins in quantities that scale with the demands of specific cell types. Importantly, this translational regulation occurs with spatiotemporal precision and is capable of recalibration as conditions change. Aberrant regulation of translation contributes to and exacerbates a wide range of diseases. Although dynamic control of translation is an essential and fundamental process shared by organisms, specific tissues and cell types can be differentially impacted by circumstances that challenge and impair basal translation, highlighting the heterogeneous nature of translational regulation. To understand how translation is differentially regulated during changing environments and across specific cells and tissues, methods capable of profiling translation in specific tissues and cells are crucial. Here, we describe a method for profiling genome-wide translation in specific tissues or cell types in *Drosophila melanogaster*, in which we combine ribosome affinity purification with ribosome profiling to enable a simplified protocol for robust analysis of translation in specific tissues.

Key words Translational regulation, Ribosome profiling, Genome-wide, Next-generation sequencing, *Drosophila*, Tissue-specific, Muscle

1 Introduction

Translational regulation is a fundamental process necessary for all cellular physiology. This regulation not only orchestrates essential functions during such core programs as embryonic development but is necessary throughout growth, maturation, and aging to modulate even subtle aspects of cellular physiology including, for example, synaptic function in mature neurons. Multiple mechanisms contribute to the regulation of mRNA translation, with modulation targeting the rates of translation initiation and elongation [1, 2], the interactions between mRNA and mRNA-binding proteins [3], and post-transcriptional modifications of mRNA [1, 4, 5]. Levels of expression of translational machinery varies across tissues [6], suggesting differential demands on translational output between tissues. Furthermore, abnormal translation has been linked to a variety of diseases affecting specific tissues such as fragile

Vyacheslav M. Labunskyy (ed.), *Ribosome Profiling: Methods and Protocols*, Methods in Molecular Biology, vol. 2252, https://doi.org/10.1007/978-1-0716-1150-0_7,

X syndrome [7], Diamond-Blackfan anemia [8], cancer [9], and other ribosomopathies [10–14]. These observations underscore the high degree of heterogeneity in cellular demands and variation in translational control among tissues.

Understanding the full complexity of translational regulation within an organism therefore requires tools and methods that are capable of profiling translation with tissue specificity. Here, we describe a method for profiling translation in specific tissues in the model organism *Drosophila melanogaster*. This approach leverages the recently developed technique called *ribosome profiling*, which utilizes a quantitative measurement of ribosome protected mRNA fragments (ribosome footprints) measured with next-generation sequencing to report translational activities [15]. The powerful method described here incorporates ribosome affinity purification into the workflow to enable selective isolation of ribosomes from tissue of interest in *Drosophila* followed by generation of sequencing library from the ribosome footprints [16]. In this method, ribosome affinity purification is achieved by expressing a 3xFlag tagged ribosomal protein, RpL3-3xFlag, in specific tissues followed by tissue lysis and pull down by anti-Flag antibody-coated magnetic beads in the presence of RNase (Fig. 1). The ribosome footprints are then isolated and converted into sequencing libraries for next-generation sequencing (Fig. 1) [16]. This method does not require ultracentrifugation or filtration-based systems to purify ribosomes, thus alleviating the technical complexities associated with those approaches. This method can also be adapted to perform RNA-seq on full length mRNAs associated with ribosomes (*see* Subheadings 3 and 4). Therefore, the method described here offers new opportunities to delineate translational mechanisms in defined tissues.

2 Materials

1. *Drosophila* line carrying the UAS-RpL3-3XFlag transgene, available as stock 77,132 at Bloomington Drosophila Stock Center.
2. Cycloheximide (100 mg/mL).
3. Protease inhibitor (EDTA-free).
4. Pellet pestles.
5. Triton X-100.
6. Protein G magnetic beads.
7. Anti-Flag antibody.
8. RNase T1 (1000 U/μL).
9. SUPERaseIn RNase Inhibitor (20 U/μL).

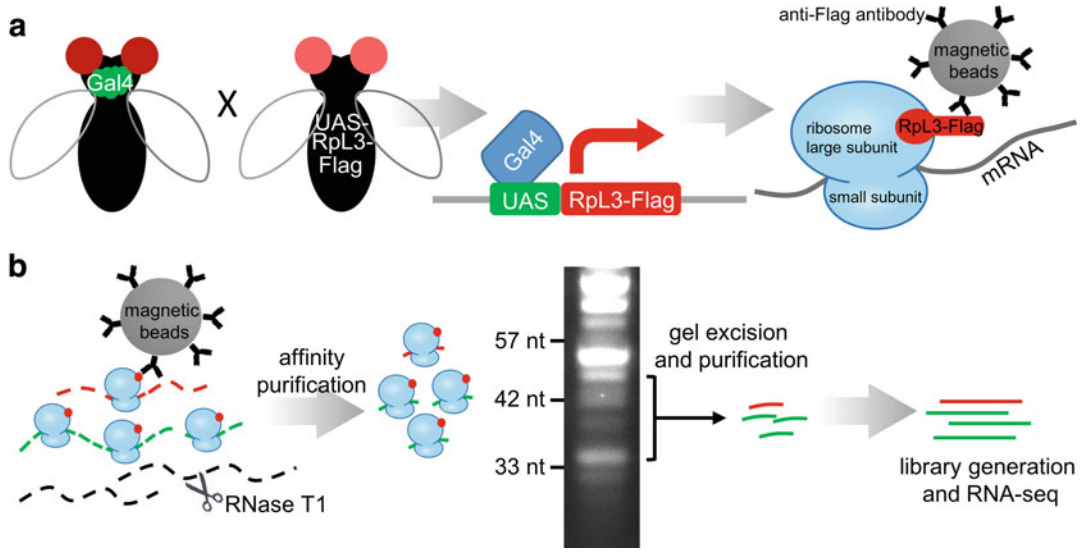

Fig. 1 Workflow for tissue-specific ribosome profiling. (**a**) The ribosome subunit RpL3 is tagged by a 3xFlag epitope and expressed in specific tissues using the Gal4/UAS system to enable affinity purification of ribosomes. (**b**) Tissue lysates are digested by RNase T1, and ribosomes are affinity purified by antibody-coated magnetic beads. Ribosome footprints are then isolated and converted into sequencing libraries

10. TRIzol reagent.
11. Chloroform.
12. Linear acrylamide (5 μg/μL).
13. Nonstick, RNase-free Microfuge Tubes, 1.5 mL.
14. 2× Novex TBE-urea sample buffer.
15. 15% TBE-urea polyacrylamide gel.
16. SYBR Green II RNA gel stain.
17. Blue light transilluminator.
18. Razor blade.
19. RNAsecure reagent.
20. 20% SDS solution.
21. Spin-X centrifuge tube filters (pore size 0.22 μm).
22. rSAP (Shrimp Alkaline Phosphatase, 1 U/μL).
23. NEBNext Multiplex Small RNA Library Prep Set for Illumina.
24. 10 mM ATP solution.
25. 50 mM DTT solution.
26. T4 PNK (T4 Polynucleotide Kinase, 10 U/μL).
27. 6% polyacrylamide gel.
28. SYBR Gold nucleic acid gel stain.

29. HL-3 saline: 5 mM HEPES, pH 7.2, 70 mM NaCl, 5 mM KCl, 10 mM $MgCl_2$, 10 mM $NaHCO_3$, 115 mM sucrose, 5 mM trehalose.
30. Lysis buffer: 10 mM HEPES, PH 7.4, 150 mM KCl, 5 mM $MgCl_2$, 100 μg/mL cycloheximide, 1× protease inhibitor (EDTA-free).
31. Wash buffer: 10 mM HEPES, PH 7.4, 150 mM KCl, 5 mM $MgCl_2$, 100 μg/mL cycloheximide, 0.1% Triton-X100, 0.1 U/μL SUPERaseIn RNase inhibitor.
32. Gel elution buffer: 10 mM Tris–HCl, pH 7.5, 250 mM NaCl, 1 mM EDTA.
33. 5× rSAP buffer: 20 mM Tris–HCl, pH 8.0, 250 mM NaCl, 1 mM $MgCl_2$.

3 Methods

3.1 Sample Collection

This method is based on expression of the 3xFlag-tagged ribosomal protein, RpL3, under the control of Gal4-dependent upstream activation sequence (UAS) (Fig. 1a). A Gal4 driver line expressed in specific cells or tissues is crossed to the UAS-RpL3-3xFlag line to enable expression of epitope-tagged ribosomes that can be purified and processed to analyze translation in specific tissues (Fig. 1b). Relevant tissues are collected and lysed before digestion by RNase. The following is a protocol based on ribosome profiling of muscle tissue from third instar *Drosophila* larvae [16] but can be easily adapted to other tissues and life stages of the fly.

1. Using a Gal4 line that expresses in the tissue of interest, cross this to the UAS-RpL3-3xFlag line, for example, the muscle driver BG57-Gal4 (also known as C57) (*see* **Note 1**) [16].
2. Pick wandering third instar larvae, and dissect in HL-3 saline to collect body walls that contain the muscle tissue with all internal organs removed. After dissection, immediately place tissue in a 1.7 mL centrifuge tube and freeze on dry ice; tissue samples can be stored at −80 °C for a month. Collect a total of eight body walls in one tube for each sample (*see* **Note 2**).
3. Add 240 μL of lysis buffer and thoroughly grind tissues using pellet pestles. Add 12 μL of 10% Triton X-100. Mix well then rotate the tube at 4 °C for 30 min (*see* **Note 3**).
4. Centrifuge the lysate for 10 min at 15,000 × *g* at 4 °C to clear the lysate, and then transfer supernatant to a new tube. Immediately proceed to Subheading 3.2, **step 1**.

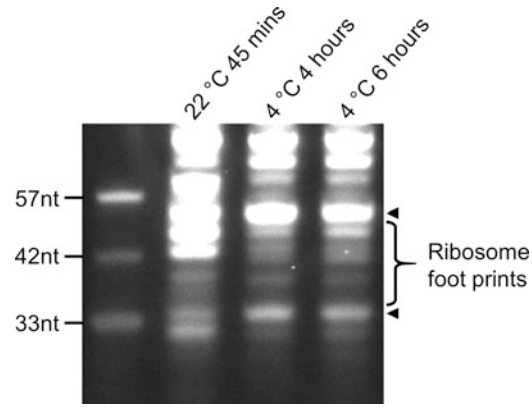

Fig. 2 Optimization of RNA digestion in tissue lysates. Polyacrylamide gel electrophoresis of RNA extracted from tissue lysates and digested by RNase T1 at the indicated temperature and duration above. 6 h digestion at 4 °C gives the best yield without excessive degradation of ribosomal RNA

3.2 Generation of Ribosome Protected mRNA Fragments (Ribosome Footprints)

The tissue lysates are incubated with magnetic beads coated with anti-Flag antibodies along with RNase T1 to simultaneously perform ribosome affinity purification and mRNA digestion.

1. Coat 100 μL protein G magnetic beads (*see* **Note 4**) with 4 μg of anti-Flag antibody according to manufacturer's instructions. Mix cleared lysate with the beads and add 10 μL of RNase T1. Rotate the mixture at 4 °C for 6 h (*see* **Note 5** and Fig. 2).
2. Place the tube on a magnet (such as the DynaMag™-2 Magnet) for 1 min, and remove all liquid. Wash beads with 200 μL of wash buffer by pipetting up and down 10 times, performing a total of three washes, and then discard all liquid.
3. Extract RNA from washed beads by adding 500 μL of TRIzol reagent, thoroughly re-suspend beads in TRIzol, and then incubate at room temperature for 5 min. Place the tube on the magnet, and transfer TRIzol solution to a 1.7 mL centrifuge tube.
4. Add 0.1 mL of chloroform, and vigorously shake the tube for 30 s to thoroughly mix, and then incubate at room temperature for 2 min.
5. Centrifuge the sample for 15 min at 12,000 × *g* at 4 °C. Carefully transfer the liquid phase above to a 1.5 mL Nonstick, RNase-free Microfuge Tube (*see* **Note 6**).
6. Add 5 μL of linear acrylamide to the liquid phase (*see* **Note 7**), then add 250 μL of isopropanol, mix well, and incubate for 10 min at room temperature.
7. Centrifuge for 10 min at 12,000 × *g* at 4 °C. RNA precipitate forms a semi-transparent pellet at the bottom of the tube. Remove all liquid, and wash the RNA pellet with 1 mL of

75% ethanol by gently inverting the tube without breaking the precipitate; the pellet should turn white upon adding 75% ethanol.

8. Remove all liquid (*see* **Note 8**), and air dry the pellet for 5 min by leaving the tube open. Dissolve pellet in 7 μL nuclease-free water (*see* **Note 9** if conducting ribosome associated RNA sequencing).
9. Mix RNA sample above with 7 μL 2× Novex TBE-urea sample buffer. Heat the mixture at 70 °C for 2 min, and then immediately place on ice.
10. Run the sample from above on a 15% TBE-urea polyacrylamide gel along with DNA oligo markers for 90 min at 200 V. Stain the gel with SYBR Green II RNA gel stain (diluted to 1× concentration in TBE buffer) for 15 min.
11. Place the gel on a blue light transilluminator, cut the gel region corresponding to the 30–50 nt range as indicated by oligo markers (Fig. 2 and **Note 10**) with a razor blade, and place the gel piece in a 1.7 mL centrifuge tube.
12. Add 200 μL of gel elution buffer. Grind the gel with pellet pestle thoroughly. Add another 300 μL of gel elution buffer, 22 μL of RNAsecure reagent, and 5 μL of 20% SDS. Mix well and then heat at 60 °C for 10 min.
13. Rotate the tube at 4 °C overnight. Transfer eluate with gel pieces to a Spin-X centrifuge tube filter, and centrifuge for 5 min at 16,000 × *g* at 25 °C. Transfer the cleared eluate to a new 1.7 mL centrifuge tube.
14. Perform RNA precipitation by adding 500 μL of isopropanol, 7 μL of linear acrylamide to the tube and mix well. Incubate at room temperature for 30 min, and then centrifuge for 15 min at 15,000 × *g* at 25 °C. Wash the pellet with 75% ethanol once, discard all liquid, and air dry the pellet for 5 min. Dissolve the pellet in 16 μL of nuclease-free water; this is the gel purified ribosome footprints (RFs) sample (*see* **Note 11**).

3.3 Phosphatase Treatment and 3′ Ligation for Library Generation

The process of generating the library for RNA-seq from ribosome footprints loosely follows a standard small RNA library generation protocol with additional phosphatase and kinase treatments. Purified ribosome footprints contain 3′ phosphate groups and 5′ hydroxyl groups. They are first treated by phosphatase to remove the 3′ phosphate group and are then ligated with 3′ adaptor oligos.

1. Perform phosphatase treatment of purified ribosome footprints by combining the following in a PCR tube:

15 μL	Gel purified RFs
10 μL	5× rSAP buffer
2.5 μL	rSAP
2.5 μL	SUPERaseIn
20 μL	Nuclease-free water

Mix and then incubate at 37 °C for 30 min.

2. Add 2 μL of RNAsecure, 0.5 μL of 500 mM EDTA to the phosphatase treatment reaction. Mix and then heat at 65 °C for 10 min to inactivate rSAP.
3. Add 50 μL of isopropanol. Mix and incubate at room temperature for 10 min, and then centrifuge for 10 min at 15,000 × *g* at 4 °C. Wash pellet with 75% ethanol, discard all liquid, and air dry pellet for 5 min. Dissolve the pellet in 7 μL Nuclease-free water.
4. Perform 3′ adaptor ligation using the corresponding component of NEBNext Multiplex Small RNA Library Prep Set for Illumina and follow manufacturer's instructions. Briefly, mix 6 μL of phosphatase treated RNA with 1 μL of 3′ SR Adaptor for Illumina. Heat the tube at 70 °C for 2 min, and then immediately transfer to ice. Add 10 μL of 3′ ligation reaction buffer, 3 μL of 3′ ligation enzyme mix to the tube. Mix and then Incubate at 25 °C for 1 h.

3.4 Kinase Treatment, 5′ Ligation and Reverse Transcription for Library Generation

After 3′ adaptor ligation, the footprint RNAs are treated by T4 Polynucleotide Kinase (T4 PNK) to add a phosphate group to their 5′ end followed by 5′ ligation and reverse transcription.

1. Perform the kinase treatment by adding the following to the 3′ ligation reaction:

2.5 μL	ATP (10 mM)
1.5 μL	DTT (50 mM)
0.5 μL	T4 PNK

Mix and incubate at 37 °C for 30 min.

2. Perform reverse transcription primer hybridization by adding 1 μL of SR RT Primer for Illumina to the kinase treatment reaction, and mix well. Heat the sample for 5 min at 75 °C, and then allow the sample to slowly cool down in the thermocycler until the temperature drops to below 35 °C.

3. Perform 5′ SR adaptor ligation using the corresponding component of the NEBNext Multiplex Small RNA Library Prep Set for Illumina, and follow manufacturer's instructions. Briefly, denature enough 5′ SR Adaptor for all samples, and add the following to the reverse transcription primer hybridization reaction:

1 μL	Denatured 5′ SR adaptor for Illumina
1 μL	5′ Ligation reaction buffer
2.5 μL	Ligation enzyme

Mix well and then incubate at 25 °C for 1 h.

4. Perform reverse transcription using NEBNext Multiplex Small RNA Library Prep Set for Illumina, and follow manufacturer's instructions. Briefly, add the following to the 5′ SR adaptor ligation reaction:

8 μL	First strand synthesis reaction buffer
1 μL	Murine RNase inhibitor
1 μL	ProtoScript II reverse transcriptase

Mix well and then incubate at 50 °C for 1 h.

3.5 Library PCR Amplification and Purification

The library DNA is amplified, and index sequences are incorporated to each sample during the PCR amplification step. The library DNA is then purified using 6% polyacrylamide gel electrophoresis.

1. Perform PCR amplification (15–20 cycles, *see* **Note 12**) using NEBNext Multiplex Small RNA Library Prep Set for Illumina and follow manufacturer's instructions. Briefly, add the following to the reverse transcription reaction:

50 μL	LongAmp Taq 2× Master Mix
2.5 μL	SR Primer for Illumina
2.5 μL	Index Primer
5 μL	Nuclease-free water

Mix well and then perform PCR cycling using condition: 94 °C 30 s, 15 cycles of [94 °C 15 s, 62 °C 30 s, 70 °C 15 s], 70 °C 5 min, hold at 4 °C.

2. Perform DNA precipitation by adding the following to the 100 μL of PCR product:

3 μL	linear acrylamide
10 μL	NaCl (2.5 M)
100 μL	Isopropanol

Mix well and then incubate at room temperature for 30 min. Centrifuge for 10 min at 15,000 × *g* at 4 °C. Wash the pellet with 75% ethanol once, discard all liquid, and air dry the pellet for 5 min. Dissolve the pellet in 10 μL nuclease-free water.

3. Perform size selection of purified library PCR product using the corresponding component of the NEBNext Multiplex Small RNA Library Prep Set for Illumina and follow manufacturer's instructions. Briefly, mix 10 μL PCR product with 2 μL Gel Loading Dye, Blue. Run the sample on a 6% polyacrylamide gel along with 5 μL of Quick-Load pBR322 for 1 h at 120 V. Stain the gel with SYBR Gold nucleic acid gel stain (diluted to 1× in TBE buffer) for 15 min.
4. View the gel under a transilluminator, cut the gel region corresponding to 140–170 bp as indicated by the Quick-Load pBR322 markers (Fig. 3), and place the gel piece in a 1.7 mL centrifuge tube.

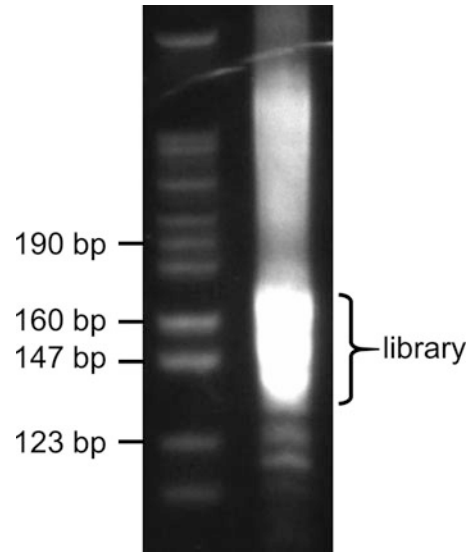

Fig. 3 Library size selection. The sequencing library is run on a 6% polyacrylamide gel to isolate the correct library DNA, a size targeted between ~130 and 170 base pairs (bp) is desired. The region to be excised is indicated in the gel

5. Add 200 μL of gel elution buffer to the tube. Grind the gel with pellet pestle. Add another 300 μL of gel elution buffer and 5 μL of 20% SDS. Mix well and rotate the tube at 4 °C overnight.
6. Transfer eluate with gel pieces to a centrifuge tube filter, and centrifuge for 5 min at 16,000 × *g* at 25 °C. Transfer the cleared eluate to a new 1.7 mL centrifuge tube.
7. Perform DNA precipitation by adding 500 μL of isopropanol, 5 μL of linear acrylamide to the tube and mix well. Incubate at room temperature for 30 min, and then centrifuge for 15 min at 15,000 × *g* at 25 °C. Wash the pellet with 75% ethanol once, discard all liquid, and air dry the pellet for 5 min. Dissolve the pellet in 15 μL of nuclease-free water; this is the library sample (*see* **Note 13**).
8. Submit the library sample for next-generation sequencing on an Illumina platform or store at −20 °C (*see* **Note 14**).

3.6 Recommendations for Next-Generation Sequencing and Data Analysis

1. Single read 50–75 cycles sequencing format is recommended. This read length is sufficient to cover the ribosome footprint RNA, which averages 35 nt (Fig. 4).
2. Reads generated by the sequencer should first be adaptor trimmed to remove adaptor sequences and then trimmed based on quality score with a minimal Q score cutoff of 20, and only retain reads that are at least 20 nt.

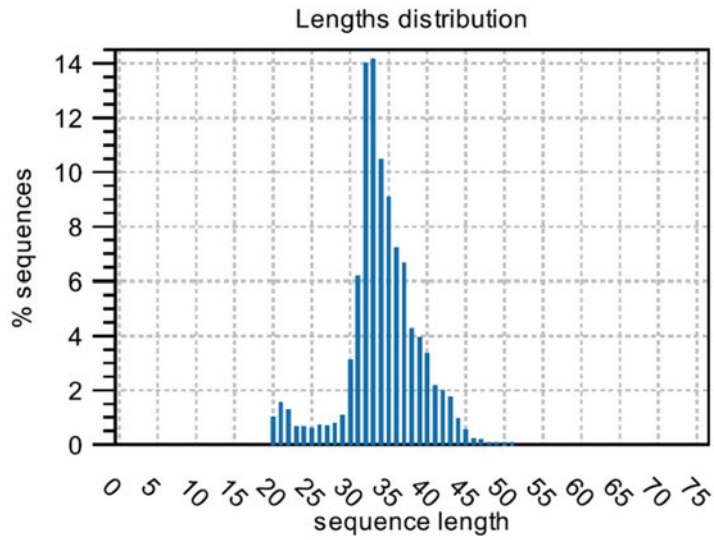

Fig. 4 Distribution of sequence lengths of ribosome footprints. Ribosome footprints verified by mapping have lengths distributed between 30 and 50 nucleotides, with a peak at 33

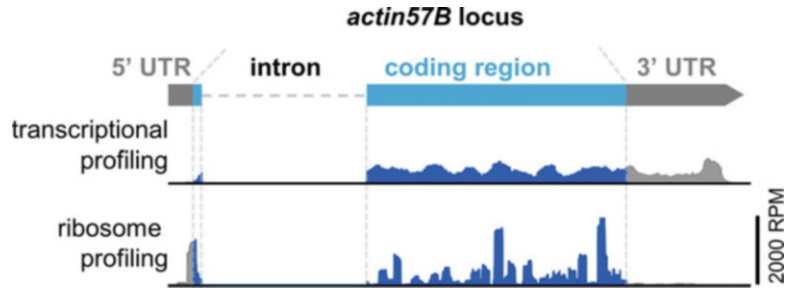

Fig. 5 Ribosome profiling reads predominantly map to genomic coding regions. Coverage of ribosome profiling reads is high in coding regions in the *Drosophila* genome and drops sharply after stop codons. The 3′ untranslated regions (UTRs) have comparatively little coverage by ribosome profiling reads. In contrast, transcriptional profiling using RNA-seq of total RNA exhibit similar levels of coverage across 5′ UTR, coding, and 3′ UTR genomic regions

3. The trimmed reads should then be mapped to the *Drosophila* genome to calculate the number of mapped reads per mRNA transcript. The commercially available software CLC Genomics Workbench (QIAGEN) can be used for reads trimming and mapping.
4. To evaluate how well the sequencing reads represent real ribosome footprint versus free mRNA fragments, check the reads coverage around the predicted stop codon of representative genes, or use a meta-analysis of all translated genes. The reads coverage should drop sharply from high coverage within the coding sequence to very low coverage in the 3′ untranslated regions (UTRs) after the stop codon (Fig. 5).
5. Normalized ribosome profiling reads, typically in RPKM (Reads Per Kilobase of transcript, per Million mapped reads), can be used as a proxy for relative translation level.
6. If RNA-seq of total RNA was carried out in parallel, translation efficiency can be calculated as RPKM from ribosome profiling reads divided by RPKM from total RNA RNA-seq reads.

4 Notes

1. For a given tissue of interest, Gal4 lines that are strongly expressed should be prioritized. Weak and/or leaky drivers will compromise specificity as there is a low level of non-specific binding of untagged ribosomes to magnetic beads during purification.
2. To increase tissue specificity, remove as much of the surrounding tissues that are irrelevant to the analysis as possible.
3. Triton X-100 should be added after tissue grinding to avoid excessive foam formation during this process.

4. Scale the amount of beads to the volume of tissue, with care to avoid the use of excessive beads. This will help to reduce non-specific ribosome binding, as the level of background ribosome binding will be proportional to the volume of beads.
5. RNase T1 was used to minimize degradation of ribosomal RNA. RNase T1 cleaves single-stranded RNA at G residues, and therefore are less prone to cleave ribosomal RNA compared to RNase I. In addition, digestion at 4 °C instead of 37 °C further reduces ribosomal RNA cleavage (Fig. 2). Due to the base preference of RNase T1, ribosome footprints produced by RNase T1 digestion are slightly longer than what was reported for footprints generated by RNase I, with an average length of 35 nt (Fig. 4).
6. Using nonstick tubes helps the formation of more compact pellets and reduces the likelihood of accidental sample loss.
7. Linear acrylamide acts as a co-precipitate, the inclusion of linear acrylamide during isopropanol precipitation increases recovery of nucleic acids.
8. It is important to remove as much liquid as possible after the 75% ethanol wash. First discard most of the wash volume using a 1 mL pipet tip, and then briefly centrifuge the tube to concentrate any droplet on the side of the tube to the bottom. Then use a 10–20 μL pipet tip to discard all liquid collected at the bottom of the tube. Use this procedure for all 75% ethanol wash steps described in this method.
9. If performing full-length sequencing of ribosome associated RNA, perform all previous steps with two modifications to Subheading 3.2, **step 1**: first, do not add RNase T1; second, the duration of the 4 °C incubation can be reduced to 2 h. Full-length ribosome-associated RNA can be processed by standard RNA-seq library generation protocols for next-generation sequencing.
10. RNase T1 digestion produces a stereotypical pattern of bands, with the two bands indicated by black arrow heads shown in Fig. 2 used as internal gel excision boundary markers. Do not include the two strong bands in the excised gel, as they are ribosomal RNA fragments and will cause a high percentage of ribosomal RNA reads in the sequencing result.
11. This method does not require ribosomal RNA removal due to a reduced abundance of ribosomal RNA in the ribosome footprints sample. However, if necessary, ribosomal RNA depletion by oligos complementary to the ribosomal RNA may be included after this step to reduce ribosomal RNA reads. Without the removal of ribosomal RNA, 10–20% of reads are mRNA reads.

12. The number of PCR cycles used should be minimized, typically falling between 15 and 20. Using amount of input sample detailed by this protocol, 15 PCR cycles should provide a robust yield of library DNA.
13. An alternative protocol to generate the library starting from purified ribosome footprints (Subheading 3.2, **step 14**) can also be used [15].
14. The following is the estimated time required to complete Subheadings 3.1–3.5. Subheading 3.1: 3 h. Subheading 3.2: 2.5 days. Subheading 3.3: 3 h. Subheading 3.4: 4 h. Subheading 3.5: 1.5 days.

Acknowledgments

This work was supported by a National Institutes of Health grant (R01NS111414) to DD.

References

1. Kong J, Lasko P (2012) Translational control in cellular and developmental processes. Nat Rev Genet 13(6):383–394. https://doi.org/10.1038/nrg3184
2. Richter JD, Coller J (2015) Pausing on polyribosomes: make way for elongation in translational control. Cell 163(2):292–300. https://doi.org/10.1016/j.cell.2015.09.041
3. Glisovic T, Bachorik JL, Yong J, Dreyfuss G (2008) RNA-binding proteins and post-transcriptional gene regulation. FEBS Lett 582(14):1977–1986. https://doi.org/10.1016/j.febslet.2008.03.004
4. Choe J, Lin S, Zhang W, Liu Q, Wang L, Ramirez-Moya J, Du P, Kim W, Tang S, Sliz P, Santisteban P, George RE, Richards WG, Wong KK, Locker N, Slack FJ, Gregory RI (2018) mRNA circularization by METTL3-eIF3h enhances translation and promotes oncogenesis. Nature 561(7724):556–560. https://doi.org/10.1038/s41586-018-0538-8
5. Meyer KD, Patil DP, Zhou J, Zinoviev A, Skabkin MA, Elemento O, Pestova TV, Qian SB, Jaffrey SR (2015) 5′ UTR m(6)A promotes Cap-independent translation. Cell 163(4):999–1010. https://doi.org/10.1016/j.cell.2015.10.012
6. Geiger T, Velic A, Macek B, Lundberg E, Kampf C, Nagaraj N, Uhlen M, Cox J, Mann M (2013) Initial quantitative proteomic map of 28 mouse tissues using the SILAC mouse. Mol Cell Proteomics 12(6):1709–1722. https://doi.org/10.1074/mcp.M112.024919
7. Richter JD, Bassell GJ, Klann E (2015) Dysregulation and restoration of translational homeostasis in fragile X syndrome. Nat Rev Neurosci 16(10):595–605. https://doi.org/10.1038/nrn4001
8. Ludwig LS, Gazda HT, Eng JC, Eichhorn SW, Thiru P, Ghazvinian R, George TI, Gotlib JR, Beggs AH, Sieff CA, Lodish HF, Lander ES, Sankaran VG (2014) Altered translation of GATA1 in diamond-Blackfan anemia. Nat Med 20(7):748–753. https://doi.org/10.1038/nm.3557
9. Aspesi A, Ellis SR (2019) Rare ribosomopathies: insights into mechanisms of cancer. Nat Rev Cancer 19(4):228–238. https://doi.org/10.1038/s41568-019-0105-0
10. Bolze A, Mahlaoui N, Byun M, Turner B, Trede N, Ellis SR, Abhyankar A, Itan Y, Patin E, Brebner S, Sackstein P, Puel A, Picard C, Abel L, Quintana-Murci L, Faust SN, Williams AP, Baretto R, Duddridge M, Kini U, Pollard AJ, Gaud C, Frange P, Orbach D, Emile JF, Stephan JL, Sorensen R, Plebani A, Hammarstrom L, Conley ME, Selleri L, Casanova JL (2013) Ribosomal protein SA haploinsufficiency in humans with isolated congenital asplenia. Science 340(6135):976–978. https://doi.org/10.1126/science.1234864

11. Farley-Barnes KI, Ogawa LM, Baserga SJ (2019) Ribosomopathies: old concepts, new controversies. Trends Genet 35(10):754–767. https://doi.org/10.1016/j.tig.2019.07.004
12. Khajuria RK, Munschauer M, Ulirsch JC, Fiorini C, Ludwig LS, McFarland SK, Abdulhay NJ, Specht H, Keshishian H, Mani DR, Jovanovic M, Ellis SR, Fulco CP, Engreitz JM, Schutz S, Lian J, Gripp KW, Weinberg OK, Pinkus GS, Gehrke L, Regev A, Lander ES, Gazda HT, Lee WY, Panse VG, Carr SA, Sankaran VG (2018) Ribosome levels selectively regulate translation and lineage commitment in human hematopoiesis. Cell 173 (1):90–103. e119. https://doi.org/10.1016/j.cell.2018.02.036
13. Mills EW, Green R (2017) Ribosomopathies: there's strength in numbers. Science 358 (6363). https://doi.org/10.1126/science.aan2755
14. Tahmasebi S, Khoutorsky A, Mathews MB, Sonenberg N (2018) Translation deregulation in human disease. Nat Rev Mol Cell Biol 19 (12):791–807. https://doi.org/10.1038/s41580-018-0034-x
15. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324 (5924):218–223. https://doi.org/10.1126/science.1168978
16. Chen X, Dickman D (2017) Development of a tissue-specific ribosome profiling approach in Drosophila enables genome-wide evaluation of translational adaptations. PLoS Genet 13 (12):e1007117. https://doi.org/10.1371/journal.pgen.1007117

Chapter 8

Measuring Organ-Specific Translation Elongation Rate in Mice

Maxim V. Gerashchenko and Vadim N. Gladyshev

Abstract

Modern methods of genome editing enable the rapid generation of mouse models to study the regulation of protein synthesis. At the same time, few options are available to study translation in rodents as the animal's complexity severely limits the repertoire of experimental tools. Here we describe a method to monitor translation in mice and other small animals. The technique is based on a ribosome profiling and specifically tailored toward measuring translation elongation. However, it can be easily applied for short upstream reading frames discovery. The advantage of this method is the ability to study translation in fully developed animals without extracting and subculturing cells, therefore, maintaining unperturbed physiological conditions.

Key words Ribosome profiling, Translation, Elongation, Mouse, Organ, Injection, Harringtonine

1 Introduction

Due to unprecedented breakthroughs in genome editing methods in the past decade, it is now easier than ever to study fine details of protein synthesis in mammalian animal models. Most of our understanding of translation comes from unicellular organisms (bacteria and budding yeast) and cultured mammalian cells. Modern in vitro methods are capable of tracking real-time translation dynamics of a single molecule and making transcriptome-wide snapshots of translation with a single nucleotide resolution by ribosome profiling [1–3]. While it greatly advanced our understanding of molecular mechanisms of protein synthesis, there is an ever-growing demand for methods applicable to whole vertebrate organisms. Some congenital human disorders are associated with dysfunctional translation machinery, mutations introducing premature stop codons, and hard-to-translate repetitive codons on the mRNA. Examples include Blackfan-Diamond anemia [4] and cystic fibrosis [5]. Several approaches to treating these diseases suggest intentionally decreasing accuracy or the rate of protein synthesis. Slowing

Vyacheslav M. Labunskyy (ed.), *Ribosome Profiling: Methods and Protocols*, Methods in Molecular Biology, vol. 2252, https://doi.org/10.1007/978-1-0716-1150-0_8,

ribosome velocity partially restored folding of the mutant cystic fibrosis transmembrane conductance regulator protein in the mouse model [5].

To examine the effects of these therapies on translation in vivo, we developed a technique based on the high-throughput sequencing (ribosome profiling). It relies on the consequent intravenous delivery of two translation inhibitors – harringtonine and cycloheximide. Harringtonine inhibits translation at the initiation stage of translation whereas cycloheximide at elongation [6]. It has to be noted that there are two initiation-specific translation inhibitors commercially available: harringtonine and lactimidomycin [7]. However, harringtonine is the only one available at a scale required for injecting many mice. According to the in vitro comparison of these inhibitors, lactimidomycin is more specific, while harringtonine allows some ribosomes to slip past the start codon resulting in less sharp ribosomal footprint peaks at the start codons [7]. The difference is not prominent enough to affect the elongation rate estimates on the time scale of experiments discussed here.

The pair of translation inhibitors was first used by Weissman lab to measure translation rates in a culture of mouse embryonic cells [2]. We adapted it to work in animals by injecting into the bloodstream [8]. First, we inject harringtonine and stall newly initiating ribosomes at the translation start sites and few amino acids downstream of it. Second, after a specified time interval (under 1 min), we inject cycloheximide to freeze elongating ribosomes in place. These interventions are done to alive sedated mice. According to our observations, both drugs rapidly spread across all major organs unprotected by blood-tissue barriers. After that, organs can be collected, and translation snapshots acquired by standard ribosome profiling. The time-dependent gap in ribosome occupancy over mRNA is representative of the elongation rate. Depending on the number of time points, average or gene-specific elongation rates can be inferred.

Little is known about drug diffusion rates in different mouse organs. Typically, the rapid rate of diffusion is not a major concern for pharmacological studies, and it depends on the chemical structure, charge, and the ability to pass through cellular membranes. It is reasonable to expect different drug uptake rates across organs and even cell types based on their adjacency to capillaries and many other factors. To avoid these issues, the elongation rates' calculations do not include the no-drugs time point. Upon closer inspection of aggregate ribosome coverage profiles from various organs and cell types (Fig. 1), we certainly see the difference in the width of the gap between the no-drug profile and the earliest injection time point. This gap reflects the time required for translation inhibitors to permeate cells and interact with ribosomes. This organ-specific delay applies to all time points obtained for that organ; therefore

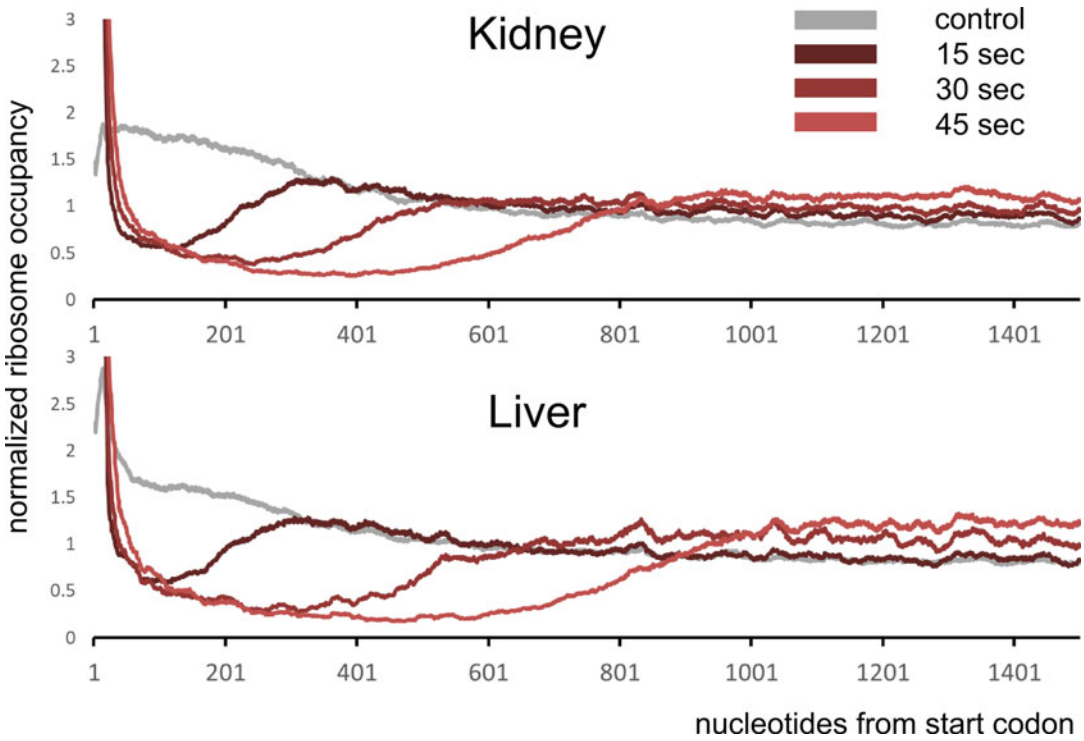

Fig. 1 Representative time-dependent ribosome occupancy plots for the kidney and the liver. 3 time points were tested: 15, 30, and 45 s between harringtonine and cycloheximide injections

the ribosomal run-off distance between time points is not affected by this delay. Consequently, the minimal number of injection time points is 3, and the no-injection control is not strictly required.

This method is not specifically designed to search for alternative translation initiation sites. The use of harringtonine and subsequent ribosomal run-off leads to saturation of the proper translation start sites, while there is a surplus of the available ribosomes that would initiate at anything that even remotely resembles a start site.

In this chapter, we present a detailed experimental protocol for measuring organ-specific translation elongation rates along with a discussion of possible advantages and disadvantages.

2 Materials

2.1 Animal Injections

1. Phosphate-buffered saline (PBS).
2. Dimethyl sulfoxide (DMSO).
3. Harringtonine: 100 mg/mL solution in DMSO.
4. Cycloheximide: 20 mg/mL solution in PBS.
5. Optional: physiological saline as a replacement for PBS.
6. Low flow vaporizer for isoflurane (*see* **Note 1**).

7. Oxygen tank.
8. Isoflurane.
9. Catheters suitable for the tail vein.
10. 0.3 mL insulin syringes.
11. Rodent anesthesia system (i.e., mouse gas dispensing mask, tubing, heating pad, heating lamp).
12. ECG monitor to track heart beat rate.

2.2 Tissue Lysis

1. Potter-Elvehjem PTFE pestle and glass tube (3 mL).
2. Ceramic mortar and motorized pestle.
3. Liquid nitrogen.
4. Lysis buffer: 20 mM Tris–HCl pH 7.5, 50 mM KCl, 50 mM NaCl, 5 mM $MgCl_2$, 5 mM $CaCl_2$, 1 mM DTT, 1% Triton, 0.1 mg/mL cycloheximide.
5. Detergents: Triton, deoxycholate, Tween-20.

2.3 Sucrose Gradients Fractionation

1. Polypropylene centrifuge tubes (14 × 89 mm).
2. Beckman ultracentrifuge compatible with SW41 rotor.
3. Fraction collection system (optional).
4. Tube piercing stand.
5. UV detector with 254 nm filter (separate or as a part of an automated gradient fractionator).
6. Syringe pump.
7. BioComp Gradient Master.
8. Gradient sucrose solutions: 10% and 50% sucrose in polysome buffer without detergents, no $CaCl_2$, and increased $MgCl_2$ to 10 mM.
9. Sucrose chase solution: 60% sucrose, 10 mM $MgCl_2$, 100 mg/mL cycloheximide.
10. 100 kDa filter units.

2.4 Footprint Extraction and Library Preparation

1. RNase I 100 U/μL.
2. RNase S7 10 mg/mL. RNase S7 is shipped as lyophilized powder, prepare a 10 mg/mL stock solution in 10 mM Tris–HCl pH 7.5, 50 mM NaCl, 1 mM EDTA, 50% glycerol.
3. RNase T1 1000 U/μL.
4. T4 RNA ligase 2 truncated 200 U/μL.
5. 5′ Deadenylase 50 U/μL.
6. Lambda nuclease 5 U/μL.
7. SuperScript III 200 U/μL.
8. 0.5 M Ethylene glycol-bis(β-aminoethyl ether)-*N*,*N*,*N*′,*N*′-tetraacetic acid (EGTA) solution.

9. EDTA-free protease inhibitors cocktail.
10. Superase-In RNase inhibitor.
11. Trizol LS reagent.
12. Direct-zol RNA purification kit.
13. 0.22 μm cellulose acetate centrifuge filter.
14. Phusion DNA polymerase.
15. Blue pestles that fit 1.5 mL microtube.
16. 10% and 15% denaturing polyacrylamide TBE-urea gels.
17. 10× TBE running buffer.
18. TBE-urea sample buffer.
19. SYBR Gold.
20. Glycogen (5 mg/mL).
21. 100% ethanol.
22. RNA gel elution buffer: 10 mM Tris–HCl pH 7.0, 1/10 volume of 3 M sodium acetate pH 5.5, 2 mM EDTA, 1 μL of Superase-In per 1 mL.
23. Blue light transillumination box.

2.5 Adapter and PCR Primer Sequences

1. Preadenylated ligation adapter.
   5′-/5rApp/AGATCGGAAGAGCACACGTCT-3′.
2. Reverse transcription oligonucleotide.
   5′-pGATCGTCGGACTGTAGAACTCTGAACGTGTAG ATCTCGGTGGTCGCCGTATCATT/iSP18/GTGACTGG AGTTCAGACGTGTGCTCTTCCGATCT-3′.
3. Forward PCR primer.
   5′-AATGATACGGCGACCACCGAGATCTACACGTTC AGAGTTCTACAGTCCGACG-3′.
4. Barcoded reverse PCR primer.
   5′-CAAGCAGAAGACGGCATACGAGAT/barcode/GT GACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3′.

3 Methods

3.1 Animal Injections

Here, we use 3-month-old C57BL mice as an example. We have successfully tried this procedure on 20-month-old mice and even different species such as the naked mole-rat. To measure the average elongation rate, at least 3 time points are required. A single mouse provides one time point; however, many organs can be collected. All animal experiments followed the guidelines of the Institutional Animal Care and Use Committee (IACUC) at Brigham and Women's Hospital, Harvard Medical School.

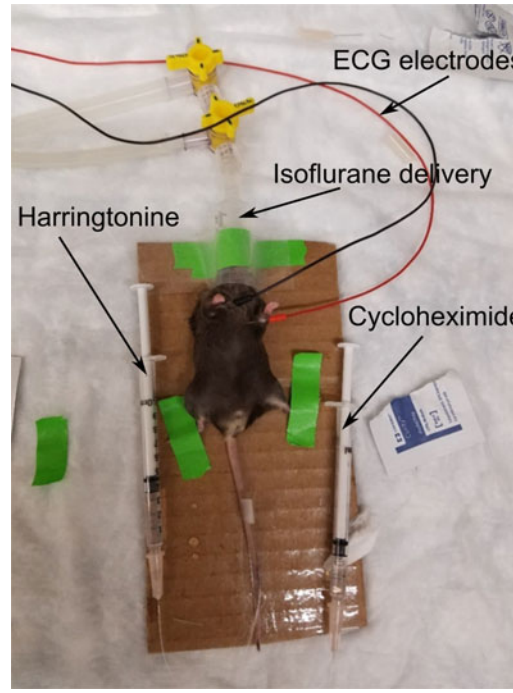

Fig. 2 Schematic of mouse injection setup

1. Sedate a mouse by placing it in a tightly closed container connected to the isoflurane vaporizer. Use 1% isoflurane mixed with oxygen as a sedative.
2. When the animal falls asleep, transfer it to a heated pad and put an isoflurane dispensing mask over the head to keep it sedated.
3. Connect ECG electrodes to monitor heartbeat rate. This is only required during the first few trials to make sure the mice used in the experiment tolerate injections with no loss in the beat rate (Fig. 2).
4. Prepare 0.3 mL insulin syringe with translation inhibitors. We recommend using 200 μL of harringtonine (5 mg/mL) and 100 μL of cycloheximide (20 mg/mL) in PBS (*see* **Note 2**).
5. Warm up the tail with a heating lamp for a minute to dilate tail veins.
6. If only a single translation inhibitor has to be delivered, use a 0.3 mL insulin syringe to deliver it via tail vein injection in 100–200 μL volume. For time series, two inhibitors must be used within a short time interval. We recommend either using a single tail vein catheter with a fork split that allows connecting two syringes or two separate catheters each connected to its own syringe since a mouse has two lateral tail veins. Make sure there are no left-over air bubbles in syringes (Fig. 2).

7. After catheters are inserted into tail veins and syringes are attached, make sure you have an open container with liquid nitrogen, timer, and surgical instruments easily accessible. This step is better done by two people. One performs injections in a timely manner and later transfers organ pieces to microtubes, and another removes organs.
8. First, inject harringtonine in 3–4 s. Start the timer simultaneously with the injection. Wait a specified time and proceed with injecting cycloheximide. We recommend selecting time points ranging from 10 to 60 s. According to our experience with over 100 mice, time points obtained within this time interval nicely follow a linear dependence between time and the ribosome run-off distance (*see* **Note 3**).
9. Wait 2 min after injecting cycloheximide. Make sure the heart is still beating normally. Sacrifice the mouse by cervical dislocation. Cut open the abdomen and a rib cage with sharp surgical scissors and harvest organs. At this point, cells are infused with cycloheximide so as much time can be taken as needed with no negative effects on a ribosome occupancy pattern. Cut collected organs in ~20 mg pieces and put them in microtubes and freeze in liquid nitrogen. Store at −80 °C.

3.2 Tissue Lysis

Preparing high-quality ribosomes from animal tissues is more challenging than from cultured cells. Tissues contain high amounts of the extracellular matrix, collagen, and lipids. Therefore, different organs require individual lysis conditions. In the protocol below, we focus on three organs, liver, kidney, and skeletal muscle, and describe several organ-specific lysis techniques.

1. Prepare fresh lysis buffers. Common components of every buffer include 20 mM Tris–HCl pH 7.5, 50 mM KCl, 50 mM NaCl, 5 mM $MgCl_2$, 5 mM $CaCl_2$, 1 mM DTT, 0.1 mg/mL cycloheximide, and protease inhibitors. The difference is in detergents used to lyse cells and solubilize lipids. We use 1% Triton X-100 for the liver and the kidney, 1% Tween-20 together with 0.25% deoxycholate for skeletal muscle, heart, brain, and testes (*see* **Note 4**).
2. Take frozen tissues from the −80 °C storage. Keep them on dry ice or liquid nitrogen. Soft tissues (liver and kidney) can be directly homogenized and lysed in a Potter-Elvehjem homogenizer filled with 1.3 mL of lysis buffer. We recommend using no more than 20 mg of the liver and no more than 40 mg of kidney for ribosome purification. Skeletal muscle requires mechanical grinding under liquid nitrogen. We had a lot of success with CryoGrinder kit which comes as a set of tall narrow ceramic mortars and pestles attachable to a handheld

motor. After 100–200 mg of skeletal muscle are grounded to fine powder, transfer it to a Potter-Elvehjem homogenizer and proceed with lysis.

3. Transfer the lysate to a new microtube and spin down in a table-top centrifuge for 2 min. Transfer cleared supernatant in a new microtube. Dilute with lysis buffer if needed to reach 1 mL volume. It is not necessary to normalize ribosome amounts for preparative purposes. However, normalization can be done to simplify the comparison between sucrose gradient profiles. The easiest and the least accurate is normalizing based on 254 or 280 nm absorption units. The accuracy is unacceptable if the ribosome content is low due to the fact that Triton X-100 also absorbs in the UV spectrum. Better approaches include specialized chemistry such as Qubit staining kits for nucleic acids and proteins.

3.3 Sucrose Gradient Fractionation

Ribosome quality is critical for downstream sequencing. In this protocol, we use a combination of RNase T1 and RNse S7 to achieve the complete digestion of polysomes while avoiding over digestion of monosomes. Other ribonucleases, such as RNase I, can also be used but require more fine-tuning for different organs and tissues.

1. Add 4 μL of RNase T1 and 1 μL RNase S7. Incubate 1 h at room temperature with gentle rotation.
2. Meanwhile, prepare sucrose gradients in Gradient Master following manufacturer's instructions.
3. Optional: add EGTA (5 mM final concentration) to inactivate RNase S7.
4. Spin down lysates in a table-top centrifuge for 1 min at max speed and transfer the supernatant in a new microtube.
5. Carefully overlay 1 mL from each sample lysate on top of the sucrose gradients. Perform ultracentrifugation in SW41 rotor for 3 h at 35,000 rpm (average RCF 151,263 $\times g$).
6. Plug the sample tube in a tube piercing device connected to UV detector. Pump 60% sucrose solution through the tube at 1 mL/min rate. Heavy sucrose solution will displace the sucrose gradient upward making it flow through UV detector. Collect the peak with monosomes to a clean microtube (1.5–2 mL).
7. Concentrate the monosome fraction with 100 kDa centrifugal filters until the final volume reaches 100 μL.
8. Add 3 volumes of Trizol LS and proceed by following the manual of the Direct-zol RNA purification kit. Store eluted RNA at −80 °C.
9. *See* **Note 5** for quality control instructions.

3.4 Footprint Extraction and Library Preparation

Ribosome profiling has a variety of ways to make a sequencing library. Here we use the adapter ligation followed by circularization.

1. Load up to 15 μg of RNA in a single well of 15% polyacrylamide TBE-Urea gel. Leave one empty well between samples to minimize cross-contamination. Use RNA oligonucleotides (25–32 nt) as markers. Run electrophoresis for 65 min at 180 V. Cut the band in the 25–32 nucleotides range. Elute RNA in 300 μL elution buffer and precipitate RNA footprints (*see* **Note 6**).
2. Set up dephosphorylation: dilute RNA in 16.75 μL with water, add 2 μL of T4 kinase A buffer, 1 μL T4 kinase, and 0.25 μL SuperaseIn. Incubate 30 min at 37 °C, precipitate.
3. Set up adapter ligation: 4 μL PEG8000, 1 μL ligation buffer, 1 μL adenylated adapter (100 ng/μL), 0.25 μL Superase-In, 1 μm T4 RNA ligase 2 truncated, bring the total volume to 10 μL with water. Incubate 3 h at 25 °C.
4. Remove excess of the adapter: add 1 μL of lambda nuclease and 0.5 μL deadenylase directly to the ligation mix. Incubate 1 h at 37 °C and inactivate for 10 min at 75 °C. No precipitation required.
5. Set up reverse transcription. Add the following components directly to the reaction mix from the previous step: 1 μL Reverse Transcription Oligo (10 pmol/ μL), 1 μL dNTP mix (10 mM). Incubate 5 min at 65 °C then place on ice. Add 4 μL of reverse transcription buffer, 2 μL DTT, 0.5 μL Superase-In, and 0.5 μL SuperScript III (all components are from the SuperScript III reverse transcription kit). Incubate 30 min at 48 °C, 1 min at 65 °C, and 5 min at 80 °C. To hydrolyze RNA, add 0.8 μL of 2 M NaOH and incubate 30 min at 98 °C, then add 0.8 μL 2 M HCL to neutralize the reaction. Precipitate.
6. Dissolve the pellet in 5 μL of water with 5 μL of TBE sample buffer. Incubate 2 min at 75 °C, cool on ice. Load in a single well of a 15% polyacrylamide TBE-Urea gel. Leave one empty well between samples to minimize cross-contamination. Run electrophoresis for 50 min at 180 V. Cut the band above the leftover reverse transcription primer. Elute DNA in 300 μL elution buffer and precipitate (*see* **Note 7**).
7. Set up circularization. Resuspend the pellet in 16.5 μL of water, add 2 μL of Circligase reaction buffer, 1 μL of 50 mM $MnCl_2$, 0.5 μL of Circligase II. Incubate 2 h at 60 °C, inactivate 10 min at 80 °C. Store the product at −20 °C for 1–2 weeks. Avoid long storage because the reaction mix will turn yellow and DNA will get degraded (most likely due to the presence of Mn). Precipitate to get rid of Mn for long-term storage.

8. Amplify the sequencing library. PCR reaction with Phusion polymerase: 1 μL PCR forward primer (10 pmol/μL), 1 μL PCR barcoded reverse primer (10 pmol/μL), 10 μL HF Buffer, 1 μL dNTP (10 mM), 1 μL DNA template from the Circligase reaction, nuclease-free water to 50 μL final volume. PCR settings: initial melting, 1 min at 98 °C, and cycling 20 s at 94 °C, 10 s at 55 °C, 15 s at 72 °C. Set up several reactions with 8–14 cycles.
9. Analyze PCR products on 8% TBE non-denaturing gel and select the one with an optimal number of cycles (strong amplicon band with no unspecific bands). Elute and precipitate DNA (*see* **Note 7**). The library is ready for sequencing.

3.5 Data Analysis

Sequencing read mapping is performed using Bowtie or STAR aligners and subsequent data analysis is performed with custom R and Perl scripts. To get started, refer to our publication for software parameters and analytical pipelines [8].

To plot time-dependent ribosome occupancy (Fig. 1), we employed the following strategy. First, we extracted RefSeq and BestRefSeq records for every protein-coding gene in the mouse genome annotation from NCBI. Among them, we identified the longest isoform for every gene, prioritized as CDS > 5′UTR > 3′UTR. Longer 5′-UTRs were trimmed to be 100 nucleotides in length. If either or UTRs were shorter than 100 nucleotides, we filled it with up to 100 based on genomic coordinates. To prepare a list of non-redundant genes, we run blast of all vs. all (blastall -p blastn -m 8 -b 500 -v 500 -e 0.001). Gene pairs that are too similar at the level of nucleotide sequence were excluded. In addition to the e-score, we enforced a requirement of the high-homology stretch being at least 50 nt long, and if it was longer, the similarity had to be at least 90% to treat these genes as homologous and redundant. This reference set ensured the unambiguous alignment of ribosome footprints. The average translation elongation rate can be calculated from the time-dependent ribosome occupancy plot by fitting a sigmoid curve and finding the inflection point. We recommend 3 replicated per time point in order to use linear regression and obtain proper p values for the elongation rate.

To estimate the elongation rate for individual genes, increase the number of time points (Fig. 3). Simple solutions such as sigmoid curve fitting are not applicable at the single gene level and require more sophisticated modeling which is out of the scope of this protocol.

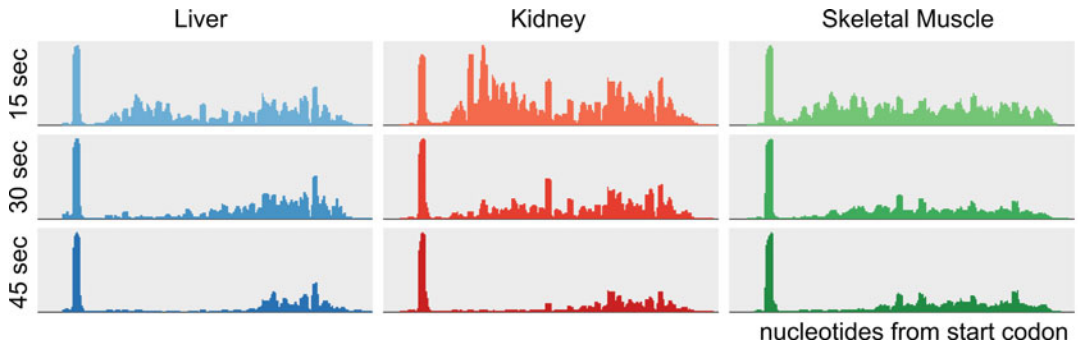

Fig. 3 Ribosome occupancy over a representative gene (*Acadm*). The more time points collected—the more accurate the translation elongation rate estimation

4 Notes

1. We strongly recommend using vaporizers equipped with flow rate control knob and not self-made contraptions with cotton balls soaked in isoflurane. The vaporizer maintains a constant flow rate and properly mixes isoflurane with oxygen which does not interfere with heart beat rate and breathing reflexes too much.
2. Unlike cycloheximide, harringtonine is not soluble in water. Dissolve it in pure DMSO to prepare 100 mg/mL stock solution, then dilute with PBS right before injecting. Harringtonine stays in-solution for about 10 min at room temperature without precipitation.
3. We are not sure why measured elongation rate declines after about 60 s. Based on our experience with over 100 mice, first 60 s display linear dependence between the elongation rate and time.
4. Deoxycholate precipitates in the presence of bivalent metal ions such as magnesium, calcium, and zinc. Tween-20 is required to stabilize it from precipitating. The protective effect does not last long; therefore deoxycholate must be added to the Tween containing lysis buffer right before use. Make sure buffers are ice-cold.
5. Prior to engaging in footprint extraction and high-throughput sequencing, make sure that the harringtonine injections work as expected. This can be verified by comparing undigested polysome profiles of injected animals (Fig. 4).
6. DNA and RNA elution from the polyacrylamide gel. Crush the gel slice with a microtube pestle and add 300 μL of the elution buffer. Incubate for at least 3 h at room temperature with active agitation.

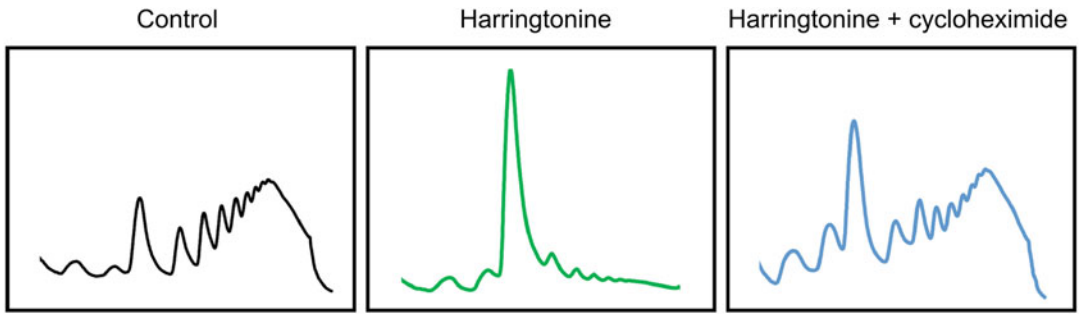

Fig. 4 Polysome profiles obtained from mouse livers with no injections (left), injected with harringtonine (center), co-injected with harringtonine and cycloheximide (right)

7. DNA and RNA precipitation. Bring the sample volume to 50 μL with water if needed. Add 1/10 volume of 3 M sodium acetate, 1/100 or 2 μL (whichever is greater) of glycogen, 2.5 volumes of 100% ethanol. Keep at −20 °C for 1 h.

References

1. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324 (5924):218–223. https://doi.org/10.1126/science.1168978
2. Ingolia NT, Lareau LF, Weissman JS (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell 147(4):789–802. https://doi.org/10.1016/j.cell.2011.10.002
3. McGlincy NJ, Ingolia NT (2017) Transcriptome-wide measurement of translation by ribosome profiling. Methods 126:112–129. https://doi.org/10.1016/j.ymeth.2017.05.028
4. Narla A, Vlachos A, Nathan DG (2011) Diamond Blackfan anemia treatment: past, present, and future. Semin Hematol 48(2):117–123. https://doi.org/10.1053/j.seminhematol.2011.01.004
5. Oliver KE, Rauscher R, Mijnders M, Wang W, Wolpert MJ, Maya J, Sabusap CM, Kesterson RA, Kirk KL, Rab A, Braakman I, Hong JS, JLt H, Ignatova Z, Sorscher EJ (2019) Slowing ribosome velocity restores folding and function of mutant CFTR. J Clin Invest. https://doi.org/10.1172/JCI124282
6. Fresno M, Jimenez A, Vazquez D (1977) Inhibition of translation in eukaryotic systems by harringtonine. Eur J Biochem 72(2):323–330
7. Lee S, Liu B, Lee S, 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(37): E2424–E2432. https://doi.org/10.1073/pnas.1207846109
8. Gerashchenko MV, Peterfi Z, Yim SH, Gladyshev VN (2021) Translation elongation rate varies among organs and decreases with age. Nucleic Acids Res 49(2):e9. https://doi.org/10.1093/nar/gkaa1103

Chapter 9

Active Ribosome Profiling with RiboLace: From Bench to Data Analysis

Massimiliano Clamer, Fabio Lauria, Toma Tebaldi, and Gabriella Viero

Abstract

Ribosome profiling is based on the deep sequencing of RNA fragments protected by ribosomes from nuclease digestion. This technique has been extensively used to study translation, with the unique ability to provide information about ribosomes positioning along transcripts at single-nucleotide resolution. Classical ribosome profiling approaches do not distinguish between fragments protected by either actively translating or inactive ribosomes. Here we describe an original method, called active ribosome profiling or RiboLace, which is based on a unique puromycin-containing molecule capable of isolating active ribosomes by means of an antibody-free and tag-free pull-down approach. This method allows reliable estimates of the translational state of any biological system, in high concordance with protein levels. RiboLace can be applied both in vitro and in vivo and generates snapshots of active ribosome footprints at single-nucleotide resolution and genome-wide level. RiboLace data are suitable for the analysis of translated genes, codon-specific translation rates, and local changes in ribosome occupancy profiles.

Key words Ribosome profiling, Translational control, Translation, Ribosome, Start codon, Codon usage, Deep sequencing, Computational analysis, RNA, Polysome

1 Introduction

The last decade has witnessed the explosion of novel methods to study translation, reviewed in [1]. The vast majority of these approaches are based on ribosome profiling (Ribo-Seq) [2], which has contributed to considerable new insights into better understanding of translation. Ribo-Seq has undoubtful discrimination power and wide applicability, but in its classical form, it cannot distinguish translationally inactive mRNAs sequestered into ribonucleoprotein particles (mRNP) and monosomes (80S), from bona fide actively translated mRNAs. In fact, mRNAs can be trapped within stalled or transiently paused ribosomes, especially in highly polarized cells as neurons [3–8]. In this scenario, Ribo-Seq does not necessarily portray the position of "authentic" mRNA footprints protected by actively translating ribosomes. This pitfall leads

Vyacheslav M. Labunskyy (ed.), *Ribosome Profiling: Methods and Protocols*, Methods in Molecular Biology, vol. 2252, https://doi.org/10.1007/978-1-0716-1150-0_9,

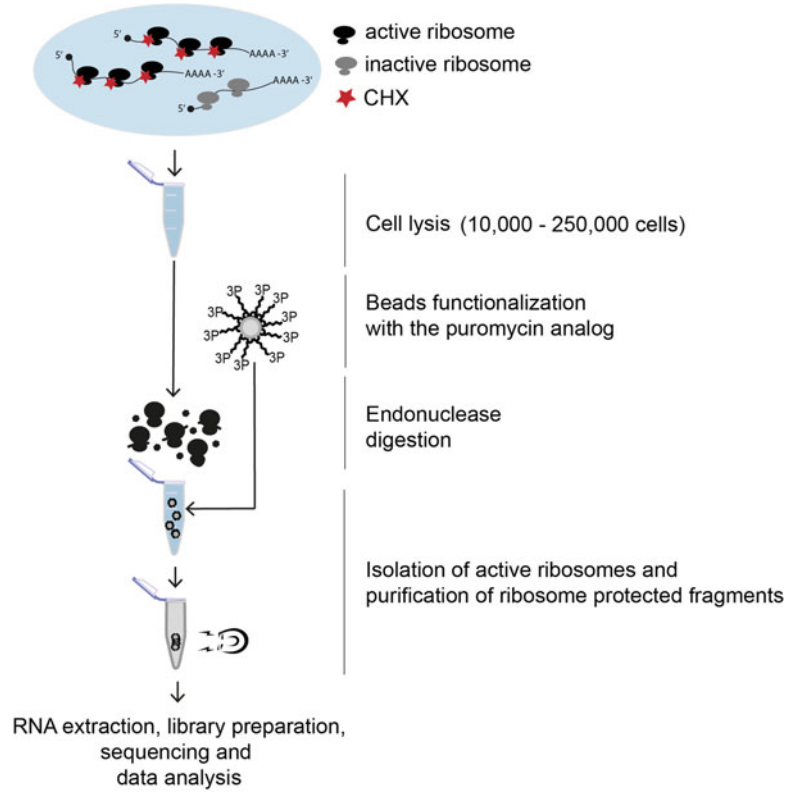

Fig. 1 Schematic of the Active-ribo-seq (or RiboLace) protocol. Cells are pretreated with cycloheximide (CHX) to immobilize ribosomes. After cell lysis, the lysate is incubated with avidin-based magnetic beads functionalized with the puromycin analog. RiboLace beads are then washed and ribosomes isolated. Finally, ribosome-protected fragments are recovered for further library preparation, sequencing, and data analysis

to possible misinterpretations of both translation occupancy profiles and translational states of the biological system under analysis. To generate insight into truly actively translated processes, Ribo-Seq can be adapted to selectively capture the occupancy signal from functionally active ribosomes.

To purify mRNA fragments protected by ribosomes which undergo active translation, we developed RiboLace, or Active-ribo-seq [9], which utilizes a puromycin analog molecule as a bait (Fig. 1). Puromycin is an aminonucleoside antibiotic which binds the ribosome [10–13] and can be incorporated into the nascent peptide chain [14, 15]. By covalently coupling puromycin to a biotin moiety via two 2,2′-ethylenedioxy-bis-ethylamine units, it is possible to isolate RNA fragments protected by puromycin-trapped translating ribosomes by simply using magnetic beads [9]. The combination of the puromycin analog and cycloheximide, which clamps ribosomes on mRNA fragments, allows to easily trap

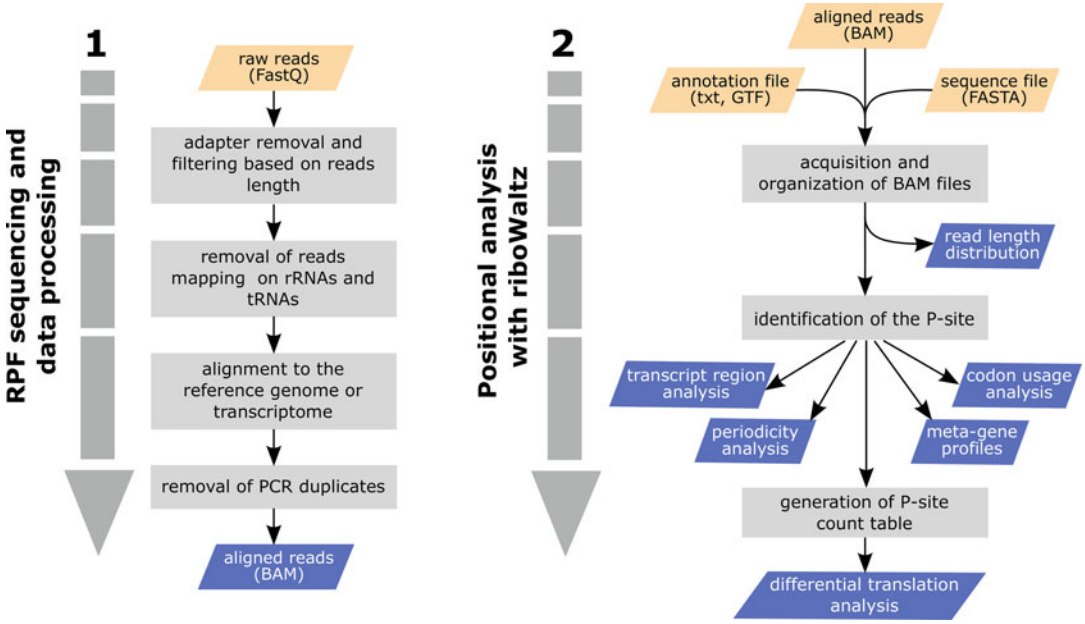

Fig. 2 Flow chart representing the basic computational steps for the analysis of active ribosome profiling data, highlighting the input requirements and the provided outputs

and purify actively translating ribosomes and their RNA protected fragments. Finally, by applying dedicated computational tools [16, 17], it is possible to selectively portray the position of bona fide active ribosomes at single-nucleotide resolution (Fig. 2). This protocol, which takes about 5 h for the isolation of active ribosomes and about 4 days for library preparation, facilitates the acquisition of reliable descriptions of active translational events in vitro and in vivo.

2 Materials

2.1 Preparation of Cell Lysates

1. Growth medium: Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS), 1% glutamine, 1% penicillin/streptomycin.
2. Phosphate-buffered saline (PBS).
3. 10 cm dishes or 6-well plates.
4. Cycloheximide: dissolve the powder in nuclease-free water to a final concentration of 10 mg/mL; keep aliquots at −20 °C.
5. Freshly prepared lysis buffer: 20 mM HEPES, 100 mM KCl, 5 mM $MgCl_2$, 20 μg/mL cycloheximide, 1% Triton X-100, 1% Na-deoxycholate, pH 7.4. All solutions must be prepared in RNase-free water. Supplement the lysis buffer with 5 U/mL DNase I and 200 U/mL of RiboLock RNase Inhibitor.
6. Nanodrop ND-1000 UV–VIS spectrophotometer.

2.2 Bead Functionalization

1. A-buffer: 50 mM NaCl, 100 mM NaOH in RNase-free water.
2. B-buffer: 2 M NaCl, 10 mM Tris–HCl, pH 7.5 in RNase-free water.
3. W-buffer: 10 mM NaCl, 10 mM $MgCl_2$, 20 mg/mL cycloheximide, 10 mM HEPES, pH 7 in RNase free-water.
4. Avidin-based magnetic beads.
5. Puromycin analog 3P (Immagina BioTechnology S.r.l., **Note 1**): dissolve in B-buffer to a concentration 10 mM. Aliquot the solution and store the aliquots at −80 °C. Avoid more than two freeze-thaw cycles.
6. Biotin-methoxypolyethylene glycol conjugate (mPEG).
7. Nanodrop ND-1000 UV–VIS spectrophotometer.
8. RNase-free water.

2.3 Endonuclease Digestion

1. 40,000 units/mL murine RNase inhibitor.
2. 100 U/μL RNase I or a convenient nuclease chosen according to [18].
3. SUPERase-In inhibitor.

2.4 Isolation of Active Ribosomes and Purification of Ribosome-Protected Fragments (RPFs)

1. W-buffer.
2. 20 mg/mL Proteinase K.
3. Sodium dodecyl sulfate (SDS): 10% solution in DNase-/RNase-free water molecular biology grade.
4. RNAse-free water.
5. Acid phenol:chloroform:isoamyl alcohol, pH 4.5.
6. 5 mg/mL glycogen coprecipitant.
7. Isopropanol for molecular biology.
8. Microcentrifuge and non-stick RNase-free microfuge tubes (0.2 mL and 1.5 mL).
9. Automatic wheel rotator.
10. Magnetic stand for 1.5 mL tubes.
11. 2 M NaCl in Rnase-free water.

2.5 RPF PAGE Purification

1. RNA control (5 μM of 25 nt and 35 nt 3′ phosphorylated RNA oligos).
2. 15% TBE-urea polyacrylamide gel.
3. 10× TBE RNase-free buffer.
4. Denaturing gel loading buffer: 2× solution of 95% formamide, 18 mM EDTA, 0.025% SDS, 0.025% xylene cyanol, and 0.025% bromophenol blue.
5. 0.5 mL RNase-free microcentrifuge tube filters, 0.22 μm.

6. Non-stick RNase-free microfuge tubes (0.2 mL and 1.5 mL).
7. SYBR Gold 10,000 × .
8. Isopropanol for molecular biology.
9. DNA marker ranging from 20 to 100 nt.
10. C-Buffer: 1 mM EDTA, 0.25% v/v SDS, 300 mM NaOAc pH 5.5.
11. 5 mg/mL glycogen coprecipitant.
12. 10 mM TRIS, pH 8.
13. Microcentrifuge and non-stick RNase-free microfuge tubes (0.2 mL and 1.5 mL).
14. Sterile 1 mL cut-tips.

2.6 T4 Polynucleotide Kinase (PNK) Dephosphorylation

1. 10 U/μL T4 PNK.
2. 10× T4 PNK buffer.
3. RNase-free water.
4. SUPERase-In inhibitor.
5. Isopropanol for molecular biology.
6. 3 M sodium acetate, pH 5.2.
7. 5 mg/mL glycogen coprecipitant.
8. Microcentrifuge and non-stick RNase-free microfuge tubes (0.2 mL and 1.5 mL).
9. 10 mM TRIS, pH 8.

2.7 Adaptor Ligation and PAGE Purification

1. 200 U/μL T4 RNA ligase 2 truncated K227Q (T4 Rnl2(tr)).
2. 10× T4 RNA ligase buffer.
3. Pre-adenylated oligonucleotide (PR primer): 5rApp/CTGTA GGCACCATCAAT/3dd.C. The initial p- indicates 5′ phosphorylation, r-ribonucleotides, and d-dexyribonucleotides at the 3′.
4. 50% PEG-8000.
5. Isopropanol for molecular biology.
6. 3 M sodium acetate, pH 5.2.
7. 5 mg/mL glycogen coprecipitant.
8. DNA marker ranging from 20 to 100 nt.
9. 15% TBE-urea polyacrylamide gel.
10. 10× TBE RNase-free buffer.
11. Denaturing gel loading buffer: 2× solution of 95% formamide, 18 mM EDTA, and 0.025% SDS, 0.025% xylene cyanol, and 0.025% bromophenol blue.
12. 0.5 mL RNase-free microcentrifuge tube filters, 0.22 μm.

13. Non-stick RNase-free microfuge tubes.
14. SYBR Gold 10,000 × .
15. C-buffer.
16. 10 mM TRIS pH 8.

2.8 Reverse Transcription and cDNA PAGE Purification

1. SuperScript III Reverse Transcriptase.
2. 5× SuperScript III buffer.
3. dNTP.
4. Reverse transcriptase oligonucleotide (RT primer): 5′-(Phos)-AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTA GATCTCGGTGGTCGC -(SpC18)-CACTCA-(SpC18)- TTC AGACGTGTGCTCTTCCGATCTATTGATGGTGCCTA CAG-3′. The designation (Phos) indicates 5′ phosphorylation and -(SpC18)- indicates a hexa-ethyleneglycol spacer.
5. C-buffer.
6. 1 N NaOH.
7. Isopropanol for molecular biology.
8. 3 M sodium acetate, pH 5.2
9. 5 mg/mL glycogen coprecipitant.
10. DNA marker ranging from 20 nt to 100 nt.
11. 15% TBE-urea polyacrylamide gel.
12. 10× TBE RNase-free buffer.
13. Denaturing gel loading buffer: 2× solution of 95% formamide, 18 mM EDTA, and 0.025% SDS, 0.025% xylene cyanol, and 0.025% bromophenol blue.
14. 0.5 mL RNase-free microcentrifuge tube filters, 0.22 μm.
15. SYBR Gold 10,000 × .
16. Microcentrifuge and non-stick RNase-free microfuge tubes (0.2 mL and 1.5 mL).
17. 10 mM Tris, pH 8.

2.9 cDNA Circularization

1. CircLigase II ssDNA ligase.
2. 10× CircLigase buffer.
3. 50 mM $MgCl_2$.
4. 5 M betaine.
5. Microcentrifuge and nonstick RNase-free microfuge tubes (0.2 mL and 1.5 mL).

2.10 PCR Amplification and PAGE Purification of the Library

1. Forward library PCR primer: 5′-AATGATACGGCGACCACC GAGATCTACAC-3′.
2. Phusion High-Fidelity PCR Master Mix.
3. Illumina Index reverse primers.

4. DNA marker (10 nt–300 nt).
5. RNase-/DNase-free water.
6. 5 mg/mL glycogen coprecipitant.
7. 6–8% TBE non-denaturing gel.
8. 10× TBE buffer.
9. DNA loading dye.
10. Microcentrifuge and nonstick RNase-free microfuge tubes (0.2 mL and 1.5 mL).
11. Agilent 2100 Bioanalyzer.
12. High-sensitivity DNA chip.

3 Methods

3.1 Preparation of Cell Lysates for Active-ribo-seq (Estimated Time, 25 min)

In this section the cell lysate is prepared (Fig. 1). It is important to block ribosomes on mRNAs by cycloheximide treatment. This step reduces the dissociation of ribosomes when the puromycin analog is added to the lysate.

1. Grow cells in growth medium, and keep in culture at 5% CO_2, 95% humidity, and 37 °C. Split the cells 3 days before the experiment, and seed them into a 6-well plate or 10 cm dish (*see* **Note 2** for details about cells number). In the case of tissue analysis, please do refer to **Note 3**.
2. When cells reach 80% confluence, treat them with cycloheximide (10 μg/mL final concentration), and keep in incubation at 37 °C for 5 min.
3. Place the plate on ice, and wash three times with cold PBS containing cycloheximide (10 μg/mL).
4. Remove completely the residual PBS with a pipette.
5. Add the freshly prepared lysis buffer as follows: 300 μL/10 cm dish or 30 μL/well of the 6-well plate and scrape vigorously.
6. Collect the cell lysate in a 1.5 mL microcentrifuge tube, vortex for 10 s, and keep on ice for 10 min.
7. Centrifuge the lysate at 20,000 × *g*, 4 °C for 5 min to remove nuclei and mitochondria.
8. Transfer the supernatant to a new tube and keep it on ice for 20 min.
9. Using a convenient UV–Vis spectrophotometer, determine the absorbance of the cell lysate at 260 nm using the lysis buffer as blank.

3.2 Bead Functionalization (Estimated Time, 90 min)

In this section the magnetic beads are functionalized with the puromycin analog to purify actively translating ribosomes, which will be prepared in Subheading 3.3.

1. Vortex the tube containing the avidin-based magnetic beads for 30 s.
2. Transfer 90 μL of bead suspension in a 1.5 mL tube, and keep them at room temperature (RT) for 10 min.
3. Place the tube on the magnet to separate the beads, and remove the beads' medium.
4. Remove the tube from the magnet, and wash the beads for 5 min with an equal volume of A-buffer; then remove the supernatant.
5. Wash the beads with 900 μL of nuclease-free water, place the tube on the magnet, and remove the supernatant. If the beads remain attached to the tube walls, add Triton X-100 (dissolved in RNase-free water) at a final concentration of 0.1%.
6. Wash the beads twice with 90 μL of B-buffer for 3 min.
7. Place the tube on the magnet for at least 1 min and remove the supernatant.
8. Dissolve the beads in 30 μL of 1 mM puromycin analog 3P. Collect 1 μL of 3P probe for "security point" (*see* **Note 4**).
9. Incubate for 1 h at RT in a constant agitation. Do not allow the beads to sediment.
10. After the incubation, place the tube on a magnet, and collect 3 μL of the supernatant (unbound probe) for "security point" (*see* **Note 4**).
11. Passivate the beads using 3 μL mPEG; mix in a shaker at RT for 15 min. Do not allow the beads to precipitate.
12. Place the tube on a magnet for 2–3 min, discard the supernatant, and wash with 500 μL of nuclease-free water.
13. Wash the beads two times with 500 μL of W-buffer; dissolve them in 200 μL of W buffer. Do not let the beads dry. Remove the buffer only before you are ready to add the cell lysate.

3.3 Endonuclease Digestion (Estimated Time, 90 min)

The cell lysate is treated with an endonuclease to digest the mRNA regions which are not protected by ribosomes. Substantial changes in the RNase activity during nuclease digestion can jeopardize the experiment.

1. Start with a total volume of lysate corresponding to 0.03–0.3 au (260 nm) diluted in W-buffer in a final volume of 150 μL. For example, if A260 = 10 au/mL = 0.01 au/μL = 30 μL needed.
2. Add 0.3 μL of murine RNase inhibitor.

3. In a 0.2 mL vial, pipet 1.5 μL of RNase I, and add 98.5 μL W-buffer to 1 U/mL.
4. Digest the sample in a 1.5 mL tube for 45 min at 25 °C with the diluted nuclease prepared in (B) using a volume according to the formula: volume of RNase I μL = A.U. (0.3 − 0.03) × 5. Trash the remaining diluted RNase I.
5. Stop digestion with 0.5 μL SUPERase-In inhibitor for 10 min on ice.

3.4 Isolation of Active Ribosomes and Purification of RPFs (Estimated Time, 3 h. Overnight Optional)

Active ribosomes are separated from the cell lysate by affinity purification with the beads prepared as described in Subheading 3.2 as follows:

1. Remove the wash buffer from the beads using a magnet.
2. Remove the tube from the magnet, and add the digested cell lysate to the beads.
3. Incubate the sample on a wheel in slow motion (3 rpm) at 4 °C for 70 min.
4. Remove the tube from the wheel. Pull down the beads by gentle hand shaking.
5. Place the tube on a magnet pre-chilled on ice and keep working at 4 °C.
6. Separate the beads with the magnet. Do not remove the tube from the magnet, and remove the remaining lysate.
7. Keeping the tube on the magnet, carefully wash the beads twice with 500 μL of W-buffer. Do not touch the beads with the pipet tip.
8. Remove the beads from the magnet, and resuspend them in W-buffer to a final volume of 200 μL. Transfer the suspension to a new nuclease-free 1.5 mL tube.
9. Add 20 μL of 10% SDS (final 1%) and 5 μL of proteinase K, and incubate the sample at 37 °C in a water bath for 75 min.
10. Add an equal volume of acid phenol:chloroform:isoamyl alcohol.
11. Vortex and centrifuge at 14,000 × *g* for 5 min at 4 °C.
12. In the case of no phase separation, add 20 μL of 2 M NaCl, mix, and repeat the centrifugation.
13. Keep the aqueous phase and transfer it into a new 1.5 mL tube.
14. Add 500 μL isopropanol and 2 μL of coprecipitant.
15. Mix and incubate at RT for 3 min, and then store at −80 °C for at least 2 h (fast procedure) or overnight (safer procedure, recommended with lysate input <0.1 au).

16. Pellet the RNA by centrifugation at 20,000 × *g* for 30 min at 4 °C.
17. Resuspend the pellet in 5 μL of RNase-free water.

3.5 RPF PAGE Purification

The RPFs are separated from rRNA and other RNA fragments using a denaturing gel and are collected via size selection. As in classical ribosome profiling, the typical RPF size to be isolated in mammalian cells is around 30 nt.

1. Pre-run the 15% TBE-urea polyacrylamide urea gel at 200 V for 30 min in 1× TBE buffer prepared with nuclease-free water. Carefully clean the wells with a syringe to remove UREA residuals.
2. Add 5 μL of denaturing gel loading buffer to 5 μL of RPFs.
3. In parallel, prepare the RNA control sample by mixing 2 μL RNA control with 3 μL of nuclease-free water and 5 μL of denaturing gel loading buffer.
4. In parallel, prepare the marker by mixing 1 μL of marker solution with 4 μL nuclease-free water and 5 μL of denaturing gel loading buffer.
5. Denature all samples for 90 s at 80 °C. Immediately place the tubes on ice.
6. Carefully clean the wells with a syringe to remove urea residuals.
7. Load the samples, RNA control, and marker on denaturing 15% TBE-urea polyacrylamide gel, and run the gel at 200 V until the bromophenol blue band reaches the bottom of the gel.
8. Stain the gel for 3 min with 1× SYBR Gold in 1× TBE running buffer on a gentle shaker.
9. Visualize the RNA using a trans-illuminator.
10. Using a scalpel, size select the RPF between 25-nt and 35-nt according to the RNA control and the marker positions. As control, excise the "RNA control" oligo bands. You can use the "RNA control" as a reference sample for the next steps of library preparation.
11. Place each gel slice in a clean non-stick RNase-free microfuge tube.
12. Add 400 μL of C-buffer and smash the gel (*see* **Note 5**); close the vial and incubate the tubes for 1 h at −80 °C.
13. Thaw the sample at RT, and then place the sample on a wheel in slow motion (3 rpm), at RT overnight.

14. Using a 1 mL cut-tip, add the gel slurry to the filter, and spin at 650 × *g* for 3 min at 4 °C to remove the gel debris. Transfer the eluted solution to a new tube.
15. Add 700 μL of isopropanol and 1.5 μL coprecipitant to the eluted sample.
16. Store at −80 °C for 2 h (fast procedure) or overnight (safe procedure).
17. Thaw the samples on ice, and pellet the RNA by centrifugation at 20,000 × *g* for 30 min at 4 °C.
18. Resuspend the pellet in 10 μL of 10 mM Tris pH 8.

3.6 T4 Polynucleotide Kinase Dephosphorylation

In this section, the RPFs are treated with T4 polynucleotide kinase (T4 PNK) in order to heal the 2′–3′ cyclic phosphate that results from RNase I cleavage, and the fragments are prepared for the ligation to the DNA linker.

1. Transfer the reaction in a 0.2 mL RNase-free PCR tube suitable for your thermocycler.
2. Set up a 50 μL reaction mixture by adding to the sample 5 μL T4 PNK buffer, 1 μL SUPERase-In (20 U/μL), 1 μL T4 PNK (10 U/μL), and 33 μL nuclease-free water.
3. Incubate the reactions in a thermocycler for 60 min at 37 °C and then for 10 min at 70 °C, and then cool the samples at RT.
4. Add 39 μL of RNase-free water, 10 μL of 3 M sodium acetate, 150 μL of isopropanol, and 1 μL coprecipitant.
5. Transfer the solution to a 1.5 mL tube.
6. Store at −80 °C for 2 h (fast procedure) or overnight (safe procedure, recommended for low-input samples).

3.7 Adaptor Ligation and PAGE Purification

In this section, the dephosphorylated fragments are ligated to a DNA linker with T4 Rnl2(tr), and the ligated product is purified by size selection using PAGE.

1. Thaw the samples on ice, and pellet the RNA by centrifugation at 20,000 × *g* for 30 min at 4 °C.
2. Resuspend the pellet in 7 μL 10 mM Tris pH 8, and transfer it to a 0.2 mL RNase-free PCR tube.
3. Add 1 μL PR-primer, denature the sample for 90 s at 80 °C in a thermocycler, and then cool the reactions at RT.
4. Set up a 20 μL reaction mixture by adding 2 μL T4 RNA ligase buffer 10×, 8 μL PEG-8000, 1 μL SUPERase-In, and 1 μL T4 Rnl2(tr) (200 U/μL).
5. Incubate for 3 h at 22 °C in a thermocycler.
6. Transfer the reaction to a new 1.5 mL tube.

7. Add 338 µL of nuclease-free water, 40 µL sodium acetate, 500 µL of isopropanol, and 1.5 µL coprecipitant.
8. Store at −80 °C for 2 h (fast procedure) or overnight (safe procedure, recommended for low-input samples).
9. Pre-run the gel at 200 V for 30 min in 1× TBE prepared with RNase-free water. Clean very carefully the gel wells with a syringe to remove urea residuals, and repeat before loading the samples.
10. Pellet the RNA by centrifugation 20,000 × *g* for 30 min at 4 °C.
11. Resuspend the pellet in 5 µL 10 mM Tris pH 8, and add 5 µL of denaturing gel loading buffer.
12. Denature all samples for 90 s at 80 °C. Immediately place the tubes on ice.
13. Load the samples on 15% TBE-urea polyacrylamide gel.
14. Load 1 µL of PR and marker (20–100 nt) as a reference.
15. Run the gel at 200 V until the bromophenol blue band reaches the bottom of the gel.
16. Size select the fragments at ~55 nt according to the marker and PR-primer positions. Proceed only if you see the ligation product. The efficiency depends on (i) the starting material, (ii) on the efficiency of previous purification steps, and on (iii) the translational state of cells.
17. Place each gel slice in a clean 1.5 mL non-stick RNase-free microfuge tube.
18. Add 400 µL of C-buffer and smash the gel (*see* **Note 5**); close the vial and incubate the tubes for 1 h at −80 °C.
19. Thaw the sample at RT, and then place the samples on a wheel in slow motion (3 rpm), at RT overnight.
20. With a 1 mL cut-tip, add the gel slurry to the filter, and spin at 650 × *g* for 3 min at 4 °C to remove the gel debris. Transfer the eluted solution to a new tube.
21. Add 700 µL of isopropanol and 1.5 µL coprecipitant to the eluted sample.
22. Store at −80 °C for 2 h (fast procedure) or overnight (safe procedure).
23. Thaw the samples on ice, and pellet the RNA by centrifugation at 20,000 × *g* for 30 min at 4 °C.
24. Resuspend the pellet in 10 µL 10 mM Tris pH 8, and transfer it to a new 0.2 mL RNase-free PCR tube.

3.8 Reverse Transcription and cDNA PAGE Purification

In this section, the fragments are retro-transcribed using a dedicated oligonucleotide which is complementary to the previously ligated PR primer. The cDNA is then purified by size selection after PAGE.

1. Thaw the sample on ice.
2. Add 2 μL of RT primer, denature the sample for 2 min at 80 °C in the thermocycler, and then cool the reaction on ice.
3. Set up a 20 μL reaction mixture by adding 4 μL 5× SuperScript III buffer, 1 μL 10 mM dNTPs, 1 μL 0.1 M DTT, 1 μL SUPERase-In (20 U/μL), and 1 μL SuperScript III (200 U/μL).
4. Incubate for 30 min at 55 °C in a thermocycler.
5. Add 2.2 μL of 1 N NaOH, heat the sample at 70 °C for 20 min, and then cool it at 4 °C.
6. Transfer the reaction to a new 1.5 mL tube.
7. Add 156 μL of RNase-free water, 20 μL sodium acetate, 300 μL of isopropanol, and 2 μL coprecipitant.
8. Store at −80 °C for 2 h (fast procedure) or overnight (safe procedure, recommended for low-input samples).
9. Pre-run the gel at 200 V for 30 min in 1× TBE prepared with nuclease-free water. Carefully clean the gel wells with a syringe to remove urea residuals, and repeat before loading the samples.
10. Thaw the sample on ice, then pellet the cDNA by centrifugation at 20,000 × *g* for 30 min at 4 °C, and air-dry the pellet for 10 min.
11. Solubilize the pellet in 5 μL 10 mM Tris pH 8; add 5 μL of denaturing gel loading buffer.
12. Denature all samples for 90 s at 80 °C. Immediately place the tubes on ice.
13. Load the samples on 15% TBE-urea polyacrylamide gel.
14. Load 1 μL of RT primer and of the marker prepared as in Subheading 3.5, **steps 3** and **4**.
15. Run the gel at 200 V until the bromophenol blue band reaches the bottom of the gel.
16. Size select the fragments at ~130 nt according to the RT primer and marker's band positions.
17. Place each gel slice in a clean 1.5 mL non-stick RNase-free microfuge tube.
18. Add 400 μL of C-buffer and smash the gel (*see* **Note 5**); close the vial and incubate the tubes for 1 h at −80 °C.

19. Thaw them at RT, and then place the samples on a wheel in slow motion (3 rpm) for 8–10 h at 37 °C.
20. With a 1 mL cut-tip, add the gel slurry to microcentrifuge filter, and spin at 650 × *g* for 3 min, at 4 °C to remove the gel debris. Transfer the eluted solution to a new tube.
21. Add 700 μL of isopropanol and 1.5 μL coprecipitant to the eluted sample.
22. Store at −80 °C for 2 h (fast procedure) or overnight (safe procedure).
23. Thaw the samples on ice, and pellet the cDNA by centrifugation (20,000 × *g*) for 30 min at 4 °C.
24. Solubilize the pellet in 12 μL 10 mM Tris pH 8, and transfer it to a new 0.2 mL RNase-free PCR tube.

3.9 cDNA Circularization

In this section the cDNA is circularized as follows:

1. Thaw the sample on ice.
2. Set up a 20 μL reaction mixture by adding to the cDNA 2 μL 10× CircLigase buffer, 4 μL 5 M betaine, 1 μL 50 mM $MnCl_2$, and 1 μL CircLigase II (100 U/μL).
3. Incubate for 2 h at 60 °C, followed by 10 min at 80 °C in a thermocycler, and then cool the samples at 4 °C.
4. Store the samples at −20 °C, or proceed directly with the PCR amplification.

3.10 PCR Library Amplification and Library PAGE Purification

In this section, each RPF inserted in the circular cDNA is amplified by PCR. For most samples, use 8 μl of the circularized cDNA and 10 PCR cycles that are typically enough to produce sufficient amount of library. For low-input samples, use 10 μl of the circularized cDNA from **step 7**, and increase the number of PCR cycles to 12–14 cycles. Avoid the use of too much cDNA or too many PCR cycles that can produce an "overamplification" of the library and adapter dimer-derived products.

1. Amplify the RPF in the circular cDNA template by setting up a 50 μL PCR reaction mixture containing 25 μL 2× Phusion Master Mix (2 U/μL), 8–10 μL circularized cDNA template, 2 μL Forward library PCR primer (10 mM), 2 μL Illumina Index reverse primers (10 mM), 0.5 μL $MnCl_2$ (50 mM), and nuclease-free water to bring the volume to 50 μL. The list of primers is in **Note 6**.
2. Set up the PCR reaction conditions as follows: 98 °C, 30 s; 98 °C, 10 s; 65 °C, 10 s; 72 °C, 5 s (10–14 cycles); 4 °C hold.
3. The remaining cDNA (10–12 μL) can be stored at −20 °C and can be used for other reactions.

4. Mix 10 μL PCR product and 3 μL DNA loading dye, and load the sample on a non-denaturing 6–8% TBE non-denaturing gel.
5. Run the gel at 200 V until the blue dye reaches the bottom of the gel.
6. Using a scalpel, size select the library at ~176 nt, according to the DNA marker.
7. Place each gel slice in a clean 1.5 mL non-stick RNase-free microfuge tube.
8. Add 400 μL of C-buffer and smash the gel; close the vial and incubate the tubes for 1 h at −80 °C.
9. Thaw them at RT, and then place the samples on a wheel at RT in slow motion (3 rpm) overnight.
10. With a 1 mL cut-tip, add the gel slurry to the microcentrifuge filter, and spin at 650 × *g* for 3 min at 4 °C to remove the gel debris. Transfer the eluted solution to a new tube.
11. Add 700 μL of isopropanol and 1.5 μL coprecipitant to the eluted sample.
12. Store at −80 °C overnight.
13. Thaw the samples on ice, and pellet the library by centrifugation at 20,000 × *g* for 30 min at 4 °C.
14. Resuspend the pellet in 11 μL 10 mM Tris pH 8, and transfer it to a new 0.2 mL RNase-free PCR tube.
15. Use 1 μL of the library of each sample for quality check using Bioanalyzer high-sensitivity DNA chip.
16. The libraries are ready for sequencing and can be stored indefinitely at −20 °C.

3.11 RPF Sequencing and Data Processing

Multiple computational tools have been developed for the analysis of RNA-Seq and ribosome profiling data, recently reviewed in [17]. In the following sections, we briefly describe the main steps for data analysis, and we provide, as example, the pipeline for processing the data using the riboWaltz R package [16] (Fig. 2).

Active ribosome profiling libraries are usually single-end; 50 nucleotide reads are sufficient for covering the whole ribosome-protected fragment (RPF), due to its short length (<35 nucleotides). For robust downstream analyses, at least 60–70 millions of raw reads are usually required. However, this range varies accordingly to the expected proportion of reads to be removed. In fact, frequent contaminations are caused by adapter dimers (>10%) or rRNAs (often highly dominant in the samples, from 30 to 70% of the initial number) and tRNAs (5–10%). PCR duplicates should also be considered and collapsed to obtain bona fide ribosome-protected fragments.

Raw FASTQ files should be initially processed as follows:

1. Remove adapter and linker sequences (e.g., with Cutadapt https://github.com/marcelm/cutadapt). Discard reads shorter than 20 nucleotides after adapter removal. This step also discards reads associated to adapter dimers.
2. Optionally, remove low-quality nucleotides by trimming the first base from the 5′ end of the remaining reads.
3. Remove the reads mapping on rRNAs and tRNAs, i.e., the reads that are not associated to ribosome-protected fragments. To do that, the reads can be aligned against the rRNA and tRNA sequences of the organism of interest. Keep only unmapped reads.
4. Map the remaining reads to a reference genome or transcriptome with a splice aware alignment tool (Tophat2, HISAT2, STAR). For example, use STAR (https://github.com/alexdobin/STAR) with the --quantMode TranscriptomeSAM option to map reads on chromosome sequences, and get output BAM files in transcript coordinates.
5. To avoid potential PCR duplicates, collapse the reads aligning to the very same region (e.g., with Picard tools MarkDuplicates on BAM files, https://broadinstitute.github.io/picard/).

3.12 Positional Analysis with riboWaltz

This section describes the pipeline and the main functions included in riboWaltz [16], an R package that integrates quality controls of the ribosome profiling data, P-site identification for improved interpretation of positional information, and a variety of graphical representations. If not otherwise specified, italic words in the following section identify functions included in riboWaltz (https://github.com/LabTranslationalArchitectomics/riboWaltz).

1. Acquire one or multiple BAM files by running *bamtolist* (*see* **Note** 7). This function handles and combines BAM files into a data structure that reports basic information about the reads such as length, leftmost and rightmost position with respect to the first nucleotide of the annotated transcript sequence.
2. Perform preliminary analyses and quality controls on the data:
   (a) *rlength_distr* plots the distribution of read lengths for a specified sample. It can be exploited to identify multiple populations of read length, associated to different ribosome conformations, and explore the contribution of each length to the determination of the P-site. Use *length_filter* to select the length of the reads of interest according to the resulting distribution and the aim of the experiment.
   (b) *rends_heat* displays the abundance of the 5′ and 3′ extremity of reads mapping on and around the start and the stop

codon of annotated CDSs, stratified by their length. This plot is particularly useful to identify the best read extremity to compute optimal P-site offsets (*see* **step 3**).

3. Compute the P-site offsets for all samples separately applying the function *psite* (*see* **Note 8**). After the identification of the P-site offsets, run *psite_info* to update the data structure obtained in **step 1**.
4. Verify the expected enrichment of ribosome signal in the CDS (run *region_psite*) and the trinucleotide periodicity of ribosomes along the coding regions (run *frame_psite* or *frame_psite_length*).
5. Display the trinucleotide periodicity along coding sequences using *metaprofile_psite*. This function generates metaprofiles (the merge of single, transcript-specific profiles at single-nucleotide resolution) based on P-site mapping around the start and the stop codon of annotated CDSs.
6. Run *codon_usage_psite* to investigate the empirical codon usage (i.e., which of the 64 codons display higher or lower ribosome density) and examine alterations of ribosome translocation at specific codons by comparing multiple samples.
7. Compute the number of in-frame P-site mapping on annotated coding sequences by *cds_coverage*. Such data can be used as starting point for downstream quantitative analyses (e.g., differential analyses) based on ribosome-protected fragments.

P-site count tables can be used as input to perform transcriptome-wide differential translation analysis with general RNA-Seq tools such as the Bioconductor packages edgeR [19] or DESeq2 [20]. Alternatively, if control RNA-Seq experiments were performed in parallel, ribosome profiling-specific tools such as Xtail [21] or anota2seq (https://bioconductor.org/packages/release/bioc/html/anota2seq.html) can be used to detect differentially translated genes.

4 Notes

1. The puromycin analog 3P is from Immagina BioTechnology S.r.l. The source is unique.
2. The minimum number of cells for optimal results depends on the cell type, cell confluency, and "translational state" of the biological model under consideration. We recommend not to overgrow the cells to avoid translational inhibition due to over-confluency. Optimal results are obtained with 70–80% cell confluence (6–ten million in a 10 cm dish, 0.8 million cells in a 35 mm dish). Approximately 100,000–250,000 MCF7 cells,

HEK-293 T or HeLa cells, and 10,000 mouse embryonic stem cells are required. The final amount of cells and total RNA needed is usually 20–40× less than standard ribosome profiling protocols.

3. For the lysis of fresh frozen tissues, use this protocol:
   (a) Pulverize the tissue under liquid nitrogen with mortar and pestle.
   (b) Recover the powder in a 1.5 mL tube.
   (c) Resuspend with 800 μL of tissue powder in lysis buffer: 20 mM HEPES, pH 7.4, 100 mM KCl, 5 mM $MgCl_2$, 100 μg/mL cycloheximide, 2% Triton X-100 (stock 10% in nuclease-free water), 1% Na deoxycholate, 200 U/mL RNase inhibitor, and 5 U/mL DNase.
   (d) Centrifuge at speed 10,000 × *g* for 2 min at 4 °C to remove tissue and membrane debris, and collect the supernatant.
   (e) Centrifuge again the supernatant twice for 10 min at 20,000 × *g*, 4 °C.
   (f) Collect the supernatant. Keep on ice for 20 min.
4. Security point: check the efficiency of the bead functionalization. Compare the absorbance of the unbound probe to the 1 mM RiboLace smart probe at Abs = 270 nm using Nanodrop ND-1000. The typical binding efficiency is ~60–80%.
5. To smash the gel, there are two options:
   (a) Use a disposable polypropylene RNase-, DNase-, pyrogen-free pellet pestle. Smash the gel slice for 1 min on ice without losing material on the pestle.
   (b) Pierce the bottom of an 0.5 mL RNase-free microfuge tube with a sterile 21-G needle, and cut off the cap. Nest the pierced small tube inside a 1.5 mL RNase-free microfuge tube, and place the gel slice into the inner tube. Spin the tube for 1 min at full speed in a tabletop microcentrifuge. Smash and force the gel slice through the needle hole. Transfer any remaining gel debris and discard the pierced tube.
6. List of Illumina primers for library generation.

   Index 1 PCR Primer

   5′ AAGCAGAAGACGGCATACGAGATCGTGATGT GACTGGAGTTCAGACGTGTGCTCTTCCGATCT 3′.

   Index 2 PCR Primer

   5′ CAAGCAGAAGACGGCATACGAGATACATCGGT GACTGGAGTTCAGACGTGTGCTCTTCCGATCT 3′.

   Index 3 PCR Primer

5′ AAGCAGAAGACGGCATACGAGATGCCTAAGT GACTGGAGTTCAGACGTGTGCTCTTCCGATCT 3′.

Index4 PCR Primer

5′ CAAGCAGAAGACGGCATACGAGATTGGTCAGT GACTGGAGTTCAGACGTGTGCTCTTCCGATCT 3′.

Index 5 PCR Primer

5′ CAAGCAGAAGACGGCATACGAGATCACTGTGT GACTGGAGTTCAGACGTGTGCTCTTCCGATCT3′.

Index 6 PCR Primer

5′ CAAGCAGAAGACGGCATACGAGATATTGGCGT GACTGGAGTTCAGACGTGTGCTCTTCCGATCT 3′.

Index 7 PCR Primer

5′ CAAGCAGAAGACGGCATACGAGATGATCTGGT GACTGGAGTTCAGACGTGTGCTCTTCCGATCT 3′.

Index 8 PCR Primer

5′ CAAGCAGAAGACGGCATACGAGATTCAAGTGT GACTGGAGTTCAGACGTGTGCTCTTCCGATCT 3′.

Index 9 PCR Primer

5′ CAAGCAGAAGACGGCATACGAGATCTGATCGT GACTGGAGTTCAGACGTGTGCTCTTCCGATCT 3′.

Index 10 PCR Primer

5′ CAAGCAGAAGACGGCATACGAGATAAGCTAGT GACTGGAGTTCAGACGTGTGCTCTTCCGATCT 3′.

Index 11 PCR Primer

5′ CAAGCAGAAGACGGCATACGAGATGTAGCCGT GACTGGAGTTCAGACGTGTGCTCTTCCGATCT 3′.

Index 12 PCR Primer

5′ CAAGCAGAAGACGGCATACGAGATTACAAGGT GACTGGAGTTCAGACGTGTGCTCTTCCGATCT 3′

7. If the BAM files are obtained after alignment against the transcriptome and if the sequencing library is stranded, the read mapping on the negative strand should not be present, and, if any, they are automatically removed.
8. riboWaltz identifies the optimal localization of ribosomes by computing with high precision the ribosome P-site offset for each read length, defined as the distance between the extremities of a read and the first nucleotide of the P-site itself. riboWaltz automatically recognizes the best read extremity to compute the optimal P-site offsets. By default, it is computed by starting from read mapping on the start codon of any annotated coding sequences. However, it is recommended to examine the results and, if needed, manually set the best extremity and the reference codon by using the *extremity* and *start* parameter, respectively.

References

1. Iwasaki S, Ingolia NT (2017) The growing toolbox for protein synthesis studies. Trends Biochem Sci 42(8):612–624
2. Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324:218–223
3. Darnell JC et al (2011) FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. Cell 146:247–261
4. Chapman RE, Walter P (1997) Translational attenuation mediated by an mRNA intron. Curr Biol 7:850–859
5. Verrier SB, Jean-Jean O (2000) Complementarity between the mRNA 5′ untranslated region and 18S ribosomal RNA can inhibit translation. RNA 6:584–597
6. Sivan G, Kedersha N, Elroy-Stein O (2007) Ribosomal slowdown mediates translational arrest during cellular division. Mol Cell Biol 27:6639–6646
7. Higashi S et al (2013) TDP-43 associates with stalled ribosomes and contributes to cell survival during cellular stress. J Neurochem 126:288–300
8. Graber TE et al (2013) Reactivation of stalled polyribosomes in synaptic plasticity. Proc Natl Acad Sci U S A 110:16205–16210
9. Clamer M, Tebaldi T, Lauria F, Bernabò P, Gómez-Biagi RF, Marchioretto M, Kandala DT, Minati L, Perenthaler E, Gubert D, Pasquardini L, Guella G, Groen EJN, Gillingwater TH, Quattrone A, Viero G (2018) Active ribosome profiling with RiboLace 2018. Cell Rep 25(4):1097–1108.e5
10. Pestka S, Rosenfeld H, Harris R, Hintikka H (1972) Studies on transfer ribonucleic acid-ribosome complexes. XXI. Effect of antibiotics on peptidyl-puromycin synthesis by mammalian polyribosomes. J Biol Chem 247:6895–6900
11. Odom OW, Picking WD, Hardesty B (1990) Movement of tRNA but not the nascent peptide during peptide bond formation on ribosomes. Biochemistry 29:10734–10744
12. Kukhanova M et al (1979) The donor site of the peptidyltransferase center of ribosomes. FEBS Lett 102:198–203
13. Schmeing TM et al (2002) A pre-translocational intermediate in protein synthesis observed in crystals of enzymatically active 50S subunits. Nat Struct Biol 9:225–230
14. Yarmolinsky MB, Haba GL (1959) Inhibition by puromycin of amino acid incorporation into protein. Proc Natl Acad Sci U S A 45:1721–1729
15. Welch M, Chastang J, Yarus M (1995) An inhibitor of ribosomal peptidyl transferase using the transition-state analogy. Biochemistry 34:385–390
16. Lauria F, Tebaldi T, Bernabò P, Groen EJ, Gillingwater TH, Viero G (2018) riboWaltz: optimization of ribosome P-site positioning in ribosome profiling data. PLoS Comput Biol 14 (8):e1006169
17. Kiniry SJ, Michel AM, Baranov PV (2019) Computational methods for ribosome profiling data analysis. Wiley Interdiscip Rev RNA: e1577. Advance online publication. https://doi.org/10.1002/wrna.1577
18. Gerashchenko MV, Gladyshev VN (2017) Ribonuclease selection for ribosome profiling. Nucleic Acids Res 45(2):e6
19. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26 (1):139–140
20. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15(12):550
21. Xiao Z, Zou Q, Liu Y, Yang X (2016) Genome-wide assessment of differential translations with ribosome profiling data. Nat Commun 7 (1):1–11

Chapter 10

Poly-A Tailing and Adaptor Ligation Methods for Ribo-Seq Library Construction

Jingfan Qiu and Shu-Bing Qian

Abstract

Ribosome profiling is a powerful technique that enables researchers to monitor translational events across the transcriptome. It provides a snapshot of ribosome positions and density across the transcriptome at a sub-codon resolution. Here we describe the whole procedure of profiling ribosome footprints in mammalian cells. Two methods for Ribo-seq library construction are introduced, and their advantages and disadvantages are compared. There is a room for further improvement of Ribo-seq in terms of the amount of starting material, the duration of library construction, and the resolution of sequencing results.

Key words Ribo-seq, Translation, mRNA, Codon, Deep sequencing

1 Introduction

Ribosome profiling (Ribo-seq) is a valuable approach to measure translational activities at particular moments in a genome-wide and quantitative manner [1, 2]. Ribosome profiling reveals individual mRNA molecules engaged with translating ribosomes during initiation, elongation, and termination. Relying on high-throughput sequencing, Ribo-seq captures the entire set of ribosome-protected mRNA fragments (RPFs) generated by nuclease digestion [3]. This technique has been used to predict thousands of open reading frames (ORFs) in eukaryotic cells, including cryptic overlapping ORFs and noncanonical (non-AUG) translation initiation [4–6].

However, the broad application of Ribo-seq has been slowed by the complexity, expense, and the lengthiness of the protocol. Additionally, concerns have swirled around the interpretation of Ribo-seq results as details of sample preparation may introduce bias and artifacts [7]. During library preparation, for instance, the efficiency of circularization or linker ligation could be influenced by the 5′ end nucleotide identity of RPFs. As a result, technically inflated or depleted RPFs could alter the overall pattern of ribosome footprints. Additionally, pre-treatment with the translation

Vyacheslav M. Labunskyy (ed.), *Ribosome Profiling: Methods and Protocols*, Methods in Molecular Biology, vol. 2252, https://doi.org/10.1007/978-1-0716-1150-0_10,

inhibitor cycloheximide (CHX) has been shown to skew codon densities and induce unwanted cellular responses [8]. Although omitting CHX pretreatment has become a common practice, eliminating artifacts introduced by varied protocols remains challenging. A ligation-free Ribo-seq approach was recently introduced [9]. However, the method relies on template-switching technology that is severely biased with a higher efficiency for RNA molecules having a G nucleotide in their 5′ end [10]. As a result, this ligation-free approach is not suitable for Ribo-seq due to the unacceptable 5′ end bias that will distort the subsequent data analysis. Despite the difficulties mentioned, the ribosome profiling methods have undergone continuous development. Several ribosome profiling protocols have been reported [9, 11–13]. Here we introduce two methods for Ribo-seq library construction as routinely performed in our laboratory, the poly-A tailing method and adaptor ligation method.

A hallmark of Ribo-seq is the 3-nt periodicity of RPFs thanks to the relatively precise 5′ end protection by elongating ribosomes. As a result, the percentage of reads mapped to the reading frame 0, or in-frame ratio (IFR), has been commonly used to reflect the resolution of Ribo-seq. Poly-A tailing method (method 1) sacrifices the yield to guarantee 5′ end accuracy of the library which is critical for good triplet periodicity. The IFR based on this method is approximately 50–60%. In marked contrast to method 1, adaptor ligation method (method 2) obtains a high yield; however, the 5′ end accuracy of the library is low. There are ~30% reads whose first nucleotides cannot be aligned to genome, probably due to the untemplated addition during reverse transcription (Fig. 1) [11]. Thus, removal of the first mismatch nucleotide from the 5′ end is needed to obtain accurate P-site. After that, optimization of library construction has improved the IFR of RPFs from ~50% to ~75% (Fig. 2) [11]. In general, successful application of Ribo-seq typically requires cell lysis, nuclease digestion, purification of ribosomes, extraction of RPFs, ribosomal RNA (rRNA) depletion, cDNA library construction, and deep sequencing.

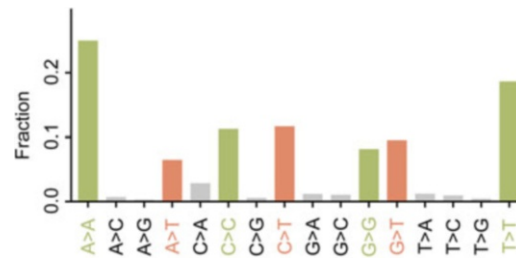

Fig. 1 The mismatch frequency of the first nucleotide from the 5′ end of the trimmed reads based on adaptor ligation method. For comparison, the frequency of matched nucleotides is shown as light green; the mismatches with highest frequency (N>T) are labeled as light red

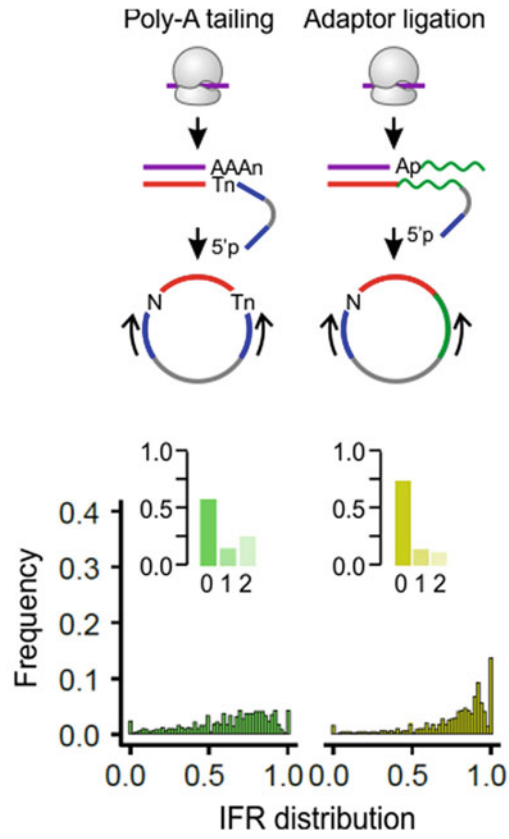

Fig. 2 Schematic representation of ribosome profiling methods using poly-A tailing and adaptor ligation. A direct comparison of the results in terms of IFR resolution is listed below. IFR, in-frame ratio of ribosome footprints

2 Materials

Prepare all solutions using RNase-free, ultrapure water and analytical grade reagents. Prepare and store all buffers and reagents at 4 °C (unless indicated otherwise). For ribosome profiling, pre-sterilized and RNase-free tubes and filter pipette tips should be used.

2.1 Buffers

1. Phosphate-buffered saline (PBS): dilute the 10× Dulbecco's PBS (no calcium, no magnesium) to 1× using RNase-free water, and adjust the pH to 7.1 ± 0.1.
2. Polysome buffer: 10 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid (HEPES), pH 7.4, 5 mM $MgCl_2$, 100 mM KCl.
3. Lysis buffer: polysome buffer with 1% Triton X-100, 5 mM dithiothreitol (DTT), 100 μg/mL CHX.
4. Sucrose solutions: 15%, 45%, and 60% sucrose solutions are prepared in polysome buffer and filtered.

5. RNA gel elution buffer: 300 mM sodium acetate (prepared using 3 M sodium acetate, pH 5.2), 1 mM ethylene diamine tetraacetic acid (EDTA; prepared using 0.5 M EDTA, pH 8.0), 0.1 U/mL SUPERase_In.
6. DNA gel elution buffer: 300 mM NaCl, 1 mM EDTA (prepared using 0.5 M EDTA, pH 8.0).
7. Linearization reaction buffer (10×): 20 mM Tris-acetate, 50 mM potassium acetate, 10 mM magnesium acetate, 1 mM DTT, pH 7.9.

2.2 Reagents

All the reagents listed here are commercially available.

1. *E. coli* RNase I (100 U/μL).
2. Trizol LS reagent.
3. Chloroform.
4. 100% ethanol.
5. 3 M sodium acetate (pH 5.2).
6. Glycogen (5 mg/mL).
7. 70% ethanol (in nuclease-free water).
8. Denaturing 15% polyacrylamide TBE (Tris base-boric acid-EDTA)-urea gel.
9. Denaturing 10% polyacrylamide TBE-urea gel.
10. Non-denaturing 8% polyacrylamide TBE gel (no urea).
11. 2× TBE sample buffer.
12. 5× non-denaturing TBE sample buffer.
13. Ultra-low range DNA ladder.
14. SYBR Gold (nucleic acid gel stain, 10,000×).
15. TBE running buffer.
16. T4 polynucleotide kinase (10 U/μL).
17. RNase inhibitor SUPERase_In (20 U/μL).
18. *E. coli* poly-A polymerase (5 U/μL).
19. SuperScript III (100 U/μL).
20. 10 mM deoxy-ribonucleoside triphosphate (dNTP) mix.
21. RNaseOUT (40 U/μL).
22. CircLigase ssDNA ligase (100 U/μL).
23. Apurinic/apyrimidinic endonuclease 1 (APE 1, 10 U/μL).
24. T4 RNA Ligase 2 (200 U/μL).
25. Phusion DNA polymerase (2 U/μL).
26. BioAnalyzer DNA 1000.

27. A 28 nt random RNA (5′-AUGUACACGGAGUCGACCCG CAACGCGA-3′) which could be used as a size marker in Ribo-seq library construction is prepared by in vitro transcription using T7 RNA polymerase and synthesized DNA template.

2.3 Primers and Linkers for Ribo-seq Library Construction

Barcodes are indicated in italics. rApp represents adenylation; Phos represents phosphorylation; SpC18 represents hexa-ethyleneglycol spacer; N represents random sequence.

1. cDNA synthesis (poly-A tailing method)
   MCA: 5′-p*CAG* ATC GTC GGA CTG TAG AAC TCT CAA GCA GAA GAC GGC ATA CGA TTT TTT TTT TTT TTT TTT TTV N-3′;
   LGT: 5′-p*GTG* ATC GTC GGA CTG TAG AAC TCT CAA GCA GAA GAC GGC ATA CGA TTT TTT TTT TTT TTT TTT TTV N-3′;
   HTC: 5′-p*AGG* ATC GTC GGA CTG TAG AAC TCT CAA GCA GAA GAC GGC ATA CGA TTT TTT TTT TTT TTT TTT TTV N-3′
   YAG: 5′-p*TCG* ATC GTC GGA CTG TAG AAC TCT CAA GCA GAA GAC GGC ATA CGA TTT TTT TTT TTT TTT TTT TTV N-3′;
2. PCR amplification (poly-A tailing method)
   qNTI200 primer: 5′-CAA GCA GAA GAC GGC ATA-3′.
   qNTI201 primer: 5′-AAT GAT ACG GCG ACC ACC GAC AGG TTC AGA GTT CTA CAG TCC GAC G-3′.
3. Linker ligation (adaptor ligation method)
   LC-Linker: rApp/ NNNNCTGTAGGCACCATCAAT / 3ddC
4. cDNA synthesis (adaptor ligation method)
   LC-RT-1: 5′-(Phos)*CTA* NNN AGA TCG GAA GAG CGT CGT GTA GGG AAA GAG TGT AGA TCT CGG TGG TCG C(SpC18)C ACT CA(SpC18) TTC AGA CGT GTG CTC TTC CGA TCT ATT GAT GGT GCC TAC AG-3′
   LC-RT-2: 5′-(Phos)*AGC* NNN AGA TCG GAA GAG CGT CGT GTA GGG AAA GAG TGT AGA TCT CGG TGG TCG C(SpC18)C ACT CA(SpC18) TTC AGA CGT GTG CTC TTC CGA TCT ATT GAT GGT GCC TAC AG-3′
   LC-RT-3: 5′-(Phos)*ATT* NNN AGA TCG GAA GAG CGT CGT GTA GGG AAA GAG TGT AGA TCT CGG TGG TCG C(SpC18)C ACT CA(SpC18) TTC AGA CGT GTG CTC TTC CGA TCT ATT GAT GGT GCC TAC AG-3′
   LC-RT-4: 5′-(Phos)*CCG* NNN AGA TCG GAA GAG CGT CGT GTA GGG AAA GAG TGT AGA TCT CGG TGG TCG C(SpC18)C ACT CA(SpC18) TTC AGA CGT GTG CTC TTC CGA TCT ATT GAT GGT GCC TAC AG-3′

5. PCR amplification (adaptor ligation method)
   LC-Seq-F: 5′-AAT GAT ACG GCG ACC ACC GAG ATC TAC AC-3′
   LC-Seq-R: 5′-CAA GCA GAA GAC GGC ATA CGA GAT GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG-3′

2.4 Equipment

1. Ultracentrifuge.
2. Gradient Master.
3. Polysome fractionation system.
4. Refrigerated microcentrifuge.
5. Dry block heater.
6. Polyacrylamide gel box.
7. Electrophoresis power supply.
8. Gel imaging system.
9. Thermal cycler.
10. Spin-X columns.

3 Methods

Carry out all procedures at room temperature unless otherwise specified.

3.1 Cell Lysis for Ribosome Profiling

The estimated time for cell lysis is approximately 1–2 h according to the number of samples.

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 into at least four 10 cm petri dishes. Adjust the splitting ratio so that the cells are at approximately 80% confluence on the day of cell lysis.
2. Aspirate medium from one dish and immediately cool the dishes on ice. Gently wash the cells with 5 mL of ice-cold PBS. Aspirate the PBS thoroughly from the dish.
3. Add 400 μL of ice-cold lysis buffer to the dish. Detach the cells by scraping the entire petri dish, and pipet up and down several times to lyse the cells.
4. Repeat **steps 2–3** until all the petri dishes are done (using the same lysis buffer).
5. Transfer the cell lysates from the last petri dish to a 1.5 mL tube, and place the tube on ice.
6. Remove the debris from the cell lysates by centrifugation for 10 min at 13,000 × *g* at 4 °C.
7. Transfer the soluble supernatant to a new 1.5 mL tube and put on ice.

3.2 Polysome Gradient

The estimated time for polysome gradient is approximately 5–6 h according to the number of samples.

1. Prepare sucrose density gradients (15–45% (wt/vol)) in ultracentrifuge tubes using a Gradient Master according to the manufacturer's instructions.
2. Load cell lysate onto sucrose gradients, followed by centrifugation for 100 min at 250,000 × *g* (in an SW41 rotor), 4 °C.
3. Fractionate separated samples at 0.375 mL/min by using a fractionation system that continually monitors OD254 values. Collect fractions into 1.5 mL tubes at 1 min intervals.

3.3 RNase I Digestion of Ribosome Fractions

The estimated time for RNase I digestion and RNA extraction is approximately 3 h.

1. Take 30 μL from each fraction from the monosome to the end of polysome. Mix thoroughly to get 300 μL total samples for library construction.
2. For RNase I digestion of ribosome fractions, add *E. coli* RNase I (100 U/μL) into samples (750 U per 100 A260 units, which is approximately 5 μL/300 μL). Then incubate at 4 °C for 1 h.
3. Extract total RNA using Trizol LS reagent (3 volume of the sample) and 200 μL of chloroform. Vigorously mix the tube and let the mixture stand for 2–3 min.
4. Centrifuge at 15,000 × *g* for 15 min at 4 °C.
5. Transfer the upper aqueous phase containing the RNA to a new tube, and then add 2.5 volume of ethanol, 0.1 volume of 3M sodium acetate (pH 5.2), and 4 μL glycogen (5 mg/mL).
6. Incubate at −20 °C for 30 min, followed by centrifuging at 15,000 × *g* for 15 min at 4 °C. Carefully take out all the liquid using a 1 mL filtered tip; do not disturb the RNA pellet.
7. Wash RNA pellet with 1 mL 70% ethanol; spin at 15,000 × *g* for 5 min at 4 °C.
8. Carefully take out all the liquid from the tube. Air-dry the RNA pellet for at least 5 min.

3.4 cDNA Library Construction (Poly-A Tailing Method)

The estimated time for cDNA library construction using poly-A tailing method is approximately 4 days (not including the time for rRNA depletion; *see* **Note 1**). In general, Subheadings 3.4.1 and 3.4.2 take 1 day; Subheadings 3.4.3–3.4.6 take 1 day; Subheadings 3.4.7–3.4.9 take 1 day; and Subheading 3.4.10 takes 1 day.

3.4.1 RNA Dephosphorylation

From this step, a 28 nt random RNA (0.5 μL) (5′-AUGUACACG GAGUCGACCCGCAACGCGA-3′) should be treated together with RNA samples in the following experiments, and it can be used as a size marker.

1. Total RNA samples (RNA dissolved in 11 μL nuclease-free water) are dephosphorylated in a 15 μL reaction containing 10× T4 polynucleotide kinase buffer (1.5 μL), 10 U RNase inhibitor SUPERase_In (20 U/μL, 0.5 μL), and 20 U T4 polynucleotide kinase (10 U/μL, 2 μL).
2. Carry out dephosphorylation in a 1.5 mL tube for 2 h at 37 °C.
3. Heat-inactivate the enzyme for 25 min at 65 °C.

3.4.2 RNA Size Selection and Extraction

1. Mix the dephosphorylated samples with 2× TBE sample buffer, heat at 70 °C for 3 min, and load on a denaturing 15% polyacrylamide TBE-urea gel. 1 μL of ultra-low range DNA ladder and 28 nt random RNA are used as size markers.
2. Run the gel at 160 V for 1.5 h. Stain the gel with 1× SYBR Gold (diluted in TBE running buffer) in the dark for 15–20 min to visualize the RNA fragments.
3. Excise gel bands containing RNA species corresponding to 28 nt region, and physically disrupt the gels using centrifugation through the holes of the tube (*see* **Note 2**).
4. Dissolve RNA fragments by soaking overnight (more than 8 h) in 400 μL RNA gel elution buffer at 4 °C.
5. Remove the gel debris using a Spin-X column by centrifuging at 13,000 × *g* for 2 min. Transfer the liquid to a new 1.5 mL tube.

3.4.3 Ethanol Precipitation

1. Mix 400 μL RNA-containing liquid, 890 μL ice-cold ethanol, 40 μL 3 M sodium acetate (pH 5.2), and 4 μL glycogen (5 mg/mL).
2. Leave the tubes at −20 °C for 30 min.
3. Centrifuge the tubes at 15,000 × *g* for 15 min at 4 °C.
4. Carefully take out all the liquid using a 1 mL filtered tip; do not disturb the RNA pellet.
5. Wash RNA pellet with 500 μL 70% ethanol (in nuclease-free water); spin at 15,000 × *g* for 5 min at 4 °C.
6. Carefully take out all the liquid from the tube. Air-dry the RNA pellet for at least 5 min at room temperature.

3.4.4 Poly-A Tailing

1. Resuspend purified RNA fragments in 5.2 μL of nuclease-free water.
2. Perform poly-A tailing reaction in an 8 μL system with 10× poly-A polymerase buffer (0.8 μL), 10 mM ATP (1 μL), 20 U/μL SUPERase_In (0.4 μL), and 5 U/μL *E. coli* poly-A polymerase (0.6 μL).
3. Carry out tailing in a 1.5 mL tube for 45 min at 37 °C.

3.4.5 cDNA Synthesis

1. Mix tailed RNA product (8 μL) with 10 mM dNTP mix (1 μL) and 2.5 mM synthesized primer (1 μL), and incubate at 70 °C for 5 min, followed by incubation on ice for 5 min.
2. Then add 10× SuperScript III buffer (2 μL), 5 mM $MgCl_2$ (4 μL), 100 mM DTT (2 μL), 40 U/μL RNaseOUT (1 μL), and 100 U/μL SuperScript III (1 μL) into the reaction mix.
3. Incubate at 50 °C for 50 min, and then heat at 85 °C for 5 min.

3.4.6 Size Selection and DNA Extraction

1. Mix the reverse transcription samples with 20 μL 2× TBE sample buffer; heat at 70 °C for 3 min.
2. Run the samples and ultra-low range DNA ladder on a 10% polyacrylamide TBE-urea gel for 1.5 h at 160 V.
3. Stain the gel with 1× SYBR Gold in the dark for 15–20 min.
4. The extended first-strand product band is expected to be about 93 nt, and the corresponding region is excised, referred to the ultra-low range DNA ladder. Physically disrupt the gels using centrifugation (*see* **Note 2**).
5. The cDNA in gel is recovered overnight in 400 μL DNA gel elution buffer.
6. Then remove the debris using Spin-X column (13,000 × *g* for 2 min).
7. Ethanol precipitate the cDNA with 890 μL ice-cold ethanol, 40 μL 3 M sodium acetate (pH 5.2), and 4 μL glycogen (5 mg/mL). Leave the tubes at −20 °C for 30 min. Centrifuge the tubes at 15,000 × *g* for 15 min at 4 °C. Carefully take out all the liquid using a 1 mL filtered tip; do not disturb the pellet. Wash the pellet with 500 μL 70% ethanol (in nuclease-free water); spin at 15,000 × *g* for 5 min at 4 °C. Carefully take out all the liquid from the tube.
8. Air-dry the pellet for at least 5 min at room temperature, and resuspend the cDNA in 11.5 μL of nuclease-free water.

3.4.7 Circularization

1. First-strand cDNA is circularized in 20 μL of reaction containing 10× CircLigase buffer (2 μL), 50 mM $MnCl_2$ (1 μL), 5 M betaine (4 μL), and 100 U/μL CircLigase ssDNA ligase (1.5 μL).
2. Perform circularization in a 1.5 mL tube at 60 °C for 1.5 h, and then heat-inactivate the reaction at 80 °C for 10 min.
3. Add 380 μL nuclease-free water to the circularized DNA.
4. Precipitate the DNA with 890 μL ice-cold ethanol, 40 μL 3 M sodium acetate (pH 5.2), and 4 μL glycogen (5 mg/mL). Leave the tubes at −20 °C for 30 min.

5. Centrifuge the tubes at 15,000 × *g* for 15 min at 4 °C. Carefully take out all the liquid using a 1 mL filtered tip; do not disturb the pellet.
6. Wash the pellet with 500 μL 70% ethanol (in nuclease-free water); spin at 15,000 × *g* for 5 min at 4 °C. Carefully take out all the liquid from the tube.
7. Air-dry the pellet for at least 5 min at room temperature, and resuspend in 8.25 μL of nuclease-free water.

3.4.8 Linearization

1. Circular single-strand DNA (8.25 μL) is relinearized with 10× linearization reaction buffer (1 μL) and 10 U/μL APE 1 (0.75 μL).
2. Carry out the reaction in a 1.5 mL tube at 37 °C for 1 h.

3.4.9 Size Selection and DNA Extraction

1. Mix the linearized single-strand DNA samples with 10 μL 2× TBE sample buffer; heat at 70 °C for 3 min. Separate the linearized single-strand DNA and ultra low range DNA ladder on a denaturing 10% polyacrylamide TBE-urea gel for 1.5 h at 160 V.
2. Stain the gel with 1× SYBR Gold in the dark for 15–20 min.
3. Then excise the expected 93 nt product bands according to the ultra-low range DNA ladder, and physically disrupt the gels using centrifugation (*see* **Note 2**). Recover them overnight in DNA gel elution buffer.
4. Remove the debris using Spin-X column (13,000 × *g* for 2 min).
5. Precipitate the DNA with 890 μL ice-cold ethanol, 40 μL 3 M sodium acetate (pH 5.2), and 4 μL glycogen (5 mg/mL). Leave the tubes at −20 °C for 30 min.
6. Centrifuge the tubes at 15,000 × *g* for 15 min at 4 °C. Carefully take out all the liquid using a 1 mL filtered tip; do not disturb the pellet.
7. Wash the pellet with 500 μL 70% ethanol (in nuclease-free water); spin at 15,000 × *g* for 5 min at 4 °C. Carefully take out all the liquid from the tube.
8. Air-dry the pellet for at least 5 min at room temperature, and resuspend in 10 μL nuclease-free water.

3.4.10 PCR Amplification and Size Selection

1. Set up the PCR reaction including 11.75 μL nuclease-free water, 5 μL 5× Phusion buffer, 0.5 μL 10 mM dNTP mix, 1.25 μL 10 mM qNTI200 primer, 1.25 μL 10 mM qNTI201 primer, 5 μL template DNA, and 0.25 μL of 2 U/μL Phusion DNA polymerase.

2. Perform PCR amplification as follows: 98 °C 30 s, 98 °C 10 s, 60 °C 20 s, 72 °C 10 s for 10–12 cycles; 72 °C 10 min; 4 °C hold (*see* **Note 3**).
3. Add 5× non-denaturing TBE sample buffer to the PCR products, no heat.
4. Separate PCR products on a non-denaturing 8% polyacrylamide TBE gel (no urea) for 1 h at 160 V using ultra-low range DNA ladder as size marker.
5. Stain the gel with 1× SYBR Gold in the dark for 15–20 min.
6. Then excise the expected DNA at 120 bp, and physically disrupt the gels using centrifugation (*see* **Note 2**). Recover overnight using DNA gel elution buffer.
7. Remove the debris using Spin-X column (13,000 × *g* for 2 min).
8. Precipitate the DNA with 890 μL ice-cold ethanol, 40 μL 3 M sodium acetate (pH 5.2), and 4 μL glycogen (5 mg/mL). Leave the tubes at −20 °C for 30 min.
9. Centrifuge the tubes at 15,000 × *g* for 15 min at 4 °C. Carefully take out all the liquid using a 1 mL filtered tip; do not disturb the pellet.
10. Wash the pellet with 500 μL 70% ethanol (in nuclease-free water); spin at 15,000 × *g* for 5 min at 4 °C. Carefully take out all the liquid from the tube.
11. Air-dry the pellet for at least 5 min at room temperature, and resuspend DNA pellet with 10 μL nuclease-free water.

3.5 cDNA Library Construction (Adaptor Ligation Method)

The estimated time for cDNA library construction using adaptor ligation method is approximately 4 days (not including the time for rRNA depletion; *see* **Note 1**). In general, Subheading 3.5.1 takes 1 day; Subheadings 3.5.2–3.5.4 take 1 day; Subheadings 3.5.5 and 3.5.6 take 1 day; Subheadings 3.5.7 and 3.5.8 take 1 day.

3.5.1 RNA Size Selection and Extraction

1. Resuspend RNA pellet in 10 μL nuclease-free water.
2. Add 10 μL 2× TBE sample buffer; heat at 70 °C for 3 min.
3. Run the RNA samples, a 28 nt random RNA control, and ultra-low range DNA ladder on a denaturing 15% polyacrylamide TBE-urea gel for 90 min at 160 V.
4. Stain the gel with 1× SYBR Gold for 15 min on the rotator in dark. Visualize the gel and cut the band between 26 and 34 nt.
5. Physically disrupt the gel using centrifugation through the holes of the tube (*see* **Note 2**).
6. Add 400 μL RNA gel elution buffer to the sample tube. Let it agitate at 4 °C overnight.

7. Remove the gel debris with Spin-X column centrifuging at 13,000 × *g* for 2 min.
8. Purify RNA using 890 μL ice-cold ethanol, 40 μL 3 M sodium acetate (pH 5.2), and 4 μL glycogen (5 mg/mL). Leave the tubes at −20 °C for 30 min.
9. Centrifuge the tubes at 15,000 × *g* for 15 min at 4 °C. Carefully take out all the liquid using a 1 mL filtered tip; do not disturb the RNA pellet.
10. Wash the RNA pellet with 500 μL 70% ethanol (in nuclease-free water); spin at 15,000 × *g* for 5 min at 4 °C. Carefully take out all the liquid from the tube.
11. Air-dry the RNA pellet for at least 5 min at room temperature, and resuspend the RNA pellet in 11 μL of nuclease-free water.

3.5.2 RNA Dephosphorylation

1. Set up a 15 μL dephosphorylation reaction in a 1.5 mL tube including the RNA sample (11 μL), 10× T4 polynucleotide kinase buffer (1.5 μL), 10 U RNase inhibitor SUPERase_In (20 U/μL, 0.5 μL), and 20 U T4 polynucleotide kinase (10 U/μL, 2 μL).
2. Incubate the sample at 37 °C for 1 h to perform the dephosphorylation, followed by heat inactivation at 65 °C for 10 min.
3. Purify the dephosphorylated RNA using 890 μL ice-cold ethanol, 40 μL 3 M sodium acetate (pH 5.2), and 4 μL glycogen (5 mg/mL). Leave the tubes at −20 °C for 30 min.
4. Centrifuge the tubes at 15,000 × *g* for 15 min at 4 °C. Carefully take out all the liquid using a 1 mL filtered tip; do not disturb the RNA pellet.
5. Wash the RNA pellet with 500 μL 70% ethanol (in nuclease-free water); spin at 15,000 × *g* for 5 min at 4 °C. Carefully take out all the liquid from the tube.
6. Air-dry the RNA pellet for at least 5 min at room temperature, and resuspend RNA in 9 μL of nuclease-free water.

3.5.3 Linker Ligation

1. Add 1 μL of LC linker (0.15 μg/μL) to the sample (9 μL) in a 1.5 mL tube.
2. Incubate at 80 °C for 90 s, and let it cool to room temperature.
3. Set up a 20 μL linker ligation reaction: 50% PEG8000 6 μL, 10× T4 RNA Ligase 2 buffer 2 μL, sample + linker 10 μL, RNase inhibitor SUPERase_In (20 U/μL) 1 μL, and T4 RNA Ligase 2 (200 U/μL) 1 μL.
4. Incubate at 22 °C for 2.5 h.
5. Purify the dephosphorylated RNA with 890 μL ice-cold ethanol, 40 μL 3 M sodium acetate (pH 5.2), and 4 μL glycogen (5 mg/mL). Leave the tubes at -20 °C for 30 min.

6. Centrifuge the tubes at 15,000 × *g* for 15 min at 4 °C. Carefully take out all the liquid using a 1 mL filtered tip; do not disturb the RNA pellet.
7. Wash the RNA pellet with 500 μL 70% ethanol (in nuclease-free water); spin at 15,000 × *g* for 5 min at 4 °C. Carefully take out all the liquid from the tube (*see* **Note 4**).

3.5.4 RNA Size Selection and Extraction

1. Air-dry the pellet for at least 5 min at room temperature, and resuspend RNA pellet in 10 μL nuclease-free water.
2. Add 10 μL 2× TBE sample buffer; heat at 70 °C for 3 min.
3. Run the samples and ultra-low range DNA ladder on a denaturing 10% polyacrylamide TBE-urea gel for 90 min at 160 V.
4. Stain the gel with 1× SYBR Gold in the dark for 15–20 min.
5. Visualize the gel and cut the band between 40 and 60 nt. Physically disrupt the gels using centrifugation through the holes of the tube (*see* **Note 2**).
6. Add 400 μL RNA gel elution buffer to the gel piece, and let it agitate overnight at 4 °C.
7. Remove the gel debris with Spin-X column centrifuging at 13,000 × *g* for 2 min.
8. Purify the RNA with 890 μL ice-cold ethanol, 40 μL 3 M sodium acetate (pH 5.2), and 4 μL glycogen (5 mg/mL). Leave the tubes at −20 °C for 30 min.
9. Centrifuge the tubes at 15,000 × *g* for 15 min at 4 °C. Carefully take out all the liquid using a 1 mL filtered tip; do not disturb the RNA pellet.
10. Wash the RNA pellet with 500 μL 70% ethanol (in nuclease-free water); spin at 15,000 × *g* for 5 min at 4 °C. Carefully take out all the liquid from the tube.
11. Air-dry the pellet for at least 5 min at room temperature, and resuspend the RNA pellet in 11 μL of nuclease-free water.

3.5.5 cDNA Synthesis

1. Add 1 μL of 2.5 μM of LC-RT primer to the RNA sample (11 μL).
2. Denature at 80 °C for 2 min and on ice for 5 min.
3. Set up a 20 μL reverse transcription reaction: 5× SuperScript III buffer 4 μL, sample + primer 12 μL, 100 mM DTT 1 μL, 10 mM dNTP mix 1 μL, 40 U/μL RNaseOUT 1 μL, and 100 U/μL SuperScript III 1 μL.
4. Carry out reverse transcription at 48 °C for 30 min, followed by 10 min at 75 °C.

5. Purify the cDNA with 890 μL ice-cold ethanol, 40 μL 3 M sodium acetate (pH 5.2), and 4 μL glycogen (5 mg/mL). Leave the tubes at −20 °C for 30 min.
6. Centrifuge the tubes at 15,000 × *g* for 15 min at 4 °C. Carefully take out all the liquid using a 1 mL filtered tip; do not disturb the pellet.
7. Wash the pellet with 500 μL 70% ethanol (in nuclease-free water); spin at 15,000 × *g* for 5 min at 4 °C. Carefully take out all the liquid from the tube.

3.5.6 Size Selection and DNA Extraction

1. Air-dry the pellet for at least 5 min at room temperature, and resuspend cDNA in 10 μL nuclease-free water.
2. Add 10 μL of 2× TBE sample buffer; heat at 70 °C for 3 min.
3. Run the samples, LC-RT primer, and ultra-low range DNA ladder on a denaturing 10% polyacrylamide TBE-urea gel for 90 min at 160 V.
4. Stain the gel with 1× SYBR Gold in the dark for 15–20 min.
5. Visualize the gel and cut the band ~200 nt, above the 125 nt primer. Physically disrupt the gel (*see* **Note 2**).
6. Add 400 μL DNA gel elution buffer to the sample tube. Let it agitate at 4 °C overnight.
7. Remove the gel debris with Spin-X column centrifuging at 13,000 × *g* for 2 min.
8. Ethanol precipitate the DNA with 890 μL ice-cold ethanol, 40 μL 3 M sodium acetate (pH 5.2), and 4 μL glycogen (5 mg/mL). Leave the tubes at −20 °C for 30 min.
9. Centrifuge the tubes at 15,000 × *g* for 15 min at 4 °C. Carefully take out all the liquid using a 1 mL filtered tip; do not disturb the pellet.
10. Wash the pellet with 500 μL 70% ethanol (in nuclease-free water); spin at 15,000 × *g* for 5 min at 4 °C. Carefully take out all the liquid from the tube.
11. Air-dry the pellet for at least 5 min at room temperature, and resuspend the DNA pellet in 15 μL of nuclease-free water.

3.5.7 Circularization

1. Set up a 20 μL circularization reaction in a 1.5 mL tube: single-stranded cDNA 15 μL, 10× CircLigase buffer 2 μL, 1 M ATP 1 μL, 50 mM $MnCl_2$ 1 μL, and CircLigase ssDNA ligase (100 U/μL) 1 μL.
2. Incubate at 60 °C for 1 h followed by 80 °C for 10 min.
3. Purify the cDNA with 890 μL ice-cold ethanol, 40 μL 3 M sodium acetate (pH 5.2), and 4 μL glycogen (5 mg/mL). Leave the tubes at −20 °C for 30 min.

4. Centrifuge the tubes at 15,000 × *g* for 15 min at 4°C. Carefully take out all the liquid using a 1 mL filtered tip; do not disturb the pellet.
5. Wash the pellet with 500 μL 70% ethanol (in nuclease-free water); spin at 15,000 × *g* for 5 min at 4 °C. Carefully take out all the liquid from the tube.
6. Air-dry the pellet for at least 5 min at room temperature, and resuspend cDNA pellet in 12.75 μL nuclease-free water.

3.5.8 PCR Amplification and Size Selection

1. Set up a 20 μL reaction: 5× Phusion buffer 4 μL, 10 mM dNTP mix 1 μL, cDNA 12.75 μL, 10 μM LC-Seq-F primer 1 μL, 10 μM LC-Seq-R primer 1 μL, and 2 U/μL Phusion DNA polymerase 0.25 μL.
2. Perform PCR reaction: 98 °C for 30 s; 98 °C for 10 s, 65 °C for 20 s, 72 °C for 10 s, 10–12 cycles; 72 °C for 5 min; 4 °C hold (*see* **Note 3**).
3. Add 5× non-denaturing TBE sample buffer to the sample, no heat.
4. Load the samples and ultra-low range DNA ladder on 8% non-denaturing polyacrylamide TBE gel (no urea), 160 V for 60 min.
5. Stain the gel with 1× SYBR Gold for 15 min on the rotator in the dark.
6. Visualize the gel and cut the band ~180 bp. Physically disrupt the gel (*see* **Note 2**).
7. Add 400 μL DNA gel elution buffer to the sample tube, and let it agitate at 4 °C overnight.
8. Remove the debris using Spin-X column centrifuging at 13,000 × *g* for 2 min.
9. Precipitate the DNA with 890 μL ice-cold ethanol, 40 μL 3 M sodium acetate (pH 5.2), and 4 μL glycogen (5 mg/mL). Leave the tubes at −20 °C for 30 min.
10. Centrifuge the tubes at 15,000 × *g* for 15 min at 4 °C. Carefully take out all the liquid using a 1 mL filtered tip; do not disturb the pellet.
11. Wash the pellet with 500 μL 70% ethanol (in nuclease-free water); spin at 15,000 × *g* for 5 min at 4 °C. Carefully take out all the liquid from the tube.
12. Air-dry the pellet for at least 5 min at room temperature, and resuspend DNA pellet with 15 μL nuclease-free water.

3.6 Deep Sequencing

Measure DNA concentration and length by Bioanalyzer DNA 1000, and then pool equal amount of each barcoded sample into one tube. Send mixed DNA samples for sequencing using Illumina

sequencing primer with a read length of 51 nt. The loading concentration of mixed DNA samples is approximately 1.5–1.8 pM in 1.3 mL.

3.7 Sequencing Data Analysis

Raw sequencing reads are first processed by trimming 3′ adapters and low-quality bases. The trimmed reads with length <15 nucleotides are excluded, and the remaining reads are then mapped to the transcriptome. For read alignment, the reads mapped to multiple positions or with more than two mismatches are excluded. The 13th position (12-nt offset from the 5′ end) of the uniquely mapped read is defined as the ribosome P-site (*see* **Note 5**). For each individual transcript, reads per kilobase per million mapped reads (RPKM) is used to compute the ribosome density. To exclude the effect of RNA level, ribosome density is normalized by corresponding RNA level.

4 Notes

1. rRNA contamination is a common phenomenon in Ribo-seq library construction. It derives largely from RNase I cleavage of surface-exposed regions of ribosomes, which generates rRNA fragments similar in size to RPFs [5]. To reduce rRNA contamination, commercial rRNA depletion kits can be used for the depletion of rRNA from Ribo-seq libraries [6].
2. To physically disrupt the gels, punch five to six holes using a needle in the 0.5 mL tube. Put the excised gel in the tube. Then put the tube above a 1.5 mL tube, and centrifuge at 13,000 × *g* for 10 min or more.
3. During PCR amplification, sometimes it takes 12 cycles to see a clear band. Sometimes 10 cycles are enough, and 12 cycles can make the band becoming smear-like. Thus, it is recommended to try both to find the best condition.
4. Purification can be omitted if the starting material is scarce.
5. Here we describe two methods of Ribo-seq library construction. As each methodology has its advantages and disadvantages, method for Ribo-seq library construction should be chosen according to experimental purpose and specific condition. Poly-A tailing method (method 1) sacrifices the yield to guarantee 5′ end accuracy of the library which is critical for good triplet periodicity. In marked contrast to method 1, adaptor ligation method (method 2) obtains a high yield; however, the 5′ end accuracy of the library is low, which should be adjusted during subsequent bioinformatic analysis by removal of the first mismatch nucleotide from the 5′ end.

Acknowledgments

This work was supported by the US National Institutes of Health (R01GM1222814 and R21CA227917) and HHMI Faculty Scholar (55108556).

References

1. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324 (5924):218–223. https://doi.org/10.1126/science.1168978
2. Brar GA, Weissman JS (2015) Ribosome profiling reveals the what, when, where and how of protein synthesis. Nat Rev Mol Cell Biol 16(11):651–664. https://doi.org/10.1038/nrm4069
3. Wolin SL, Walter P (1988) Ribosome pausing and stacking during translation of a eukaryotic mRNA. EMBO J 7(11):3559–3569
4. Stern-Ginossar N, Weisburd B, Michalski A, Le VT, Hein MY, Huang SX, Ma M, Shen B, Qian SB, Hengel H, Mann M, Ingolia NT, Weissman JS (2012) Decoding human cytomegalovirus. Science 338(6110):1088–1093. https://doi.org/10.1126/science.1227919
5. Ingolia NT, Lareau LF, Weissman JS (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell 147(4):789–802. https://doi.org/10.1016/j.cell.2011.10.002
6. Gao X, Wan J, Liu B, Ma M, Shen B, Qian SB (2015) Quantitative profiling of initiating ribosomes in vivo. Nat Methods 12(2):147–153. https://doi.org/10.1038/nmeth.3208
7. Gerashchenko MV, Gladyshev VN (2014) Translation inhibitors cause abnormalities in ribosome profiling experiments. Nucleic Acids Res 42(17):e134. https://doi.org/10.1093/nar/gku671
8. Santos DA, Shi L, Tu BP, Weissman JS (2019) Cycloheximide can distort measurements of mRNA levels and translation efficiency. Nucleic Acids Res 47(10):4974–4985. https://doi.org/10.1093/nar/gkz205
9. Hornstein N, Torres D, Das Sharma S, Tang G, Canoll P, Sims PA (2016) Ligation-free ribosome profiling of cell type-specific translation in the brain. Genome Biol 17(1):149. https://doi.org/10.1186/s13059-016-1005-1
10. Turchinovich A, Surowy H, Serva A, Zapatka M, Lichter P, Burwinkel B (2014) Capture and Amplification by Tailing and Switching (CATS). An ultrasensitive ligation-independent method for generation of DNA libraries for deep sequencing from picogram amounts of DNA and RNA. RNA Biol 11 (7):817–828. https://doi.org/10.4161/rna.29304
11. 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 (8):1534–1550. https://doi.org/10.1038/nprot.2012.086
12. McGlincy NJ, Ingolia NT (2017) Transcriptome-wide measurement of translation by ribosome profiling. Methods 126:112–129. https://doi.org/10.1016/j.ymeth.2017.05.028
13. Reid DW, Shenolikar S, Nicchitta CV (2015) Simple and inexpensive ribosome profiling analysis of mRNA translation. Methods 91:69–74. https://doi.org/10.1016/j.ymeth.2015.07.003

Chapter 11

Choice of Ribonucleases for Ribosome Profiling Experiments

Maxim V. Gerashchenko

Abstract

In the past 10 years, standard transcriptome sequencing protocols were optimized so well that no prior experience is required to prepare the sequencing library. Often, all enzymatic steps are designed to work in the same reaction tube minimizing handling time and reducing human errors. Ribosome profiling stands out from these methods. It is a very demanding technique that requires isolation of intact ribosomes, and thus there are multiple additional considerations that must be accounted for (McGlincy and Ingolia, Methods 126:112–129, 2017). In this chapter, we discuss how to select a ribonuclease to produce ribosomal footprints that will be later converted to the sequencing library. Several ribonucleases with different cutting patterns are commercially available. Selecting the right one for the experimental application can save a lot of time and frustration.

Key words Ribosome profiling, Translation, Ribonuclease, rRNA

1 Introduction

Ribosome-mRNA complexes have to be treated with a ribonuclease (RNase) to eliminate mRNA parts located outside of the ribosome and to generate so-called ribosomal footprints. However, the ribosome itself contains ribosomal RNA which is also exposed to the ribonuclease. During the RNase treatment, the ribosome can suffer damage and lose structural integrity in which case the quality and the composition of ribosomal footprints will be compromised. First ribosome profiling studies were conducted on yeast ribosomes, and they happened to tolerate any kind of nuclease-inflicted damage [2, 3]. Ribosomes from other organisms differ in their stability (Fig. 1). Moreover, even ribosomes isolated from different organs of the same species might display different tolerance to the same ribonuclease [2]. The exact nature of such discrepancies remains unclear. It is particularly intriguing considering the fact that rRNA and protein sequence and composition of the eukaryotic ribosome are highly similar across species and nearly identical within a species.

Vyacheslav M. Labunskyy (ed.), *Ribosome Profiling: Methods and Protocols*, Methods in Molecular Biology, vol. 2252, https://doi.org/10.1007/978-1-0716-1150-0_11,

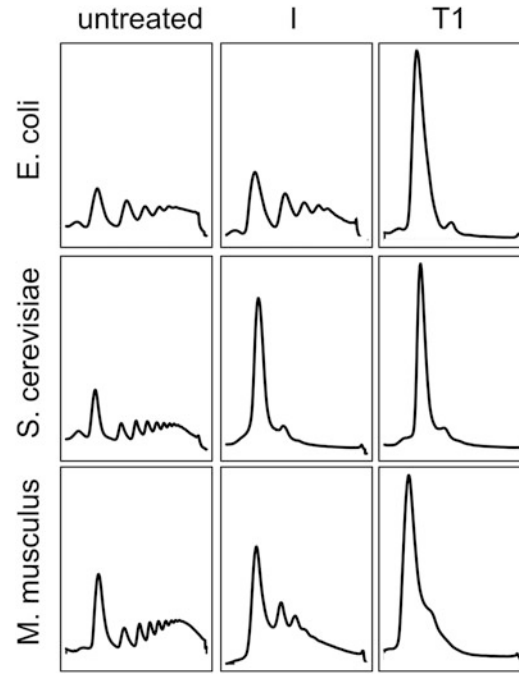

Fig. 1 Sucrose gradients of polysomes digested with RNase I or RNase T1. RNase I is inactive in *E. coli* lysates. RNase I aggressively over digests polysomes from the mouse liver

Nevertheless, it has direct implications for any experimental scientist who is interested in assessing translation by ribosome profiling. Most of the published ribosome profiling studies use RNase I in eukaryotes and RNase S7 in prokaryotes. However, there are several other ribonucleases commercially available and using them may save a lot of time depending on the application.

Ribonuclease must maintain activity at the physiological conditions required to support ribosome structural integrity. Several commercially available nucleases qualify these criteria: A, T1, S7, and I. They possess different cutting preferences and can be used individually or in a mix.

1.1 Ribonucleases and Their Properties

1. RNase I preferentially hydrolyzes single-stranded RNA to nucleoside 3′-monophosphates. It does not require metal ions for activity and has no cutting preference [4].
2. RNase A specifically degrades single-stranded RNA at C and U residues by cleaving a phosphodiester bond between the 5-′-ribose of a nucleotide and the phosphate group attached to the 3′-ribose of an adjacent pyrimidine nucleotide [5]. Every ribosomal footprint generated by this enzyme has pyrimidine (C or U) at the 3′ end.

3. RNase T1 specifically degrades single-stranded RNA at G residues. It cleaves the phosphodiester bond between the 3′-guanylic residue and the 5′-OH residue of adjacent nucleotides. The enzyme does not require metal ions for activity [6]. Every ribosomal footprint generated by this enzyme has G at the 3′ end (*see* **Note 1**).
4. RNase S7 (Micrococcal nuclease) hydrolyzes single and double-stranded RNA to yield 3′-mononucleotides and oligonucleotides. More likely to cleave AT-rich regions than GC-rich. The enzyme requires Ca^{2+} ions for activity [7].

2 Materials

2.1 Ribonucleases

1. RNase I, 100 U/μL.
2. RNase T1, 1000 U/μL.
3. RNase S7, 10 mg lyophilized powder. To prepare a stock solution, dissolve in 100 μL of the storage buffer: 10 mM Tris–HCl pH 7.5, 50 mM NaCl, 1 mM EDTA, 50% glycerol.
4. RNase A, 1 mg/mL.

2.2 Other Reagents

1. 0.5 M ethylene glycol-bis(β-aminoethyl ether)-*N,N,N′,N′*-tetraacetic acid, sodium salt (EGTA).
2. Superase-In RNase inhibitor.
3. EDTA-free protease inhibitor cocktail.
4. Cycloheximide.
5. Triton X-100.

2.3 Sucrose Gradients

1. Polypropylene centrifuge tubes (14 × 89 mm).
2. Beckman ultracentrifuge compatible with SW41 rotor.
3. Fraction collection system (optional).
4. Brandel tube piercing stand.
5. BioComp Gradient Master.
6. UV detector with 254 nm filter (separate or as a part of an automated gradient fractionator).
7. Syringe pump.
8. Gradient sucrose solutions: 10% and 50% sucrose in polysome buffer without detergents, no $CaCl_2$, and increased $MgCl_2$ to 10 mM.
9. Sucrose chase solution: 60% sucrose, 10 mM $MgCl_2$, 100 mg/mL cycloheximide.

2.4 Subtractive Hybridization

1. Streptavidin magnetic beads.
2. Magnetic stand.
3. 20× saline sodium citrate (SSC) buffer: 3 M NaCl in 0.3 M sodium citrate (pH 7.0).
4. Magnetic beads wash buffer: 10 mM Tris–HCl pH 7.5, 0.5 mM EDTA, 1 M NaCl.

3 Methods

There are many ribosome profiling studies performed in various model organisms, tissues, and cell lines which provides a rich reference resource for selecting RNase treatment compatible with the specimen origin. However, since ribosome isolation requires experience, the experimental conditions that worked well in one laboratory may not reproduce in another one. Therefore, it is important to outline experimental goals before doing ribosome profiling for the first time. Generally, ribosome profiling can be used to address the translation efficiency of certain genes as well as the fine structural details of translation such as the presence of upstream regulatory reading frames, ribosome pausing, and stalling sites. The first part is straightforward and similar to the standard transcriptome analysis of gene expression. It only requires a basic understanding of computational biology; results are easily interpretable, and many studies are exclusively interested in this aspect alone. The second part is quite challenging, it requires extensive experience in bioinformatics, and many studies do not invest in this sort of analysis.

RNase I is the best choice when the fine details of translation are studied. It has no cutting preference; therefore, ribosomal footprints have the same offset of codons located in the A/P/E (aminoacyl, peptidyl, exit) sites regardless of the transcript sequence. It is very convenient for computational analysis as it allows to generate ribosome occupancy plots and to identify the proper reading frame. RNase I cannot be used in *E. coli* and many other prokaryotes because they express it natively along with the inhibitor which renders exogenous RNase I inactive even if added in large quantities (Fig. 1). Despite its benefits, RNase I has a serious drawback – it easily compromises the structural integrity of the ribosome causing over digestion. It requires a series of sucrose gradient profiles done prior to attempting footprint isolation and sequencing. In some species, for example, *Drosophila melanogaster*, obtaining high-quality monosomes is hardly possible even after rigorous fine-tuning of ribonuclease digestion conditions [8, 9].

Spending time on testing various digestion conditions is often unnecessary if the researcher is only interested in translation efficiency and gene expression estimates. In which case RNase I can be

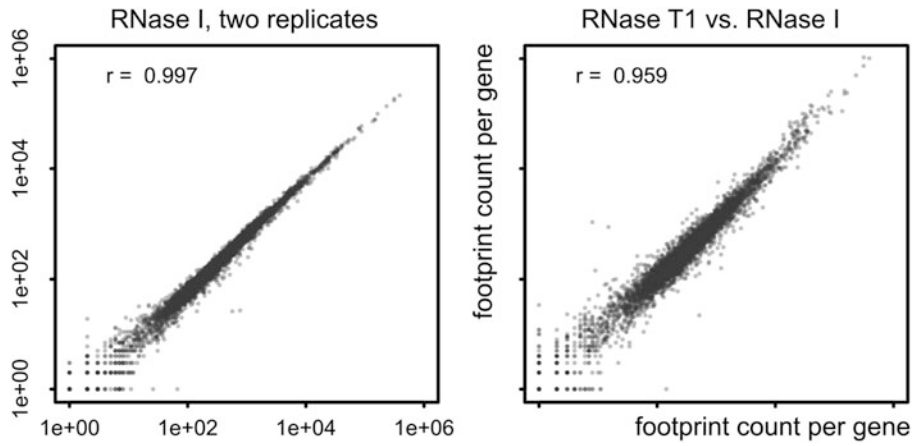

Fig. 2 Footprints per gene count in the sample treated with different ribonucleases

substituted with another nuclease. Although other ribonucleases have a cutting preference, it does not change the footprint per gene count compared to RNase I (Fig. 2). The reason is rather simple—an average protein-coding gene is much longer than a footprint and lacks low-complexity regions [2].

3.1 Test Ribosome Stability in the Presence of Ribonucleases

Ribosome quality is critical for sequencing. Ideally, the complete conversion of polysomes to monosomes without the loss of monosomes is desired (Fig. 1).

1. Prepare a lysate in the 2 mL of the buffer of your choosing. Different species favors slightly different lysis buffers; here is the standard one to get started: 20 mM Tris–HCl pH 7.5, 50 mM KCl, 50 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 1% Triton, 0.1 mg/mL cycloheximide, protease inhibitors.
2. Meanwhile, prepare sucrose gradients in the Gradient Master from BioComp following manufacturer's instructions.
3. Spin down lysates in a table-top centrifuge for 1 min, and transfer the supernatant in a new microtube.
4. Split the lysate into 6 equal parts and bring the volumes to 1 mL with the fresh lysis buffer.
5. Set-up the ribonuclease digestion test (*see* **Note 2**):
   - Untreated control.
   - RNase I.
   - RNase A.
   - RNase S7.
   - RNase T1.
   - RNase S7 + RNase T1.

Keep the reaction at room temperature for 30 min with gentle rotation. The exact quantities of ribonucleases do not matter much at this stage. We suggest adding 1 μL of each enzyme per 10 units of the lysate (OD_{260}).

6. Inactivate ribonucleases. Add 5 mM $CaCl_2$ to inactivate RNase S7. Add 2 μL Superase-In to inactivate RNase I, RNase T1, and RNase A.
7. Carefully overlay 1 mL from each sample lysate on top of the sucrose gradients. Perform ultracentrifugation in SW41 rotor for 3 h at 35,000 rpm (151,000 × *g*).
8. Plug the sample tube in a tube piercing device connected to UV detector. Pump 60% sucrose solution through the tube at 1 mL/min rate. Heavy sucrose solution will displace the sucrose gradient upward making it flow through UV detector.

3.2 Quantify the Number of Monosomes and Polysomes

A nuclease digests the mRNA regions unprotected by ribosomes. As a result, polysomes disappear and the monosome peak grows proportionally. Several techniques can be used to calculate areas under the peaks on a sucrose gradient track. We recommend using the custom R script (*see* **Note 3**). Make sure the starting amount of ribosomes was the same in every sample. It greatly helps with peak calculations because sucrose profiles have the same absorbance baseline. In our experience, RNase T1 was the only nuclease that consistently preserved ribosome integrity while thoroughly converting polysomes to monosomes in a variety of species, tissues and cell lines [2]. RNase S7 holds the second-best place.

3.3 Fine-Tuning Nuclease Digestion Conditions

Based on the ribonuclease testing, there is a good chance RNase T1 or RNase S7 deliver acceptable performance. However, there are several parameters that can be further changed to increase or decrease digestion efficacy.

- Temperature. It often helps to reduce over digestion by incubating reaction mix on ice or in a cold room instead of the room temperature.
- Incubation time can be reduced to 20 min or extended to 1 h.
- Enzyme quantity has the least impact.
- Order of the treatments. In the instance where the cutting bias of individual RNases has to be avoided and RNase I cannot be used, a mixture of RNase S7 and T1 produces acceptable results. These nucleases can be added as a mixture at the same time or sequentially. In the latter case, the treatment can be done with RNase S7 first, inactivated with $CaCl_2$, then RNase T1 can be added to the same lysate.

3.4 Improve Footprints Yield by Subtractive Hybridization

Ribonucleases have distinct cutting preference, and it affects the repertoire of ribosomal RNA (rRNA) fragments that will contaminate the resulting ribosomal profiling libraries. There exact percentage of this contamination greatly varies depending on the species, tissue and cell type, and proportion of ribosomes engaged in translation. A common strategy to remove unwanted rRNA fragments is to do subtractive hybridization. It is usually done right after ribosomal footprints were reverse transcribed into single stranded DNA counterparts. Custom DNA oligonucleotide designed to be reverse-complement to the rRNA serve as a bait to fish them out. There are several commercially available kits to do this kind of rRNA depletion; however they are typically designed for enriching full-length mRNA and eliminating full-length or at least partially degraded rRNA. As a result, they perform poorly when rRNA is highly fragmented and fragments are short as in the case of ribosome profiling. These kits also target only commonly used model organisms and may not be suitable for other species.

Custom oligonucleotide baits are a viable alternative to commercial kits. We recommend sequencing one ribosome profiling library with low coverage just to characterize the content of rRNA contaminants. Figure 3 demonstrates rRNA coverage produced by different ribonucleases in *S. cerevisiae* [2]. Although the original sample was identical, ribonucleases generated fragments from different rRNA regions, and it has to be accounted for while designing the baits. We usually select 5–7 regions that yield the highest number of contaminants and design biotinylated oligonucleotides 50–60 nt in length to target them. The location of the biotin is not important; it can be at the 3′ as well as 5′ terminus.

Interestingly, the same ribonuclease can have different rRNA digestion patterns in different tissues and cell types. For instance, Fig. 4 shows examples from several mouse organs treated with the same mixture of RNase T1/S7 [10]. Therefore, hybridization baits must be selected carefully to maximize the yield of useful footprints. Of course, it only makes sense for medium to large-scale projects, while smaller projects can be sequenced to sufficient depth even if the rRNA content is high.

1. Prepare a mix of biotinylated oligonucleotide baits in the 1× SSC buffer. If the rRNA digestion pattern is known, maintain the same proportion of individual baits as the composition of the rRNA fragments. Refer to the magnetic beads manufacturer to learn the binding capacity.
2. Prepare streptavidin magnetic beads by washing 3–4 times with a high salt buffer: 10 mM Tris–HCl pH 7.5, 0.5 mM EDTA, 1 M NaCl. If the beads are not certified as RNAse/DNAse-free, wash 3 times in 0.1 M NaOH, 0.05 M NaCl solution, repeat 3 times with just 0.05 M NaCl. Add biotinylated baits

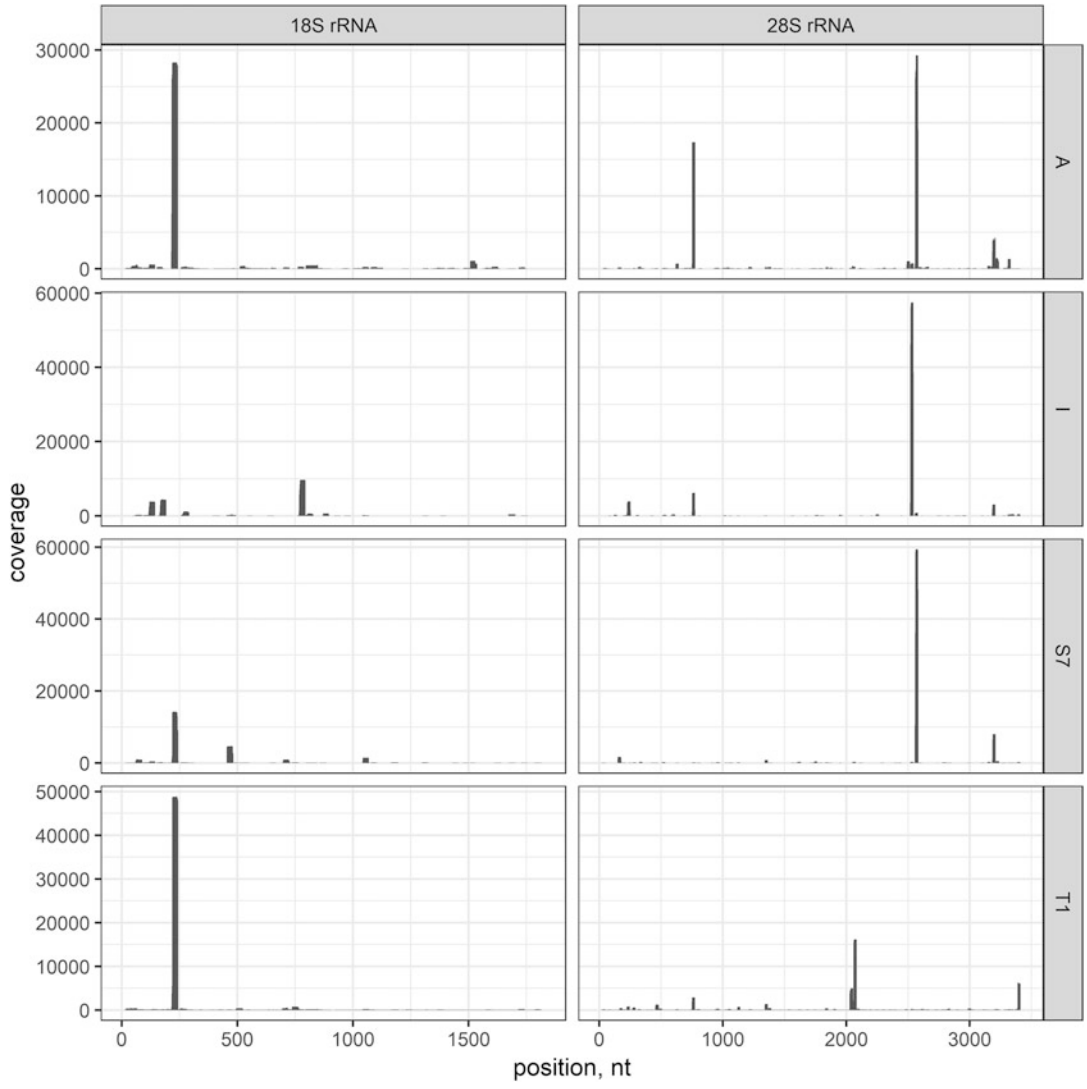

Fig. 3 rRNA contamination generated by ribonucleases. RNAse digestion patterns of 18S and 28S rRNA in *S. cerevisiae*. High peaks represent major contaminants and should be targeted by subtractive hybridization

and incubate 30 min at room temperature with gentle agitation. Wash off the excess of unbound baits by rinsing beads 3 times with the high salt buffer. Repeat washing three times with 1× SSC buffer (a commonly used buffer for nucleic acid hybridization).

3. Set up hybridization by combining ssDNA from ribosome profiling with ssDNA baits. Make sure you have at least four-fold excess of baits over the ribosome profiling ssDNA. Adjust the total volume to 50 μL and maintain 1× SSC. Mix well, then

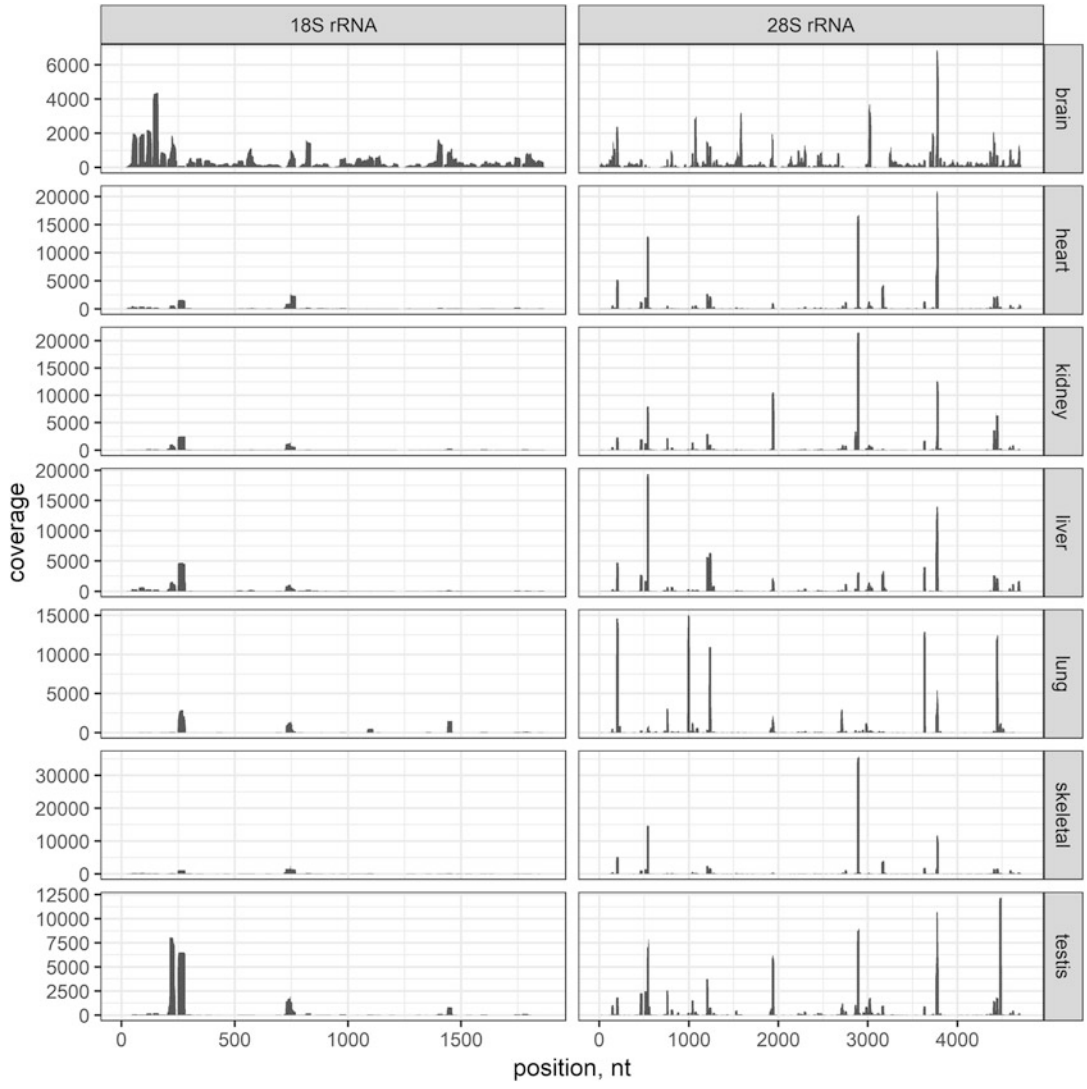

Fig. 4 rRNA contamination generated by ribonucleases. RNAse digestion patterns of 18S and 28S rRNA in various mouse organs. The mixture of RNAse T1/S7 is used in each case

place at 80 °C for 5 min in a PCR machine. Slowly decrease the temperature to 50 °C and transfer to streptavidin magnetic beads. Incubate at 37 °C for 15 min with gentle rotation. For increased annealing specificity incubate at 50 °C if possible.

4. Place the tube on a magnetic stand for 1 min and transfer the supernatant containing ribosome depleted ssDNA library in a new microtube. Proceed with ethanol precipitation and ribosome profiling library preparation of your choice.

4 Notes

1. Enzymes are never 100% accurate, so strictly speaking, some fraction of footprints will have nucleotides other than specified at 3′ end. Moreover, different ribosome sources may have a high content of endogenous ribonucleases, skewing the representation of terminal nucleotides. In the case of the ribonuclease T1, we observed anywhere from 95% to 85% of all footprints ending with guanine.
2. This layout is optimized for an SW41 ultracentrifuge rotor which holds six samples. The experimenter is free to use any combination of ribonucleases as he deems fit.
3. The R code required for plotting and quantifying sucrose gradient profiles is available from https://github.com/germaximus/SucroseProfiler.

References

1. McGlincy NJ, Ingolia NT (2017) Transcriptome-wide measurement of translation by ribosome profiling. Methods 126:112–129. https://doi.org/10.1016/j.ymeth.2017.05.028
2. Gerashchenko MV, Gladyshev VN (2017) Ribonuclease selection for ribosome profiling. Nucleic Acids Res 45(2):e6. https://doi.org/10.1093/nar/gkw822
3. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324 (5924):218–223. https://doi.org/10.1126/science.1168978
4. Shen V, Schlessinger D (1982) 16 RNases, I, II, and IV of *Escherichia coli*. In: Boyer PD (ed) The enzymes, vol 15. Academic Press, pp 501–515. https://doi.org/10.1016/S1874-6047(08)60288-7
5. Raines RT (1998) Ribonuclease A. Chem Rev 98(3):1045–1066. https://doi.org/10.1021/cr960427h
6. Pace CN, Heinemann U, Hahn U, Saenger W (1991) Ribonuclease T1: structure, function, and stability. Angew Chem Int Ed English 30 (4):343–360. https://doi.org/10.1002/anie.199103433
7. Heins JN, Suriano JR, Taniuchi H, Anfinsen CB (1967) Characterization of a nuclease produced by *Staphylococcus aureus*. J Biol Chem 242(5):1016–1020
8. Dunn JG, Foo CK, Belletier NG, Gavis ER, Weissman JS (2013) Ribosome profiling reveals pervasive and regulated stop codon readthrough in Drosophila melanogaster. elife 2:e01179. https://doi.org/10.7554/eLife.01179
9. Miettinen TP, Bjorklund M (2015) Modified ribosome profiling reveals high abundance of ribosome protected mRNA fragments derived from 3′ untranslated regions. Nucleic Acids Res 43(2):1019–1034. https://doi.org/10.1093/nar/gku1310
10. Gerashchenko MV, Peterfi Z, Yim SH, Gladyshev VN (2021) Translation elongation rate varies among organs and decreases with age. Nucleic Acids Res 49(2):e9. https://doi.org/10.1093/nar/gkaa1103

Part II

Bioinformatics and Statistical Analysis

Chapter 12

Codon Resolution Analysis of Ribosome Profiling Data

Alexander Bartholomäus and Zoya Ignatova

Abstract

Translation is a central biological process in living cells. Ribosome profiling approach enables assessing translation on a global, cell-wide level. Extracting versatile information from the ribosome profiling data usually requires specialized expertise for handling the sequencing data that is not available to the broad community of experimentalists. Here, we provide an easy-to-use and modifiable workflow that uses a small set of commands and enables full data analysis in a standardized way, including precise positioning of the ribosome-protected fragments, for determining codon-specific translation features. The workflow is complemented with simple step-by-step explanations and is accessible to scientists with no computational background.

Key words Ribo-Seq, Ribosome profiling, Deep sequencing, Translation and regulation, Analysis pipeline, Calibration, Nucleotide resolution, Codon

1 Introduction

Translation of the genetic information from a nucleotide sequence into a functional protein is a central biological process in all living cells. A large proportion of the cellular resources is dedicated to translation: it consumes up to the half of the energy production of the cell, and approximately 35–45% of the genomes are assigned to proteins of the translation apparatus [1, 2]. Systematic studies quantifying transcripts and proteins at genomic scales frequently reveal poor correlation, which is partly due to translational regulation [3–5]. Translation is a multilayered process, orchestrated by the ribosome. A new emerging concept suggests that the dynamics of the ribosome coordinates the multivalent functions from decoding of the genetic information to producing a functional protein. Examples with various proteins have demonstrated that the ribosomes translate mRNAs with non-uniform speed [6] which kinetically synchronizes processes downstream of translation, including co-translational folding, translocation, assembly, and processing of the nascent protein [6–13]. Thus, direct analyses of translation

Vyacheslav M. Labunskyy (ed.), *Ribosome Profiling: Methods and Protocols*, Methods in Molecular Biology, vol. 2252, https://doi.org/10.1007/978-1-0716-1150-0_12,

provide a more accurate and complete measure of gene expression and protein biogenesis compared to analyzing mRNA expression levels alone [3].

Ribosomes cover a stretch of 25–30 nucleotides with slightly larger fragments in eukaryotes than in prokaryotes. Currently, the state-of-the-art technology for transcriptome-scale quantification of translating ribosomes is ribosome profiling (Ribo-Seq) [14]. It is based on a deep sequencing of ribosome-protected fragments (RPFs), following treatment with ribonucleases that digest unprotected, ribosome-free mRNA segments [15] allowing the ribosome position to be monitored with a single-nucleotide or sub-codon precision on a transcriptome-wide scale [14]. Assignment of the RPFs to the transcriptome, known also as mapping, at its most basic level provides information of the regions of RNA being translated, and from the RPF density, the level of translation on open-reading frames can be inferred [16]. In permissive or nutrient-balanced growth conditions, variations in the RPF density within an open-reading frame reflect differences in translation speed or efficiency. Adverse growth conditions (e.g., environmental stress) or aberrant mRNAs are sensitively sensed by the translation apparatus, activate ribosome-associated quality control mechanisms, and cause ribosomal stalling and/or queuing [17–19]; hence non-uniform distribution of the RPFs within open-reading frames would report on aberrant translation events. Ribosome profiling sets, generated in conditions to report on genuine translation, further precise positioning of the RPFs (also known as calibration) at the ribosomal A, P, and E sites (the three sites the ribosome binds tRNAs, i.e., aminoacyl-tRNA, peptidyl-tRNA, and tRNA-exit site, respectively) [20], can provide insights into codon-specific features of translation [21, 22] and mechanisms of specific reading frames being translated [23].

Currently, several resources and tools are available to analyze ribosome profiling data; each of them, however, bears a specialized focus. These include repository databases, enabling comparison and integration of ribosome profiling data sets [24], or tools to detect open-reading frames [25, 26], extract translation regulation regimes [27], and calibrate RPFs [28–30]. Attempts for interactive analysis and processing have also been made [31]. Yet, most of these tools require different level of programming knowledge and are thus inaccessible to the large community of experimentalists.

The precise positioning of RPFs within the open-reading frame, or calibration of RPFs, is the key step to obtain codon resolution and extract codon-dependent translation features. In this analysis, the RPFs are anchored using their 5′ or 3′ ends to common translation positions to determine the position of the start codon in the ribosomal P site and the stop codon in the A site, respectively. Few exiting tools implement automatic procedures to determine offset of the RPF read ends to the ribosomal A or P site

[28–30]. However, the offset strongly depends on different factors, e.g., used nucleases, RPFs length distributions, and variations in the sequencing protocols among laboratories [32, 33]. Thus, manual inspection is still necessary to validate and correct automatically determined offsets.

In this chapter, we demonstrate how the ribosome profiling data can be used to assess the position of actively translating ribosomes in cells growing under permissive conditions, thereby inferring the codon position of the ribosomal A, P, or E sites. Our descriptions cover the full range of data processing with special focus on precise positioning of the RPFs in the ribosomal A, P, or E sites. Our manual approach implements detection at translation start and stop for each read length separately. We show a computational pipeline with step-by-step instructions to process and analyze ribosome profiling data to obtain codon-specific information from the RPFs. We also provide brief descriptions of how tools can be used and combined at each step. Our detailed explanations and schematic visualizations along with example data and commands provide a workflow that is easy to use, in particular by experimentalists without programming background, and simple to modify to achieve high precision in ribosome profiling data analysis. Using only a dozen of simple commands, we cover the entire pipeline from reads mapping and exact RPF positioning to generation of standard plots and data export. This enables non-expert users to perform the analysis in a standardized fashion.

2 Materials

2.1 Software Dependencies

The pipeline requires software that runs on Linux operating system (OS). macOS should work too but was not tested. For compatibility reasons we recommend using the most recent versions. We provide the versions we used to verify the pipeline. The explicit installation commands are given on the GitHub page (see link below).

The software for sequencing reads mapping is optional. Other mapping procedures can be applied too.

1. Linux operation system (Ubuntu 18.04. LTS).
2. R (version 3.5.0) (https://www.R-project.org) using the following packages:
   - foreach (version 1.4.7).
   - doParallel (version 1.0.15).
3. bedtools (version 2.26.0) (https://bedtools.readthedocs.io) [34]
4. samtools (version 1.7) (http://www.htslib.org/) [35]

5. [Optional] For the sequencing reads processing and mapping/alignment to reference:
   - cutadapt (version 2.5) (https://cutadapt.readthedocs.io) [36].
   - bowtie (version 2.26.0) [37].

2.2 Download and Prepare the Pipeline

1. Download. The pipeline is written as a set of R commands and functions that often wrap up system commands. All scripts and configuration files used in this chapter can be downloaded from `https://github.com/AlexanderBartholomaeus/MiMB_ribosome_profiling`,
   or the git clone can be used:

```
git clone git://github.com/AlexanderBartholomaeus/
MiMB_ ribosome_profiling.git
```

2. Verify that the necessary tools are installed. Open a terminal window, and navigate to the downloaded or cloned directory, and verify that all tools listed in Subheading 2.1 are installed by executing the following: `Rscript verify_dependencies.R`
   If the script finishes without errors, the system is ready. If errors occur, check whether the tools are installed correctly, or change the default path to a tool in the config/tools_location.csv. This file can be opened and modified with a text editor (e.g., gedit) or office program (e.g., OpenOffice). Some tools might be installed locally instead of globally; the latter requires system's administrator rights. For example, if cutadapt is updated with a local user rights, it might be located in `/home/user/.local/bin/cutadapt`. The line #3 in the `tools_location.csv` file would look like this: `cutadapt, /home/user/.local/bin/cutadapt`. If you change the location of any tool, run the `verify_dependencies.R` script again. If you damaged config or other files, the original files should be downloaded again.
3. Install R dependencies packages. To install the necessary R packages, the installation script can be run from the command line: Rscript `install_R_packages.R`
   After installing all tools, continue with the analyses.

2.3 Genome and Sequencing Data

The following data and information are needed to apply the pipeline:

1. Genome/transcriptome sequence(s): Those are usually in a form of a FASTA file containing chromosomes or transcripts, respectively. This information is needed to map the sequencing reads (RPFs).

2. Gene/transcript annotation of the coding sequence: This information is usually provided in `GFF/GFF3` (`GFF` version 3), `GBK` (genebank), or `BED` (bedtools) format. We recommend using `BED` format which is the simplest. BED format is used with six columns that cover all necessary information, including the DNA strand (*see* **Note 1**).
3. Ribosome profiling sequencing reads: Raw reads are provided in `FASTQ` format, while mapped reads are generally stored as `BAM` files. If mapped data are generated by in-house procedures, verify that the `FASTA` and `BED` files are matching, e.g., contain the exact same transcript/gene names. Files of each format are included in the example data folder.

2.4 Example Sequencing Data and Code

To provide a set of explicit examples, we assembled data files and working code examples. The example data files and the example code can be found in `example_data/` folder and at the end of each subsection in the Methods, respectively. These code and data were used to generate the plots for Figs. 2 and 3.

Sequencing data: We use publicly available data [38] to illustrate the workflow. For the example data file `rpf.fastq.gz`, we use the first one million reads of SRR1734437 (https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=SRR1734437) which contains the RPFs of *E. coli* MG1655. The data can be found in the `example_data/sequencing_data` folder.

Code: For the code examples, it is recommended to execute the scripts from the location of the downloaded Git folder. The output folder will be one directory higher (as indicated by `../`). However, the paths can be freely adapted to the needs of the analysis.

2.5 Hardware Requirements

We use the fast and memory-efficient tools bowtie, cutadapt, samtools, bedtools, and Linux commands for most of our operations. Few steps of the processing and plotting are implemented in R directly. Depending on the executed steps and the input data, the hardware requirements may vary, e.g., mapping of hundreds of millions of reads to large eukaryotic genomes requires more memory and CPU time as compared to mapping of few million reads to a prokaryotic genome. The majority of the intermediate BAM files are kept during the pipeline run, as they can serve as inputs at different steps.

We suggest working on a workstation with 16 GB of RAM and 8-core CPU and 10x larger disk space than that of FASTQ or BAM input file size. Our default configuration uses 4 cores. This is adjustable by changing the number in the `config/cores_max.csv` file. If no mapping is performed, 8 GB RAM and 2-core CPU are sufficient; however, the disk space for the intermediate files is still needed. After storing the results, we suggest deleting the output folders. Note that the `coverage_start_stop.R` (Subheadings 3.3, **step 2** and 3.4, **step 7**) and `calibration_count_plot.R` (Subheading 3.4, **step 2**) read large matrices of

position-based counts that can be very large if a large number of expressed genes is used. Therefore, to reduce the RAM usage, consider using smaller gene set or a higher threshold (Subheading 3.3, **step 1**), if permitted by the research question.

3 Methods

The following instructions provide a step-by-step manual. We offer detailed explanations and highlight key steps and discuss various alternative options. In addition, we provide a Git repository with all necessary files and commands to execute each step of the pipeline, which can be found in: https://github.com/AlexanderBartholomaeus/MiMB_ribosome_profiling. We use R programming language. Most commands are wrapped up into system commands using basic Linux tools or samtools and bedtools, thus enabling very fast processing. The plots are generated using base R. To benefit of the multi-core processing, we use R packages foreach and doParallel. Many processes, however, are I/O intensive, e.g., the procedures of selecting and modifying reads are computationally simple but will generate a new file of large size. We restrict these intensive processes to 8-cores.

3.1 [Optional] Sequencing Read Pre-processing and Mapping

To cover the whole workflow, from raw sequencing reads to final results including plots, we provide also a basic mapping procedure. This basic mapping strategy works well because the RPFs are short; the procedure is complex, when pursuing a multi-step mapping with rRNA removal first. Several different tools for mapping exist which are reviewed in [39] and can also be used. If the sequencing data have been already mapped, proceed with Subheading 3.2. Here, we provide a basic procedure based on bowtie [37] which is more sensitive for short reads than bowtie2. The full suite of mapping parameters and their impact on ribosome profiling data analysis is summarized in [39].

To prepare the sequencing reads for mapping, the sequencing adapters and low-quality bases must be removed. Subsequently, reads are mapped to a reference genome or transcriptome. The pipeline is designed to work with Illumina sequencing reads but can be modified to any other sequencing platform and read types.

1. Set adapter removal and quality trimming parameters. Open the `config/mapping_config.csv`. The lines starting with `cutadapt_` indicate `cutadapt` parameters. Please set the parameters according to your preferences. Most importantly, set the 3′ adapter.
2. Set sequencing read mapping/alignment parameters. In the same `config/mapping_config.csv` files, lines starting with `bowtie_` indicate `bowtie` parameters. Note that the recommended settings can be changed.

3. Pre-process and map/align reads. Start the pre-processing and mapping by calling the script, followed by the input folder containing the raw sequencing reads (`FASTQ` or `FASTQ.GZ` files), the output folder (which will be created, if not existing), and the path to the `FASTA` of the reference:

```
Rscript mapping.R input_folder output_folder reference_file
```

With our example data, the call would look like this:

```
Rscript mapping.R example_data/sequencing_data/ my_out_folder
example_data/genome_data/E_coli_genome.fa
```

In the output folder the following three different subfolders will be created:

- `adapterRemoval` – It contains the adapter and quality-trimmed sequencing reads (`FASTQ.GZ` format).
- `bowtie_index` – It contains the index of the reference that is necessary for the mapping process.
- `mapping` – It contains the mapped reads in BAM format. This folder serves as an input for several other scripts (see Fig. 1).

4. Example code:

```
Rscript mapping.R example_data/sequencing_data/
../out example_data/genome_data/E_coli_genome.fa
```

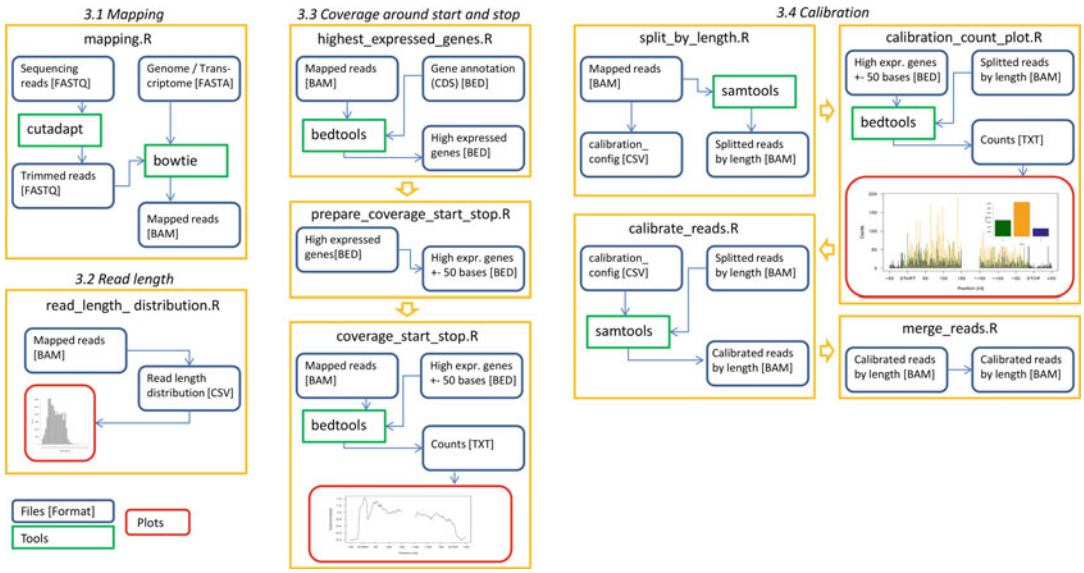

Fig. 1 Workflow overview. Scripts are shown in yellow boxes; files, tools, and plots in blue, green, and red, respectively. Yellow arrows indicate the flow of the different scripts. Blue arrows show the interaction of files and tools within scripts

3.2 Statistics and Filtering Ribosomal Reads by Length

Usually RPFs have different lengths. However, actively translating ribosomes should produce RPFs with a certain narrow length [40]. For example, some RPF lengths indicate ribosomal decay, i.e., truncated RPFs resulting from an upstream mRNA cleavage and ribosome stalling at 3′ ends [41]. The following scripts allow for obtaining the general distribution of the read length in the data set, and filter and select the range of RPF lengths of interest.

1. Create read length distribution. Read length distribution is also a suitable indicator for assessment of the quality of ribosomal profiling sets. Discrete ribosomal states are characterized by different RPF fragments [40, 41]. The input for the script is a folder with the mapped sequencing reads (as generated in Subheading 3.1 or with your own procedure).

```
Rscript read_length_distribution.R in_folder out_folder
```

   The output folder contains the read length distributions for each single input (with file name according to the input files), a summarized table (`all_read_length.csv`) for the read lengths of 20–50 nucleotides, and a `PDF` with bar charts of the read length distribution (Fig. 2a) for each input file (*see* **Note 2).**

2. Select and filter the read length(s) of interest. For some analysis only a specific range or even a single read length is suitable, e.g., only the read length characteristic for genuinely translating ribosomes. To select read length(s) of interest, use the following script:

```
Rscript read_length_selection.R in_folder out_folder
length_range_start- length_range_end
```

   To select a single read length, enter the same number for start and end, e.g.:

```
Rscript read_length_selection.R in_folder out_folder 26-26
```

   The resulting `BAM` file is named according to the input with addition of the read length range. The files can be used as input for other steps of the pipeline, for example, in Subheading 3.3.

3. Example code:

```
Rscript read_length_distribution.R ../out/mapping/ ../out
```

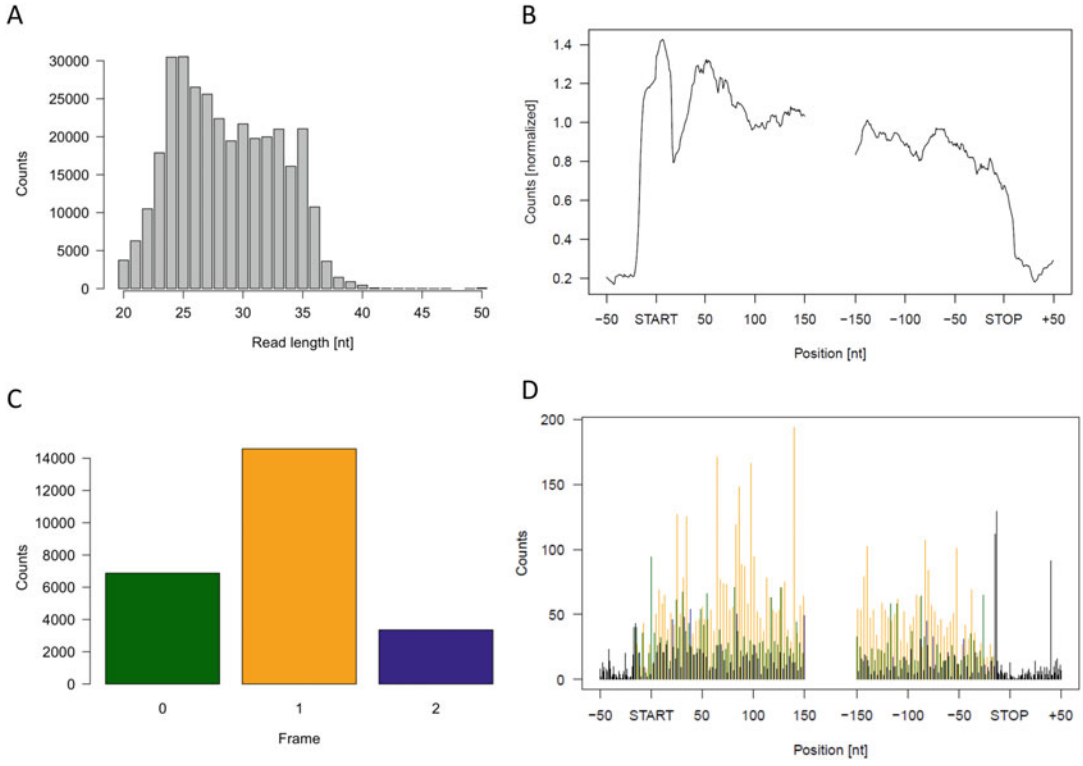

Fig. 2 Example of plots generated by the scripts of the workflow. (**a**) Read length distribution of RPFs. (**b**) Coverage plot of aggregated full-length RPFs around the start and stop codons. (**c**) Summed RPFs in the coding sequence for each frame. (**d**) RPFs of a specific length around the start and stop codons using the 5′ nucleotide of the reads (5′ calibration). The counts of the three different frames color coded as in (**c**) are shown

3.3 RPF Coverage Within Start and Stop Codon Regions

Coverage plots flanking the start and stop codons have led to different findings related to translation regulation of gene expression. In this section we create the coverage plot using all RPFs, regardless of their length. The same plots can be created using RPFs of a specific length, i.e., the results from Subheading 3.2 **step 2**.

1. Restrict the analysis to sufficiently expressed genes only. To remove lowly expressed genes, which will increase the signal-to-noise ratio, it is necessary to identify genes with a good or at least moderate expression. Choose a sample (e.g., control condition) and the corresponding annotation `BED` file. The threshold for highly expressed genes is set to 0.1 reads per kilobase length of the gene/transcript/feature as a default. However, this value can be changed to higher or lower values.

```
Rscript highest_expressed_genes.R in_bam in_bed
out_folder threshold[optional]
```

In the output folder, the `highest_expressed_genes.bed` is provided that will be used in the following steps. In addition, you can find the file containing the counts.

2. Prepare annotation that focuses at start and stop codons. To assess the nucleotides upstream of the start codon and downstream of the stop codon, it is necessary to modify the gene annotation information so that the base counting includes these positions. If you prepare a file using in-house programs, we recommend to use (−)50 bases upstream of the start codon and (+)50 bases downstream of the stop codon. Our script assumes that the `BED` annotation file represents the coding sequence, because 50 bases upstream and downstream will be added to the gene start and end, respectively. The script can handle eukaryotic annotation information with many exons per gene, but it requires the same names of all exons (*see* **Note 1**).

```
Rscript prepare_coverage_start_stop.R in_file
```

The resulting file name includes the base name of the input file and `_plus_50nt` and is written into the folder of the input file (*see* **Note 3**).

3. Count reads for each base and generate plot. We recommend using genes expressed over the threshold (see Subheading 3.3, **step 1**). The script uses bedtools to count per base using the input folder with the mapped reads and the input `BED` file with the region around that start and stop codons from Subheading 3.3, **step 2**.

```
Rscript coverage_start_stop.R in_folder in_file out_folder
```

The output folder contains one `PDF` with a plot (Fig. 2b) for each input `BAM` file showing the counts around the corresponding start and stop codon region. In addition, there is a folder named `counts/` that contains three different count files that have a specific first part of the filename extended by the name of the input `BAM` file as follows:

- `bedtools_... .txt`: This file is a result of the counting using `bedtools coverage` command and contains the raw read counts for each gene/transcript. The file can be relatively large and does not allow easy access of single genes or regions. However, this file should be kept, because it contains the raw read counts that can be easily inspected, if necessary.
- `all_genes_... .csv`: This file contains the normalized counts for start and stop codon regions, whereby each column is a position and each row a gene. The normalization is performed by dividing the raw counts by the (arithmetic) mean of counts in the coding sequence region.

- `plot_... .csv`: This file contains the final summarized counts that are used to generate the plots. The aggregation is done by the mean. These file can be used to create an own plot.

As mentioned above, there are two steps of averaging needed to weight the genes equally. The first averaging or normalization is within each gene, i.e., division by the mean. This is necessary to equalize the weight of the counts within each gene, whereby each gene is normalized to one. In the second normalization step, the genes are normalized by averaging each position separately for the gene set considered in each analysis. The values in the resulting aggregated plot dwell around a value of 1 which arises from the normalization (division) by the mean of the reads in the coding sequence of each gene. You can follow the counts in the `all_genes_....csv`.

4. Example code:

```
Rscript highest_expressed_genes.R
../out/mapping/rpf.sort.bam
example_data/genome_data/E_coli_genes.bed ../out/ 0.1
Rscript prepare_coverage.R
../out/highest_expressed_genes/highest_expressed_genes.bed
Rscript coverage_start_stop.R ../out/mapping/
../out/highest_expressed_genes/highest_expressed_genes_
plus_50nt.bed ../out/
```

3.4 Calibration of RPFs

A fascinating advantage of ribosome profiling data is the possibility to detect ribosome positions with codon or nucleotide precision. However, usually not all fragments show this precision; thus fuzzy fragments should be filtered out. Depending on organisms and protocols, i.e., used nucleases, the RPFs can vary at both ends. For bacteria the 3′ ends of the RPFs show a sharp signal (also called 3′ assignment or 3′ calibration), whereas for eukaryotes the 5′ RPF ends provide such sharp signal (called 5′ assignment or 5′ calibration) [42]. The differences are attributed to variations in the specificity of the RNases used to generate the RPFs in the prokaryotic and eukaryotic ribosome profiling data sets, respectively, and are discussed in [32]. To determine the read length that carries codon precision, each single read length should be analyzed separately. Subsequently, the offset to ribosomal A and P site can be determined by the overhang at the stop and start codon, respectively. The manual approach presented here is suitable for both 5′ and 3′ ends calibration and thus applicable for prokaryotic and eukaryotic ribosome profiling data sets. Thereby, we recommend 3′ end calibration for procaryotic and 5′ end calibration for eukaryotic data sets.

1. Splitting reads by length and keeping first or last base of each read. To split mapped read into different read lengths and modify the BAM files in such way that each read is represented by its first or last nucleotide only, the following script has to be called. By keeping the first nucleotide, the 5′ calibration is performed, whereas keeping the last nucleotide will result in 3′ calibration. The input folder must contain the mapped BAM files. As input for the `read_base` parameter, one of the following strings should be entered: first, last, 5prime, or 3prime. First and 5prime will take the first nucleotide and, thus, perform calibration using the 5′ read ends. Last and 3prime enable taking the last nucleotide of each read, and hence, calibration using the 3′ ends of the RPFs will be done. You may also specify a read length range of interest. If this is not given, it is set to 24–30, which covers the most common RPF lengths.

```
Rscript split_by_length.R in_folder
out_folder read_base length_range_start-
length_range_stop[optional]
```

   The output folder contains a folder `split_by_length/` with BAM files for each length according to each input file, with additional name tag for firstBase or lastBase. In addition, two files for later calibration configuration are created, `calibration_5prime_config.csv` and `calibration_3prime_config.csv`, each containing an empty table with input BAM file names and the selected read length range. This file will be used in Subheading 3.4, **steps 3** and **4** to exactly determine the ribosomal P site and correct the offset that can be performed individually for each read length.

2. Counting reads for each base and generating plots. This step assumes that Subheading 3.3, **steps 1** and **2** are accomplished. You need a modified BED file that covers the 50 bases upstream and downstream of the start and stop of the coding sequence to determine the offsets later. The input folder should contain the split and modified BAM files from Subheading 3.4, **step 1**. The input file is the modified BED from Subheading 3.3, **step 2**.

```
Rscript calibration_count_plot.R in_folder in_file out_folder
```

   For each input BAM file, a PDF with two plots is generated. In the next step we will explain and use these plots to manually identify the offset of the ribosomal P site.

3. Determining reliable read length and ribosomal P site codon for each read length. To manually determine reliable read length, the two plots from Subheading 3.4, **step 2** are used. The first plot shows the summed counts for each reading frame

(Fig. 2c). We recommend selecting read length showing a non-uniform distribution, i.e., with one frame exhibiting significantly higher counts compared to the other two (in some cases two frames are significantly higher than the third one). If the frame criteria are fulfilled, the offset can be determined from the second plot in which the counts of first/last base of each read are plotted (Fig. 2d) (*see* **Note 4**). Bases at the offsets are color coded (CDS −15 bases upstream for 5′ calibration or CDS +15 bases downstream for 3′ calibration; Fig. 3). The offset should be determined for each input `BAM` file and each read length; usually this offset is the same for same read length among different libraries generated with the same experimental setup (i.e., using the same RNase). Enter the manually determined offset into the `calibration_5prime_config.csv` or the `calibration_3prime_config.csv` table created in Subheading 3.4, **step 1** (depending on whether 5′ or 3′ calibration is performed, respectively). If you do not want to calibrate a specific read length, e.g., because it exhibits not well-defined frame pattern or unclear offset, leave the field in table NA (or 0 or empty). Figure 3 shows an example of calibration using the 5' ends of the reads with a length of 24 nucleotides (*see* **Note 4**). The example of `calibration_5prime_config.csv` file is in the example data and will be used in the example code.

4. Calibrating selected read length. This step will correct the offset that we determined in Subheading 3.4, **step 3**. The final `BAM` files (in the created output folder) contain calibrated reads that show codon resolution. The folder with the mapped and unmodified reads (e.g., from Subheading 3.1) should be used as input folder. Use the `calibration_5prime_config.csv` or `calibration_3prime_config.csv` as input file. Based on the input file name, the script automatically performs 5′ or 3′ calibration.

```
Rscript calibrate_reads.R in_folder in_file out_folder
```

In the next steps, the calibrated files for different read lengths can be merged and used for further analysis. The files can be used to plot the exact position around the start and stop codons as shown in Subheading 3.3, **steps 3** and **4**.

5. Merging calibrated read lengths. To merge different read lengths into one file for later analysis, you can use the following script. All files and thus all read lengths with the same base name in the input folder will be merged. If no out_folder is provided, the merged files are written to the input directory.

```
Rscript merge_reads.R in_folder out_folder[optional]
```

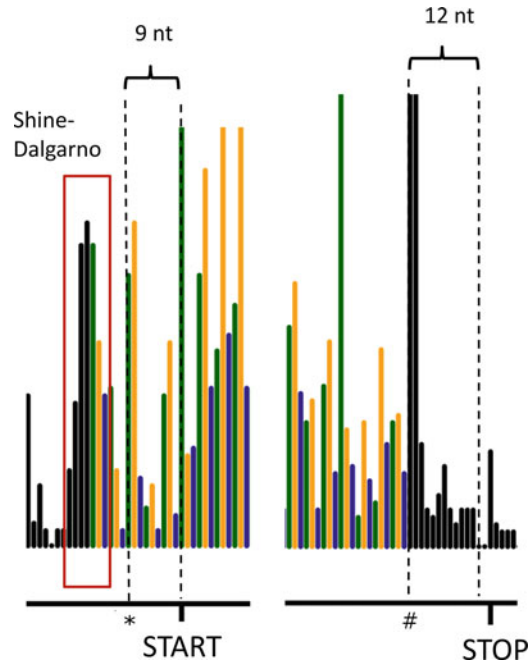

Fig. 3 Determining the offset of RPFs following calibration. To determine the offset, i.e., the distance of the first base of the reads to the first base of the ribosomal P site, we first identify the trinucleotide periodicity that is characteristic for actively translating ribosomes. Here, the yellow frame exhibits the highest signal. This frame should be placed in the middle position of the codons; thus the green frame is the nucleotide representing each first nucleotide of the codon. The offset of the first nucleotide of the trinucleotide periodic frame (*) is nine nucleotides. At the end, the offset of the first nucleotide of last codon of the trinucleotide periodic frame (#) has an offset to the first nucleotide of the STOP codon of 12. Hence, the difference between both offsets determined through the start and stop is exactly three nucleotides. If we would shift only nine bases, the P site would be placed exactly at the last translated codon. The red-framed area indicates nucleotide without trinucleotide periodicity. These are six to seven bases upstream of the first nucleotide delineating the trinucleotide periodic frame and are likely ribosomes assembling at the Shine-Dalgarno sequence

6. [Optional] Generating counts for sufficiently expressed genes using the calibrated reads. You are now ready to generate counts for single genes using the calibrated reads. The exact position of the read allows performing further analysis, e.g., open-reading frames or frameshift detection or codon usage analysis. Here, bedtools can be used directly or the script of Subheading 3.3, **step 3** using folder with calibrated reads as input and the known annotation file (*see* **Note 5**).

7. Example code:

```
Rscript split_by_length.R ../out/mapping/ ../out/ 24-30
Rscript calibration_count_plot.R
../out/calibration/split_by_length/
../out/highest_expressed_genes/highest_expressed_genes_
plus_50nt.bed ../out/
Rscript calibrate_reads.R ../out/mapping/
example_data/calibration_5prime_config.csv ../out
Rscript merge_reads.R ../out/calibration/calibrated/
Rscript coverage_start_stop.R
../out/calibration/calibrated/
../out/highest_expressed_genes/highest_expressed_genes_
plus_50nt.bed ../out/calibrated_coverage
```

4 Notes

1. Only the coding sequence position without UTRs should be used. If sequencing data from eukaryotic species are analyzed, the same gene name for all exons should be used in column #4. Tools like BioMart or the UCSC table browser allow for extracting the required information. If starting with a GFF file, we recommend to store only meaningful information in column #9 of the GFF (e.g., gene/transcript name). To create a BED file from a GFF input, the following R script and command can be used:

```
Rscript gff2bed.R input_file out_folder
```

2. If in parallel processing RNA-Seq data matching the ribosome profiling data, the RNA-Seq libraries should have a distribution similar to a normal distribution.
3. The script expects coding sequence position information without UTRs. For eukaryotic species with different exons per gene/transcript, the gene name should be the same for all exons. This is necessary for the script to correctly add 50 bases to the start and end of the transcript.
4. The base in the plot that is marked START is the first nucleotide of the start codon, whereas STOP marks the last nucleotide of the stop codon. The offset is visible at the start and at the stop, but the signal can be noisy or covered by other signals. As shown in Fig. 3, some counts can be found in the region of the Shine-Dalgarno sequence that mediates in translation initiation in prokaryotes [43]; however, these counts do not show a trinucleotide periodicity as the coding sequence regions. We

expect certain behavior of the ribosomal P site: It should be positioned precisely at the start codon. The P site should not cover the stop codon; stop codons align with the A site. This results in a different offset at the start and stop codons, which should be offset by exactly three nucleotides. We also expect a trinucleotide periodicity within the entire coding sequence.

5. If you use the script of Subheading 3.3, **step 3** and an annotation file without 50 bases added upstream of the start and downstream of the stop codon, respectively, or with only very few genes, the generated plot may not be meaningful. The counts of interest will be found in the counts/ output folder.

Acknowledgments

The work was supported by European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 764591 (SynCrop) and by Deutsche Forschungsgemeinschaft SPP2002 (IG 73/16-1) grants to Z.I.

References

1. Buttgereit F, Brand MD (1995) A hierarchy of ATP-consuming processes in mammalian cells. Biochem J 312(Pt 1):163–167. https://doi.org/10.1042/bj3120163
2. Russell JB, Cook GM (1995) Energetics of bacterial growth: balance of anabolic and catabolic reactions. Microbiol Rev 59(1):48–62
3. Ingolia NT (2014) Ribosome profiling: new views of translation, from single codons to genome scale. Nat Rev Genet 15(3):205–213. https://doi.org/10.1038/nrg3645
4. Liu Y, Beyer A, Aebersold R (2016) On the dependency of cellular protein levels on mRNA Abundance. Cell 165(3):535–550. https://doi.org/10.1016/j.cell.2016.03.014
5. Sonenberg N, Hinnebusch AG (2009) Regulation of translation initiation in eukaryotes: mechanisms and biological targets. Cell 136(4):731–745. https://doi.org/10.1016/j.cell.2009.01.042
6. Zhang G, Ignatova Z (2011) Folding at the birth of the nascent chain: coordinating translation with co-translational folding. Curr Opin Struct Biol 21(1):25–31. https://doi.org/10.1016/j.sbi.2010.10.008
7. Collart MA, Weiss B (2019) Ribosome pausing, a dangerous necessity for co-translational events. Nucleic Acids Res. https://doi.org/10.1093/nar/gkz763
8. Hanson G, Coller J (2018) Codon optimality, bias and usage in translation and mRNA decay. Nat Rev Mol Cell Biol 19(1):20–30. https://doi.org/10.1038/nrm.2017.91
9. Kirchner S, Ignatova Z (2015) Emerging roles of tRNA in adaptive translation, signalling dynamics and disease. Nat Rev Genet 16(2):98–112. https://doi.org/10.1038/nrg3861
10. Kramer G, Boehringer D, Ban N, Bukau B (2009) The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. Nat Struct Mol Biol 16(6):589–597. https://doi.org/10.1038/nsmb.1614
11. Kramer G, Shiber A, Bukau B (2019) Mechanisms of cotranslational maturation of newly synthesized proteins. Annu Rev Biochem 88:337–364. https://doi.org/10.1146/annurev-biochem-013118-111717
12. Schwarz A, Beck M (2019) The benefits of cotranslational assembly: a structural perspective. Trends Cell Biol 29(10):791–803. https://doi.org/10.1016/j.tcb.2019.07.006
13. Stein KC, Frydman J (2019) The stop-and-go traffic regulating protein biogenesis: How translation kinetics controls proteostasis. J Biol Chem 294(6):2076–2084. https://doi.org/10.1074/jbc.REV118.002814

14. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324 (5924):218–223. https://doi.org/10.1126/science.1168978
15. Steitz JA (1969) Polypeptide chain initiation: nucleotide sequences of the three ribosomal binding sites in bacteriophage R17 RNA. Nature 224(5223):957–964. https://doi.org/10.1038/224957a0
16. Ingolia NT (2016) Ribosome footprint profiling of translation throughout the genome. Cell 165(1):22–33. https://doi.org/10.1016/j.cell.2016.02.066
17. Anders M, Chelysheva I, Goebel I, Trenkner T, Zhou J, Mao Y, Verzini S, Qian SB, Ignatova Z (2018) Dynamic m(6)A methylation facilitates mRNA triaging to stress granules. Life Sci Alliance 1(4):e201800113. https://doi.org/10.26508/lsa.201800113
18. Bartholomaus A, Fedyunin I, Feist P, Sin C, Zhang G, Valleriani A, Ignatova Z (2016) Bacteria differently regulate mRNA abundance to specifically respond to various stresses. Philos Trans Series A, Math Phys Eng Sci 374(2063). https://doi.org/10.1098/rsta.2015.0069
19. Ikeuchi K, Izawa T, Inada T (2018) Recent progress on the molecular mechanism of quality controls induced by ribosome stalling. Front Genet 9:743. https://doi.org/10.3389/fgene.2018.00743
20. Mohammad F, Woolstenhulme CJ, Green R, Buskirk AR (2016) Clarifying the translational pausing landscape in bacteria by ribosome profiling. Cell Rep 14(4):686–694. https://doi.org/10.1016/j.celrep.2015.12.073
21. Avcilar-Kucukgoze I, Bartholomaus A, Cordero Varela JA, Kaml RF, Neubauer P, Budisa N, Ignatova Z (2016) Discharging tRNAs: a tug of war between translation and detoxification in Escherichia coli. Nucleic Acids Res 44(17):8324–8334. https://doi.org/10.1093/nar/gkw697
22. Kirchner S, Cai Z, Rauscher R, Kastelic N, Anding M, Czech A, Kleizen B, Ostedgaard LS, Braakman I, Sheppard DN, Ignatova Z (2017) Alteration of protein function by a silent polymorphism linked to tRNA abundance. PLoS Biol 15(5):e2000779. https://doi.org/10.1371/journal.pbio.2000779
23. Dunn JG, Foo CK, Belletier NG, Gavis ER, Weissman JS (2013) Ribosome profiling reveals pervasive and regulated stop codon readthrough in Drosophila melanogaster. eLife 2:e01179. https://doi.org/10.7554/eLife.01179
24. Wang H, Yang L, Wang Y, Chen L, Li H, Xie Z (2019) RPFdb v2.0: an updated database for genome-wide information of translated mRNA generated from ribosome profiling. Nucleic Acids Res 47(D1):D230–D234. https://doi.org/10.1093/nar/gky978
25. Calviello L, Mukherjee N, Wyler E, Zauber H, Hirsekorn A, Selbach M, Landthaler M, Obermayer B, Ohler U (2016) Detecting actively translated open reading frames in ribosome profiling data. Nat Methods 13 (2):165–170. https://doi.org/10.1038/nmeth.3688
26. Fields AP, Rodriguez EH, Jovanovic M, Stern-Ginossar N, Haas BJ, Mertins P, Raychowdhury R, Hacohen N, Carr SA, Ingolia NT, Regev A, Weissman JS (2015) A regression-based analysis of ribosome-profiling data reveals a conserved complexity to mammalian translation. Mol Cell 60(5):816–827. https://doi.org/10.1016/j.molcel.2015.11.013
27. Olshen AB, Hsieh AC, Stumpf CR, Olshen RA, Ruggero D, Taylor BS (2013) Assessing gene-level translational control from ribosome profiling. Bioinformatics 29(23):2995–3002. https://doi.org/10.1093/bioinformatics/btt533
28. Dunn JG, Weissman JS (2016) Plastid: nucleotide-resolution analysis of next-generation sequencing and genomics data. BMC Genomics 17(1):958. https://doi.org/10.1186/s12864-016-3278-x
29. Lauria F, Tebaldi T, Bernabo P, Groen EJN, Gillingwater TH, Viero G (2018) riboWaltz: optimization of ribosome P-site positioning in ribosome profiling data. PLoS Comput Biol 14 (8):e1006169. https://doi.org/10.1371/journal.pcbi.1006169
30. Popa A, Lebrigand K, Paquet A, Nottet N, Robbe-Sermesant K, Waldmann R, Barbry P (2016) RiboProfiling: a bioconductor package for standard Ribo-seq pipeline processing. F1000Res 5:1309. https://doi.org/10.12688/f1000research.8964.1
31. Michel AM, Mullan JP, Velayudhan V, O'Connor PB, Donohue CA, Baranov PV (2016) RiboGalaxy: a browser based platform for the alignment, analysis and visualization of ribosome profiling data. RNA Biol 13 (3):316–319. https://doi.org/10.1080/15476286.2016.1141862
32. Gerashchenko MV, Gladyshev VN (2017) Ribonuclease selection for ribosome profiling. Nucleic Acids Res 45(2):e6. https://doi.org/10.1093/nar/gkw822
33. O'Connor PB, Andreev DE, Baranov PV (2016) Comparative survey of the relative impact of mRNA features on local ribosome profiling read density. Nat Commun 7:12915. https://doi.org/10.1038/ncomms12915

34. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26(6):841–842. https://doi.org/10.1093/bioinformatics/btq033
35. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25 (16):2078–2079. https://doi.org/10.1093/bioinformatics/btp352
36. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal 17(1). https://doi.org/10.14806/ej.17.1.200
37. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10(3):R25. https://doi.org/10.1186/gb-2009-10-3-r25
38. Mohammad F, Green R, Buskirk AR (2019) A systematically-revised ribosome profiling method for bacteria reveals pauses at single-codon resolution. Elife 8. https://doi.org/10.7554/eLife.42591
39. Bartholomaus A, Del Campo C, Ignatova Z (2016) Mapping the non-standardized biases of ribosome profiling. Biol Chem 397 (1):23–35. https://doi.org/10.1515/hsz-2015-0197
40. Wu CC, Zinshteyn B, Wehner KA, Green R (2019) High-resolution ribosome profiling defines discrete ribosome elongation states and translational regulation during cellular stress. Mol Cell 73(5):959–970. e955. https://doi.org/10.1016/j.molcel.2018.12.009
41. Guydosh NR, Green R (2017) Translation of poly(A) tails leads to precise mRNA cleavage. RNA 23(5):749–761. https://doi.org/10.1261/rna.060418.116
42. Woolstenhulme CJ, Guydosh NR, Green R, Buskirk AR (2015) High-precision analysis of translational pausing by ribosome profiling in bacteria lacking EFP. Cell Rep 11(1):13–21. https://doi.org/10.1016/j.celrep.2015.03.014
43. Del Campo C, Bartholomaus A, Fedyunin I, Ignatova Z (2015) Secondary structure across the bacterial transcriptome reveals versatile roles in mRNA regulation and function. PLoS Genet 11(10):e1005613. https://doi.org/10.1371/journal.pgen.1005613

Chapter 13

Assessing Ribosome Distribution Along Transcripts with Polarity Scores and Regression Slope Estimates

Ilya E. Vorontsov, Artyom A. Egorov, Aleksandra S. Anisimova, Irina A. Eliseeva, Vsevolod J. Makeev, Vadim N. Gladyshev, Sergey E. Dmitriev, and Ivan V. Kulakovskiy

Abstract

During translation, the rate of ribosome movement along mRNA varies. This leads to a non-uniform ribosome distribution along the transcript, depending on local mRNA sequence, structure, tRNA availability, and translation factor abundance, as well as the relationship between the overall rates of initiation, elongation, and termination. Stress, antibiotics, and genetic perturbations affecting composition and properties of translation machinery can alter the ribosome positional distribution dramatically. Here, we offer a computational protocol for analyzing positional distribution profiles using ribosome profiling (Ribo-Seq) data. The protocol uses *papolarity*, a new Python toolkit for the analysis of transcript-level short read coverage profiles. For a single sample, for each transcript *papolarity* allows for computing the classic polarity metric which, in the case of Ribo-Seq, reflects ribosome positional preferences. For comparison versus a control sample, *papolarity* estimates an improved metric, the relative linear regression slope of coverage along transcript length. This involves de-noising by profile segmentation with a Poisson model and aggregation of Ribo-Seq coverage within segments, thus achieving reliable estimates of the regression slope. The *papolarity* software and the associated protocol can be conveniently used for Ribo-Seq data analysis in the command-line Linux environment. *Papolarity* package is available through Python *pip* package manager. The source code is available at https://github.com/autosome-ru/papolarity.

Key words Ribo-Seq, Ribosome footprint density, Ribosome footprint coverage, Ribosome distribution, Polarity, Segmentation, Linear regression

Abbreviations

CDS	Coding segments
CH	Cycloheximide
GEO	Gene Expression Omnibus
HR	Harringtonine
mESCs	Mouse embryonic stem cells
Ribo-Seq	Ribosome profiling

Vyacheslav M. Labunskyy (ed.), *Ribosome Profiling: Methods and Protocols*, Methods in Molecular Biology, vol. 2252, https://doi.org/10.1007/978-1-0716-1150-0_13,

1 Introduction

Ribosome profiling (Ribo-Seq) radically expanded an arsenal of high-throughput sequencing techniques in the field of protein synthesis research [1]. Ribo-Seq not only allows quantitative and qualitative characterization of the whole-cell translatome in specific conditions and at particular time points but also reveals novel molecular mechanisms of protein synthesis and translational control [2, 3]. In particular, it allows tracking changes in ribosome distribution along mRNA coding regions.

Ribosome motion along a transcript can sometimes slow down or even cease, leading to ribosome queuing. This can occur at some specific internal sites within the coding region (e.g., clusters of rare codons, damaged nucleotides, extremely stable secondary structures, or regions encoding stretches of proline or positively charged amino acids) [4]. Ribosome pausing also occurs at the beginning and end of the coding region, due to switching the mode of ribosome activity from initiation to elongation and from elongation to termination and recycling, respectively. Thus, ribosome footprints are usually enriched at start and stop codons of many genes, which may vary in intensity depending on mRNA, genetic background, and experimental conditions, such as the use of antibiotics [5–12]. Moreover, it has been shown that the first ~30–50 codons are translated at a low speed in many organisms due to a higher density of codons with low tRNA adaptation index [13]. It was suggested that this region serves as a "ramp" for the starting ribosome to reduce ribosomal traffic jams. As a result, ribosome occupancy in different regions of a transcript is not uniform, forming a gradient of footprint density along the coding region.

Importantly, this uneven ribosome distribution is condition-specific. For example, heat shock or drug-induced proteotoxic stresses trigger elongation pausing after translation of ~50–65 codons in mammalian cells, leading to an increased ribosome coverage of the corresponding mRNA region [8, 14]. In yeast, depletion of certain translation elongation factors shifts overall ribosome density toward 5′ ends of the coding regions [15, 16], while depletion of ribosomal proteins RPLP1/2 similarly redistributes the ribosomes in human cells [17]. In mice, knockout of the *Fmr1* gene removes the translational pausing across a large number of genes [18]. In mycobacteria, specialized ribosomes with RpsR2 are assembled upon Zn depletion and exhibit altered codon usage causing a clear polarity shift toward 5′ ends of transcripts [19].

For a quantitative analysis of footprint distribution along the transcript length, several metrics and specialized approaches have been proposed. In some cases, relatively simple metrics were used, such as the 5′ loading ratio [14], the ribosome pausing index, and the asymmetry score [20], which are calculated as the normalized

footprint density within certain segments of the transcript coding regions, e.g., near the start or stop codons or globally in the 5′ and 3′ halves of transcript. Changes in such metrics reflect the extent to which the ribosomes are preferentially accumulated or depleted at the ends of coding segments (CDS).

Later, the so-called polarity score was introduced as a whole-transcript metric to reveal ribosome positional distribution on the mRNA [16]. Ribosome polarity is analogous to a center of mass of the coverage profile [21] and assigns a value between −1 and 1 to each transcript, with positive and negative scores reflecting relative footprint enrichment at the 5′ or 3′ end of CDS, respectively. To compare transcriptome-level changes between two samples, standard statistical methods such as Wilcoxon signed-rank test can be used to assess the differences in polarity score distributions across transcriptome [15].

Simplicity of the polarity score makes it useful as a starting point for estimating changes in ribosome positional preferences from Ribo-Seq data. However, there are no mature command-line tools to compute positional metrics such as polarity from the read alignments. At the same time, in general, R framework of the Ribo-Seq analysis tools is very rich [22], whereas the Python repertoire of production-ready tools is much weaker [22], with only a few notable exceptions such as *plastid* [23].

Here we describe a protocol and a Python toolkit *papolarity* designed to estimate ribosome positional profiles from transcriptomic read alignments. *Papolarity* is not limited to computing only the standard polarity score. It also provides a more complex and reliable estimate of a relative positional footprint density in two samples (e.g., to compare a sample against its control), utilizing optimal segmentation of coverage profiles with a Poisson model. This approach is advantageous over basic polarity estimation, since it allows correcting for non-uniform footprint coverage arising from read mapping peculiarities and transcript-specific ribosome stalling, by using information from multiple available samples.

A conceptual overview of the protocol is given in Fig. 1. The first step is the preprocessing of raw sequencing data including read trimming and alignment to a reference transcriptome. The second step involves the *papolarity* toolkit, which uses read alignments and genome annotation. In the simplest scenario, it estimates polarity scores for each transcript in each Ribo-Seq sample independently. A comparative scenario for two or more samples requires several substeps.

(1) Footprint coverage profiles are pooled, and the resulting pooled profile is used for segmentation with a Poisson model with the *pasio* Python package. For each transcript, it allows obtaining an optimal partition of segments, or sequence windows, in which the footprint coverage can be considered as a result of a series of Poisson trials with a fixed Poisson parameter lambda. (2) The

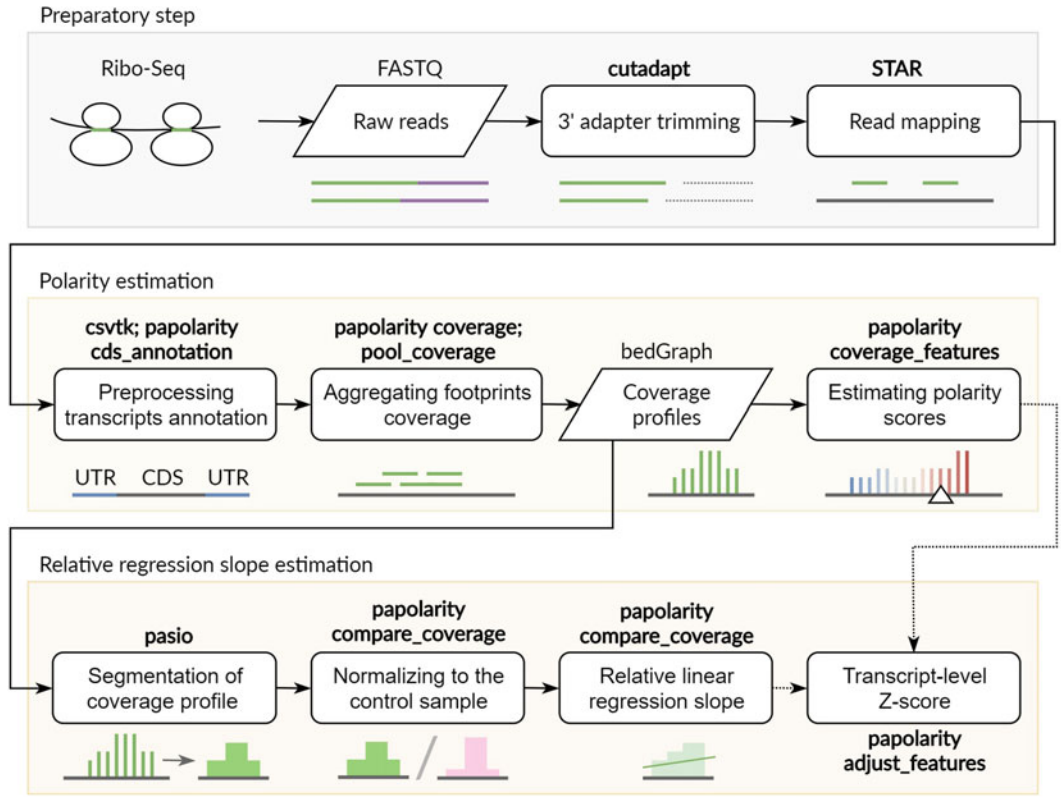

Fig. 1 A conceptual overview of the protocol

resulting segmentation is used to flatten profiles of each Ribo-Seq sample, i.e., to simplify profiles by averaging single-nucleotide resolution values within non-overlapping windows defined by optimal segments. (3) The flattened profile for each transcript is normalized to the transcript read coverage. Next, to account for ribosome pausing and technical artifacts, each profile is normalized to that of a control sample (obtaining either ratios or $\log_2$ ratios). (4) The normalized profiles are used to estimate the linear regression slope, reflecting changes in positional preference of ribosomes relative to the control samples.

For both polarity and relative linear regression slope, *papolarity* allows for an additional step of estimating transcript-level Z-scores by using mean and variance estimated for transcripts of comparable lengths. By doing so, it is possible to identify particular transcripts demonstrating extreme polarity scores or, in case of two-sample comparison, regression slopes.

The described procedure facilitates analysis of positional changes in ribosomal coverage, providing a convenient and useful facet to the assessment of ribosome profiling data. We demonstrate the described protocol in action using data from the classical study

by Ingolia et al. [5]. The authors monitored kinetics of in vivo translation in mouse embryonic stem cells (mESCs) by tracing run-off elongation. For this purpose, they first arrested de novo initiating ribosomes by harringtonine (HR) and then, at different time points, halted all elongating ribosomes by cycloheximide (CH). The result was a series of data sets characterized by the progressive 5′–3′ depletion of ribosomes causing systematic changes in footprint distributions and thus allowing evaluation and demonstration of transcript-level changes in positional densities.

2 Materials

This section describes the computing hardware and software requirements, the necessary steps to set up the software tools, and the test data preprocessing. As the input data, the *papolarity* protocol requires transcriptome alignments in BAM format and the corresponding GTF file (preferably the same as used to obtain the transcriptome alignments).

2.1 Overview of Test Data

The test data of Ingolia et al. [5] was extracted from the Gene Expression Omnibus (accession number GSE30839). The following samples (Table 1) were used in the analysis; the respective raw sequencing data were downloaded from the NCBI Short Read Archive.

2.2 Software Requirements

The protocol should be executable under any modern 64-bit Linux with BASH shell; *see* **Note 1** for details on hardware requirements. The following convention is used throughout the text. Software and toolkit names are given in *italic*. Commands, file or folder names, and parts of shell commands are given in the text in `monospace` font. Executable BASH shell commands are given in separate paragraphs written in `monospace` font. The commands presented in the protocol often do not fit a single line of code and visually formatted with automatic line breaks (*see* **Note 2**).

Basic knowledge of BASH would be helpful for a reader to understand the commands in detail. Throughout the protocol, BASH scripting is used to automate the analysis for multiple samples. All commands are presented in a way they should be run under the selected working folder containing the `fastq` folder with the raw sequencing reads in compressed (gzipped) FASTQ format (`fastq.gz`).

2.2.1 Setting Up Necessary Tools and Packages

A modern Linux distribution, especially running on a bioinformatics-oriented machine, may already have most of the necessary tools preinstalled. However, here the *conda* environment management system is used to ensure the required tools are

Table 1
Overview of the test data used for the *papolarity* protocol description and demonstration

Sample ID	Harringtonine treatment	Cycloheximide treatment	GEO ID (SRA IDs)
ES_noHR_noCH_ribo	None	None	GSM765298 (SRR315616, SRR315617, SRR315618, SRR315619)
ES_noHR_60sCH_ribo	None	100 μg/mL, 60 s	GSM765292 (SRR315601, SRR315602)
ES_90sHR_60sCH_ribo	1 μg/mL, 90 s	100 μg/mL, 60 s	GSM765297 (SRR315612, SRR315613, SRR315614, SRR315615)
ES_120sHR_60sCH_ribo	1 μg/mL, 120 s	100 μg/mL, 60 s	GSM765294 (SRR315604, SRR315605, SRR315606)
ES_150sHR_60sCH_ribo	1 μg/mL, 150 s	100 μg/mL, 60 s	GSM765295 (SRR315607, SRR315608, SRR315609)
ES_180sHR_60sCH_ribo	0.5 μg/mL, 180 s	100 μg/mL, 60 s	GSM765296 (SRR315610, SRR315611)

Table 2
List of software tools and packages used in the protocol

Software	Version
conda	4.7.12
sra-tools[a]	2.8.0
bedtools	2.29.2
samtools	1.9
cutadapt	2.7
STAR	2.7.3a
GNU parallel	20191122
csvtk	0.19.1
Python package	**Version**
pasio	1.1.2
papolarity	1.0.0

[a]*sra-tools* is necessary only to download and prepare the test data

successfully installed on a machine without root privileges and do not conflict with the system-level tools and packages. Note that `gzip` is necessary at the preparatory stage of the protocol and should be available in the host system (not installed with *conda*). To download the minimal *miniconda* installer, `wget` and `gzip` should be preinstalled in the host operating system. The protocol was tested with the following versions of the tools (Table 2); *see* also **Note 3**.

2.2.2 Setting Up Conda Environment

First, setup the *conda* package and environment management system:

```
wget https://repo.anaconda.com/miniconda/Miniconda3-latest-
Linux-x86_64.sh;
chmod +x Miniconda3-latest-Linux-x86_64.sh;
./Miniconda3-latest-Linux-x86_64.sh;
./miniconda3/bin/conda init bash;
```

Upon completion, restart your BASH session (e.g., re-login to the machine). Next, create and activate the `papolarityenv` environment. While creating the `papolarityenv`, Python 3.7 is specified to be available in the environment. The activation step is necessary not only to install the necessary tools but also later to run any part of the protocol after re-login:

```
conda create --yes --name papolarityenv python=3.7;
conda activate papolarityenv;
```

Note that changing the environment name would not affect the following protocol, except for the environment activation step.

Finally, install the basic software (*GNU parallel*), the bioinformatics software from the `bioconda` channel, and necessary Python packages (with `pip`):

```
conda install --yes -c conda-forge parallel;
conda install --yes -c bioconda star bedtools samtools csvtk
sra-tools;
pip install cutadapt pasio papolarity;
```

2.3 Preparing Test Data

The protocol is applicable for the analysis of arbitrary Ribo-Seq data set by adapting file names in the commands, where necessary. It is also possible to exactly reproduce the protocol execution on the test data. To this end, the raw data from SRA should be downloaded from GEO with `prefetch` and converted to FASTQ format with `fastq-dump` (both programs from *sra-tools* [24]). The resulting FASTQ files can be merged with the basic Linux `cat` resulting in a single gzipped FASTQ file per sample. The initial six compressed `*.fastq.gz` files (named accordingly to the respective sample names; see Sample ID in Table 1) should be placed to the `fastq` subfolder of the selected working folder (which should be constant for the whole protocol execution). Note that for the described type of analysis, the complementary RNA-Seq data is not necessary. For exact commands to reproduce the test data preparation, *see* **Note 4**.

2.4 Ribo-Seq Read Processing

Analysis of ribosome positional densities requires information regarding locations of ribosome footprints along transcripts. In turn, this entails standard preprocessing, consisting of read trimming and mapping. Note that during this stage, program parameters were set to utilize 32 computational threads to speed up the process (the respective value can be as low as 1 in case a particular machine does not allow multi-core processing, resulting in linearly decreased performance); *see* **Note 5**. Since the details of Ribo-Seq data preprocessing do not fit the scope of this protocol, we do not elaborate on the strategy, tool selection, and parameter settings used at the preprocessing stage; *see* **Note 6** for discussion.

2.4.1 Preparatory Step

To begin preprocessing, mouse genome sequence and transcript annotation should be downloaded, unpacked, and indexed for short-read mapping. First, create a new `genome` folder, download, and unpack the genome and genome annotation (this example uses the latest mouse GENCODE build of 2019):

```
mkdir ./genome; cd ./genome;
wget ftp://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_mouse/
```

```
release_M23/gencode.vM23.basic.annotation.gtf.gz;
wget ftp://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_mouse/
release_M23/GRCm38.primary_assembly.genome.fa.gz;
find . -type f -name '*.gz' -exec bash -c 'gzip -d {}' \;
```

Next, index the genome and annotation with *STAR* aligner, which will be then used for read mapping, and exit the `genome` folder:

```
STAR --runMode genomeGenerate --genomeDir . --genomeFastaFiles
./GRCm38.primary_assembly.genome.fa --sjdbGTFfile ./gencode.
vM23.basic.annotation.gtf --runThreadN 32;
cd ..;
```

As a result, the `genome` folder will contain the genome assembly `GRCm38.primary_assembly.genome.fa`, transcript annotation `gencode.vM23.basic.annotation.gtf`, and genome index of *STAR* (multiple files).

2.4.2 Trimming and Alignment

First, perform 3′ adapter trimming with *cutadapt* using the adapter sequences as specified in [5]:

```
mkdir ./trim;
find ./fastq/*.fastq.gz -maxdepth 1 -type f -exec bash -c 'bn=
$(basename {}); cutadapt -a CTGTAGGCACCATCAAT -j 32 --minimum-
length 20 -q 20 --trimmed-only {} -o trim/$bn' \; > cutadapt.
log;
```

As a result, the `trim` folder will contain FASTQ files with quality- and adapter-trimmed reads in compressed `fastq.gz` files with the ribosome footprints, one file per sample. The `cutadapt.log` can be checked for trimming statistics and possible execution errors.

Next, create separate folders for *STAR* alignments of reads of individual samples, and perform read mapping:

```
mkdir align;
find ./fastq/*.fastq.gz -maxdepth 1 -type f -exec bash -c 's=
$(basename {});bn=${s%.fastq.gz}; mkdir align/$bn' \; ;
find ./trim/*.fastq.gz -maxdepth 1 -type f -exec bash -c 's=
$(basename {});bn=${s%.fastq.gz}; STAR --genomeDir ./genome/
--readFilesCommand zcat --quantMode TranscriptomeSAM
--readFilesIn {} --outFileNamePrefix ./align/$bn/
--alignSJDBoverhangMin 1 --runThreadN 32 --outSAMtype BAM
Unsorted' \; ;
```

With the specified settings, *STAR* provides both genomic (`Aligned.out.bam`) and transcriptomic (`Aligned.`

toTranscriptome.out.bam) alignments for each sample. The transcriptomic alignments will enter the consequent analysis.

Finally, filter uniquely mapped reads, and perform sorting and indexing of alignment BAM files:

```
find ./trim/*.fastq.gz -maxdepth 1 -type f -exec bash -c 's=
$(basename {});bn=${s%.fastq.gz}; samtools view -b -q 255 ./
align/$bn/Aligned.toTranscriptome.out.bam | samtools sort - >
./align/$bn.bam ' \; ;
find ./align/*.bam -maxdepth 1 -type f -exec samtools index {}
\; ;
```

As a result, the align folder will contain six alignments (*.bam) and six corresponding BAM index (*.bai) files, a pair per sample, named according to sample identifiers. When processing your own data, please note that *papolarity* requires the alignment files to be sorted and indexed (e.g., with *samtools* as shown above).

3 Methods

Here the protocol is described step by step, first computing the basic polarity score and then moving to a more complex assessment involving segmentation of profiles and estimation of the relative linear regression slope. Many commands in this following protocol are wrapped in a loop to perform the analysis of multiple samples; *see* **Note 5** for details regarding parallel and sequential execution.

3.1 Basic Preprocessing

This section describes the common starting steps of the analysis, necessary for both the polarity score and the relative regression slope estimation.

3.1.1 Preparing Transcript Annotation

Some of the further steps require particular details from the transcript annotation (*see* **Note 7** for requirements related to GTF structure), such as CDS locations. First, extract the transcript CDS annotations (to be saved in the tab-separated plain text file genome/cds_features.tsv):

```
papolarity cds_annotation ./genome/gencode.vM23.basic.
annotation.gtf --attr-filter transcript_type=protein_coding
--attr-filter gene_type=protein_coding --output-file
./genome/cds_features.tsv;
```

Of note, at this stage, the list is filtered to include only protein-coding transcripts. Second, it is used to derive additional information, which will be then utilized on the following steps.

Estimate the transcript and CDS lengths (to be saved in genome/transcript_lengths.tsv):

```
csvtk --tabs cut genome/cds_features.tsv --fields 'transcript_id,
transcript_length,cds_length' --out-file genome/transcript_
lengths.tsv
```

Generate the mapping of transcript and gene identifiers (to be saved in `genome/transcript2gene.tsv`):

```
csvtk --tabs cut genome/cds_features.tsv --fields 'transcript_id,
gene_id' --out-file genome/transcript2gene.tsv
```

3.1.2 Preparing Coverage Profiles

The `*.bam` alignments of all samples should be converted to bedGraph coverage profiles with the following script:

```
mkdir coverage;
SAMPLES='ES_noHR_noCH_ribo ES_noHR_60sCH_ribo ES_90sHR_60sCH_
ribo ES_120sHR_60sCH_ribo ES_150sHR_60sCH_ribo ES_180sHR_
60sCH_ribo';
( for SAMPLE in $SAMPLES; do echo papolarity get_coverage
"./align/${SAMPLE}.bam" --sort --dtype int --output-file
"./coverage/${SAMPLE}.bedgraph.gz"; done ) | parallel;
```

As a result, the `coverage` folder will contain `*.bedgraph.gz` files, one file per sample. Note that the files are gzipped on-the-fly to save disk space. The resulting profiles will not be normalized for their coverage, i.e., they will provide raw footprint coverage values per transcript position obtained from each input alignment file (*see* **Note 8**).

Pooling coverage profiles: Generate the pooled coverage profile across all samples that will be used to filter transcripts by coverage and, later, to obtain the robust segmentation of the coverage profiles (`./coverage/pooled.bedgraph.gz`). Note the `--dtype int` parameter that instructs the tool to retain integer read counts (required for profile segmentation; see Subheading 3.3).

```
papolarity pool_coverage ./coverage/*.bedgraph.gz --dtype int
--output-file ./coverage/pooled.bedgraph.gz;
```

Of note, the same tool optionally can be also helpful for pooling coverage profiles of related samples, e.g., technical replicates. This is not part of the described protocol, but can be useful in other scenarios, e.g.:

```
papolarity pool_coverage sample_r1.bedgraph.gz sample_r2.
bedgraph.gz sample_r3.bedgraph.gz --dtype int --output-file
sample_pooled.bedgraph.gz;
```

Note that the command above is not directly executable on the test data and is provided only as an illustrative template.

Clipping profiles to coding segments: To assess ribosome density within coding segments only, generate clipped variants of the coverage profiles. The command below allows for excluding untranslated regions and, additionally, 30 nucleotides neighboring start and stop codons (`--drop-5-flank 30 --drop-3-flank 30`) of each transcript with the annotated CDS. Additional clipping of 30nts from each side allows excluding bias from the start and stop codons and neighboring regions (as footprint coverage on start and stop codons may be specifically dependent on experimental conditions and change in a different fashion in comparison to the CDS).

```
mkdir cds_coverage;
SAMPLES='ES_noHR_noCH_ribo ES_noHR_60sCH_ribo ES_90sHR_
60sCH_ribo ES_120sHR_60sCH_ribo ES_150sHR_60sCH_ribo
ES_180sHR_60sCH_ribo';
( for SAMPLE in $SAMPLES 'pooled'; do echo papolarity clip_cds
./genome/cds_features.tsv "./coverage/${SAMPLE}.bedgraph.gz"
--drop-5-flank 30 --drop-3-flank 30 --contig-naming original
--output-file "./cds_coverage/${SAMPLE}.bedgraph.gz"; done ) |
parallel;
```

The results will be saved to `./cds_coverage/*.bedgraph.gz` files with the file names as the initial sample names, with an additional profile for pooled data.

The transcripts without annotated CDSs are ignored and excluded at this step. Note that the resulting clipped bedGraph-files will have a shifted coordinate system where zero is the start of the retained region.

3.2 Polarity Score

3.2.1 Estimating the Scores

At this stage, it is possible to specify unique prefix when generating the resulting files so the column names will be distinguishable between different samples:

```
SAMPLES='ES_noHR_noCH_ribo ES_noHR_60sCH_ribo ES_90sHR_60sCH_
ribo ES_120sHR_60sCH_ribo ES_150sHR_60sCH_ribo ES_180sHR_
60sCH_ribo';
mkdir -p ./coverage_features/raw;
( for SAMPLE in $SAMPLES 'pooled'; do echo papolarity coverage_
features "./cds_coverage/${SAMPLE}.bedgraph.gz" --prefix
"${SAMPLE}_" --output-file "./coverage_features/raw/
${SAMPLE}.tsv"; done ) | parallel;
```

The results will be saved to `./coverage_features/raw/*.tsv` files with the file names as the initial sample names.

Note that the polarity score of complete transcripts (instead of clipped CDS) can be estimated by using files from the `./coverage/` folder (instead of the `./cds_coverage/` folder used above).

3.2.2 Transcript List Filtering

The estimates of polarity scores for lowly expressed genes might be non-accurate and noisy. This step allows obtaining the list of reliably expressed transcripts and, additionally, selecting a single major transcript isoform per gene.

First, select transcripts with mean pooled coverage no less than 1 and having the third quartile (75th percentile) of the coverage values distribution greater than 0; *see* **Note 9**. This is performed with `csvtk` (result saved to `coverage_features/pooled/pooled.filtered_1.tsv`):

```
mkdir -p ./coverage_features/pooled;
csvtk --tabs filter2 "coverage_features/raw/pooled.tsv"
--filter '($pooled_mean_coverage >= 1) && ($pooled_coverage_q75 >
0)' --out-file "coverage_features/pooled/pooled.filtered_1.
tsv";
```

Second, add gene identifiers and transcript lengths (result saved to `coverage_features/pooled/pooled.filtered_1.with_gene_id.tsv`):

```
csvtk --tabs join --fields transcript_id "coverage_features/
pooled/pooled.filtered_1.tsv" "genome/transcript2gene.tsv"
"genome/transcript_lengths.tsv" --out-file "coverage_
features/pooled/pooled.filtered_1.with_gene_id.tsv";
```

Third, as a non-mandatory step, select a single major (with the highest coverage) transcript per gene (result saved to `coverage_features/pooled/pooled.filtered_2.tsv`):

```
papolarity choose_best "coverage_features/pooled/pooled.
filtered_1.with_gene_id.tsv" pooled_mean_coverage max
--group-by gene_id --header --output-file "coverage_features/
pooled/pooled.filtered_2.tsv";
```

Finally, select the column with the transcript IDs and supplementary transcript lengths, and save it as the final filtered transcript list (`transcripts_list.tsv`):

```
csvtk --tabs cut "coverage_features/pooled/pooled.filtered_2.
tsv" --fields transcript_id,transcript_length,cds_length
--out-file ./transcripts_list.tsv;
```

3.2.3 Finalizing Polarity Score Lists

By using the transcript list from the previous step, filter the polarity score lists, and concatenate the transcript length information, which will be used on the next step to estimate polarity Z-scores (results saved to `./coverage_features/filtered/*.tsv`):

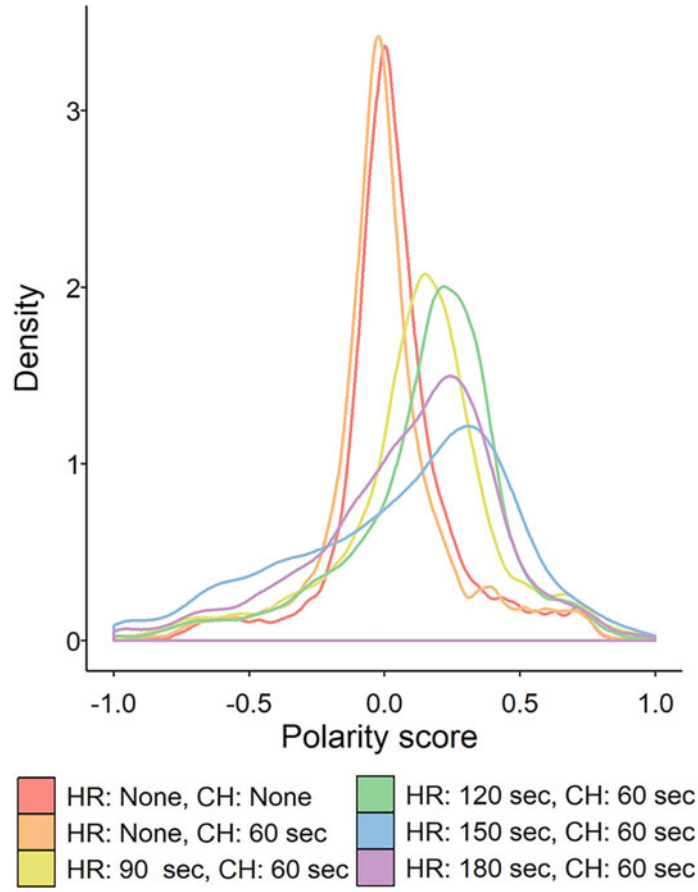

Fig. 2 Distribution of polarity scores estimated for different samples of the test data. X axis: polarity score value. Y axis: distribution density; smoothed density estimate is shown for clarity. Different colors correspond to different samples (as shown in the figure legend)

```
SAMPLES='ES_noHR_noCH_ribo ES_noHR_60sCH_ribo ES_90sHR_
60sCH_ribo ES_120sHR_60sCH_ribo ES_150sHR_60sCH_ribo
ES_180sHR_60sCH_ribo';
mkdir -p ./coverage_features/filtered;
for SAMPLE in $SAMPLES; do csvtk --tabs join ./transcripts_
list.tsv "./coverage_features/raw/${SAMPLE}.tsv" --out-file
"./coverage_features/filtered/${SAMPLE}.tsv"; done;
```

The resulting distributions of polarity scores are visualized in Fig. 2, where the change of distribution shape becomes more exhibited upon longer HR treatment, except for 180 sec. timepoint with a lower HR concentration (*see* Table 1).

3.2.4 Polarity Z-Score Estimation

The mean and variance in polarity score values depend on the transcript lengths. This does not strongly affect the comparison of polarity score distributions between different samples (which can be done with a Wilcoxon signed-rank test) but hinders the identification of particular genes with unusual or extreme polarity scores. To resolve this issue, compute the polarity Z-score of each transcript by estimating the expected mean and variance from polarity scores of transcripts of similar lengths. To this end, sort transcripts by CDS length (or total length, depending on clipping strategy). Then, apply *papolarity* to traverse the list and estimate mean and variance from 500 transcripts in a sliding window centered on each transcript under consideration (results saved in `./coverage_features/adjusted/*.tsv`):

```
SAMPLES='ES_noHR_noCH_ribo ES_noHR_60sCH_ribo ES_90sHR_
60sCH_ribo ES_120sHR_60sCH_ribo ES_150sHR_60sCH_ribo
ES_180sHR_60sCH_ribo';
mkdir -p ./coverage_features/adjusted;
for SAMPLE in $SAMPLES; do papolarity adjust_features
"./coverage_features/filtered/${SAMPLE}.tsv" --sort-field
'cds_length' --fields "${SAMPLE}_polarity" --mode z-score
--window 500 --prefix 'zscore_' --output-file "./coverage_
features/adjusted/${SAMPLE}.tsv"; done;
```

3.2.5 Visualizing Per-Sample Score Distribution

For convenience, *papolarity* allows generating draft plots for polarity score distribution separately for samples (results saved as `coverage_features/plot/*.png`):

```
SAMPLES='ES_noHR_noCH_ribo ES_noHR_60sCH_ribo ES_90sHR_60s
CH_ribo ES_120sHR_60sCH_ribo ES_150sHR_60sCH_ribo
ES_180sHR_60sCH_ribo';
mkdir -p ./coverage_features/plot/;
for SAMPLE in $SAMPLES; do papolarity plot_distribution
"coverage_features/filtered/${SAMPLE}.tsv" --fields "${SAMPLE}_
polarity" --no-legend --title "${SAMPLE} polarity distribution"
--zero-line green --xlim -1.0 1.0 --output-file "coverage_
features/plot/${SAMPLE}.png"; done;
```

3.3 Segmentation and Relative Slope Estimation

Estimation of relative slopes requires transcript coverage data and the filtered transcript list obtained upon successful completion of the first two steps of the polarity estimation protocol.

3.3.1 Simplifying Coverage Profiles

To reliably estimate the relative regression slope when comparing a pair of samples, start from de-noising and simplifying the coverage profiles by optimal segmentation with a Poisson model (results saved as `segmentation.bed.gz`):

```
pasio ./coverage/pooled.bedgraph.gz --alpha 1 --beta 1
--output-file ./segmentation.bed.gz --output-mode bed;
```

Here the *pasio* Python package is used with default parameters; *see* **Note 10**.

3.3.2 Clipping Segmentation to Coding Segments

In case the CDS clipping was applied for coverage estimates at the previous steps (see the starting steps of the polarity score estimation), the same procedure should be applied to segmentation (result saved to `cds_segmentation.bed.gz`):

```
papolarity clip_cds ./genome/cds_features.tsv ./segmentation.
bed.gz --drop-5-flank 30 --drop-3-flank 30 --contig-naming
original --output-file ./cds_segmentation.bed.gz;
```

3.3.3 Generation of Flattened Coverage Profiles

Of note, this is a non-mandatory step. During the next steps, the segmentation is used by *papolarity* internally, without creating additional intermediate files. However, it might be useful to flatten the coverage profiles according to the segmentation for visualization purposes or for external usage. This can be done with the following command (result saved to `./cds_coverage_flattened/*.bedgraph.gz`):

```
SAMPLES='ES_noHR_noCH_ribo ES_noHR_60sCH_ribo ES_90sHR_60s
CH_ribo ES_120sHR_60sCH_ribo ES_150sHR_60sCH_ribo
ES_180sHR_60sCH_ribo';
mkdir cds_coverage_flattened;
( for SAMPLE in $SAMPLES 'pooled'; do echo papolarity
flatten_coverage ./cds_segmentation.bed.gz "./cds_coverage/
${SAMPLE}.bedgraph.gz" --only-matching --output-file "./
cds_coverage_flattened/${SAMPLE}.bedgraph.gz"; done ) |
parallel;
```

This command flattens CDS-clipped coverage profiles, but the same procedure can be applied to initial whole-transcript profiles as well. To illustrate the segmentation results, *svist4get* [25] was used to visualize original and flattened CDS-clipped profiles (Fig. 3).

3.3.4 Estimating Segmentation-Based Comparative Metrics

The proposed procedure compares positional profiles against a control sample by estimating a linear regression slope along transcript from a relative normalized profile.

In some experimental setups, the direct control data might be non-available or have low coverage, for various reasons. In this case, it is possible to use the complete profile pooled across samples or an average profile instead. In this protocol demonstration, `ES_noHR_noCH_ribo` sample is used as the control.

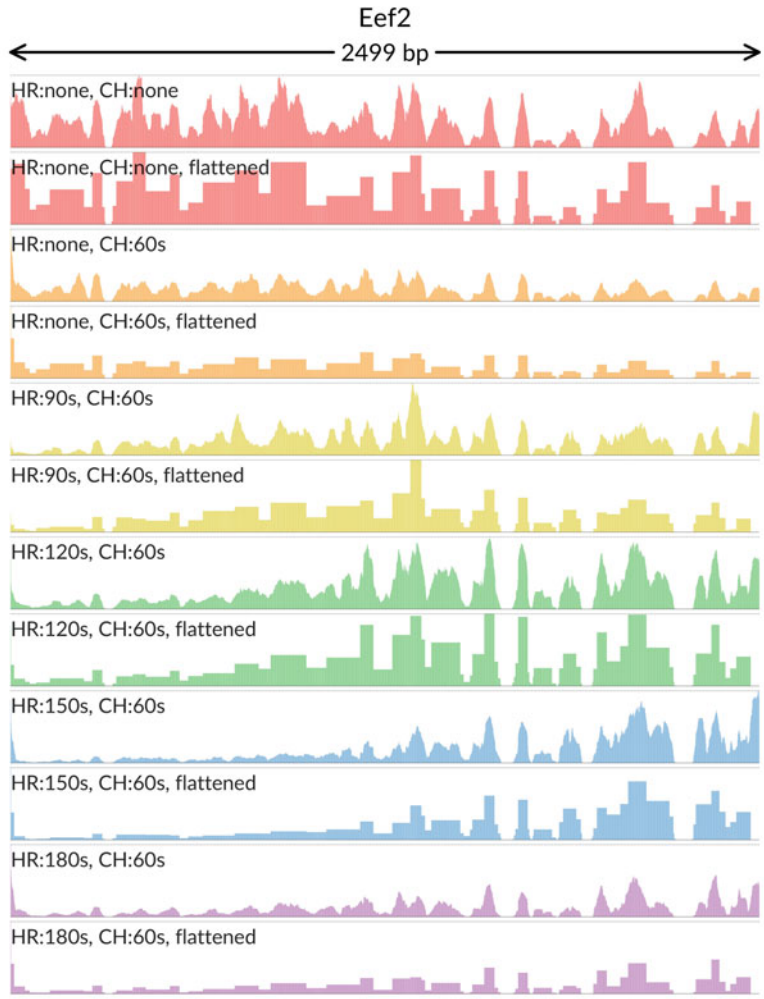

Fig. 3 Original and flattened profiles of ribosome footprint coverage of the major *Eef2* transcript in the test data. Different samples (names shown in track labels) are plotted with different colors (as shown in the figure legend). The plot was generated by svist4get [25]

For each transcript the relative normalized profile is obtained in the following way: (1) flattening the profile of interest (POI) and the control profile on a basis of a given set of segments; (2) normalizing the POI and the control profile to the sum of the respective profile; and (3) computing per-segment ratios or $\log_2$ ratios between the normalized POI and the control profile. The transcript length is then scaled to [0,1] interval, the per-segment relative normalized values are assigned to segment centers, and the respective points are used for linear regression. In addition to `slope` (for ratios) and `slopelog` (for log ratio) estimates, *papolarity* provides

segment-level L1 distance (`l1_distance`) estimated between normalized profiles obtained at the step (2) above.

To properly handle lowly covered transcripts, a "pseudocount" of 1 is added to the total coverage of each segment before profile normalization. However, in the case sample and control coverage values of a particular segment are both zero, the segment is completely excluded from the slope estimation to account for possible coverage gaps induced by read mapping limitations arising in regions of low complexity or in the case of paralogs. Additionally, transcripts for which the first quartile (25th quantile) of segment coverage values is <1 are also excluded.

This algorithm is implemented in `papolarity compare_coverage` which is applied to compare all samples versus the control sample (the results are saved to `comparison/raw/*.tsv`):

```
CONTROL='ES_noHR_noCH_ribo';
EXPERIMENTS ='ES_noHR_60sCH_ribo ES_90sHR_60sCH_ribo
ES_120sHR_60sCH_ribo ES_150sHR_60sCH_ribo ES_180sHR_60sCH_
ribo';
mkdir -p ./comparison/raw;
( for EXPERIMENT in $ EXPERIMENTS ; do echo papolarity
compare_coverage ./cds_segmentation.bed.gz "./cds_coverage/
${CONTROL}.bedgraph.gz" "./cds_coverage/${ EXPERIMENT }.
bedgraph.gz" --segment-coverage-quantile 0.25 1 --prefix
"${EXPERIMENT}_" --output-file "comparison/raw/${EXPERIMENT}.
tsv"; done ) | parallel;
```

An illustration of the slope estimation is shown in Fig. 4.

3.3.5 Filtering Results

The list of transcripts passing coverage filters, and, additionally, selecting only one major transcript per gene, was prepared during the polarity score estimation. This list of transcripts can be used for filtering in the same way as previously (results saved to `comparison/filtered/*.tsv`):

```
EXPERIMENTS ='ES_noHR_60sCH_ribo ES_90sHR_60sCH_ribo
ES_120sHR_60sCH_ribo ES_150sHR_60sCH_ribo ES_180sHR_60sCH_
ribo';
mkdir -p ./comparison/filtered;
for EXPERIMENT in $ EXPERIMENTS ; do csvtk --tabs join ./
transcripts_list.tsv "comparison/raw/${ EXPERIMENT }.tsv"
--out-file "./comparison/filtered/${EXPERIMENT}.tsv"; done;
```

The resulting distributions of `slopelog` scores are visualized in Fig. 5, where the change of distribution shape becomes more exhibited upon longer HR treatment, as previously observed for polarity scores.

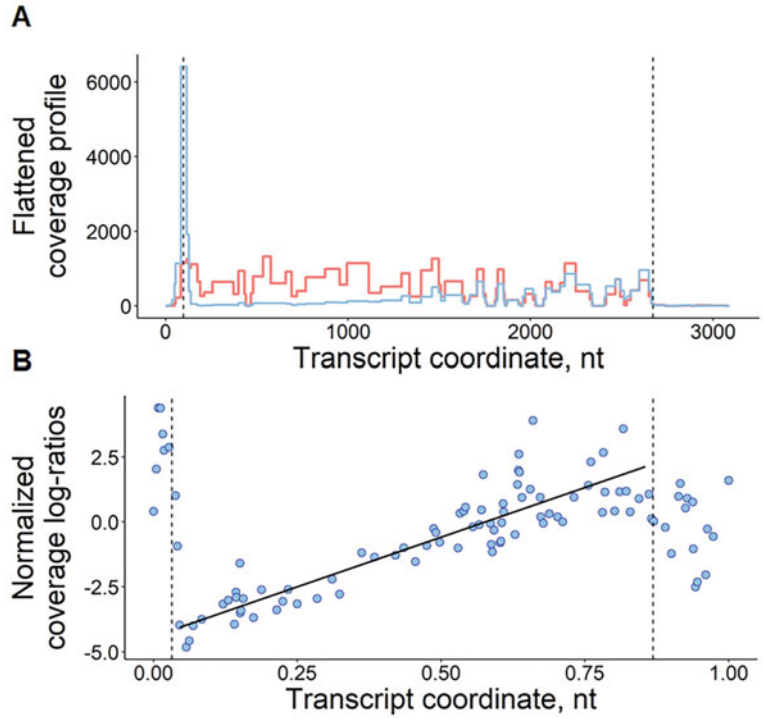

Fig. 4 An illustrative example of relative normalized profile computed for *Eef2* transcript. A. Flattened profile of the selected sample (`ES_150sHR_60sCH_ribo`) and the control profile (`ES_noHR_noCH_ribo`). B. Per-segment $\log_2$ ratios between the normalized POI (`ES_150sHR_60sCH_ribo`) and the control (`ES_noHR_noCH_ribo`) and the linear regression line

3.3.6 Estimating Z-Scores

As for polarity estimates, it is possible to obtain transcript-level Z-scores (results saved to `comparison/adjusted/*.tsv`):

```
EXPERIMENTS ='ES_noHR_60sCH_ribo ES_90sHR_60sCH_ribo
ES_120sHR_60sCH_ribo ES_150sHR_60sCH_ribo ES_180sHR_60sCH_
ribo';
mkdir -p ./comparison/adjusted;
for EXPERIMENT in $EXPERIMENTS; do papolarity adjust_features
"comparison/filtered/${ EXPERIMENT }.tsv" --sort-field
'cds_length' --fields "${ EXPERIMENT }_slope" "${ EXPERIMENT }
_slopelog" "${EXPERIMENT}_l1_distance" --mode z-score --window
500 --prefix 'zscore_' --output-file "./comparison/adjusted/${
EXPERIMENT}.tsv"; done;
```

3.3.7 Visualizing Per-Sample Distributions

As for polarity estimates, *papolarity* allows generating draft plots for resulting slope distributions separately for samples (results saved as `comparison/plot/*.png`):

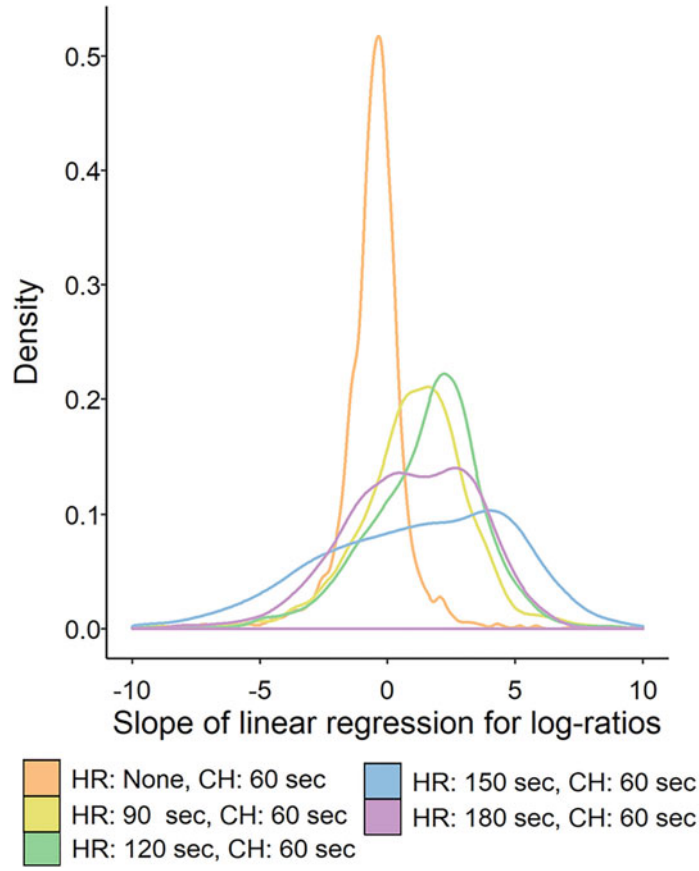

Fig. 5 Distribution of linear regression slope from profile $\log_2$ ratios estimated versus the control sample (`ES_noHR_noCH_ribo`). X axis: slope value. Y axis: distribution density; smoothed density estimate is shown for clarity. Different samples are shown in different colors (as shown in the figure legend)

```
EXPERIMENTS ='ES_noHR_60sCH_ribo ES_90sHR_60sCH_ribo
ES_120sHR_60sCH_ribo ES_150sHR_60sCH_ribo ES_180sHR_60sCH_
ribo';
mkdir -p ./comparison/plot;
for EXPERIMENT in $ EXPERIMENTS ; do papolarity plot_
distribution "comparison/adjusted/${EXPERIMENT}.tsv" --fields
"${EXPERIMENT}_slopelog" --no-legend --title $'Distribution of
linear regression slope\nfor normalized coverage log-ratios'
--zero-line green --xlim -10 10 --output-file "./comparison/
plot/${EXPERIMENT}_slopelog.png"; done;
```

4 Notes

The most recent development version of *papolarity* is available at https://github.com/autosome-ru/papolarity. For convenient copying-and-pasting, there is also a downloadable complete list of bash commands presented in this protocol. A stable version of *papolarity* can be installed as described in Materials.

4.1 Notes on Software Usage

1. **Computing hardware requirements and runtime estimates:** The data preprocessing steps (e.g., adapter trimming and read mapping; see Subheadings 2.2–2.4) are computationally intensive. The raw test data contained approximately 350 million short reads. The total amount of disk space used to preprocess the test data was approximately 40Gb (including SRR files and unpacked, merged, and trimmed FASTQ files). Everything except the trimmed FASTQ files (10Gb) is not necessary for testing the *papolarity* protocol. The maximum amount of RAM required was also 35–40Gb at the preprocessing stage (read mapping with *STAR*). The preprocessing runtime heavily depends on hardware performance (from an hour to several hours) and scales well with faster storage and more computational threads. Depending on the data volume, the slowest part can be genome indexing for read mapping. In the case of read alignments precalculated on a high-performance machine, the analysis of ribosome positional density can be performed on any modern personal computer in a few hours. For the test data, the total amount of disk space (excluding the input data and GTF annotation) was less than 400Mb; the maximum RAM usage peaked at 2Gb.
2. **Troubleshooting execution of commands:** For clarity, individual BASH commands in the protocol (from Subheading 2.2 and later) were separated by a semicolon `;` which is not necessary under normal usage (when each command is followed by a newline symbol, e.g., using the Enter key in the command prompt). A user should not mistake an optional semicolon separating bash commands `;` for mandatory escape sequence of semicolon `\;` which ends `find` command specification that is used for serial processing of multiple files at some steps. Importantly, the commands presented in the protocol are often longer than a single line of the code and thus are automatically formatted with line breaks. If the commands are retyped manually, visible line breaks should be **ignored**. For convenience, the complete code presented in this protocol is available at the *papolarity* GitHub page: https://github.com/autosome-ru/papolarity. There should be no major issues during the protocol execution, given the software and hardware requirements are met. However, minor mistypes will break the

chain of commands. Typically, the problems arise from incorrect file paths, non-activated *conda* environment, or special symbols (such as but not limited to dashes, commas, and spaces) in file and folder names. Thus alphanumeric naming of samples is strongly recommended (underscore _ and point . symbols are also acceptable). There might be some complex errors related to particular features of transcriptomic alignments from some rarely used alignment tools or incorrectly handled peculiarities of a particular GTF transcript annotation leading to problematic extraction of key transcript information.

3. **Optimal software versions:** Newer versions of the tools should be generally usable, except for possible major changes in the command line syntax, requiring the user to consult the respective software manuals. Despite this inconvenience, we strongly recommend using Table 2 only as a baseline reference and install the latest software packages when possible (including the presented *papolarity* and its required dependency *pasio*), since the updated versions often introduce critical bugfixes and major improvements: see also Subheading 2.2.
4. **Reproducing test data:** It might be useful to explore the presented protocol using exactly the same test data (see Subheading 2.3). The code below is not the most efficient or elegant way to download and unpack GEO data, but should be fully functional. The code will download, unpack, and prepare the test data allowing to reproduce the *papolarity* protocol execution step by step:

```
SRR='SRR315616 SRR315617 SRR315618 SRR315619 SRR315601
SRR315602 SRR315612 SRR315613 SRR315614 SRR315615 SRR315604
SRR315605 SRR315606 SRR315607 SRR315608 SRR315609 SRR315610
SRR315611';
 prefetch $SRR;
 mkdir raw;
 for SRR_FILE in $SRR; do fastq-dump -O ./raw --gzip $SRR_FILE
--split-files; done;
 cat ./raw/SRR315616_1.fastq.gz ./raw/SRR315617_1.fastq.gz ./
raw/SRR315618_1.fastq.gz ./raw/SRR315619_1.fastq.gz > ./
fastq/ES_noHR_noCH_ribo.fastq.gz;
 cat ./raw/SRR315601_1.fastq.gz ./raw/SRR315602_1.fastq.gz >
./fastq/ES_noHR_60sCH_ribo.fastq.gz;
 cat ./raw/SRR315612_1.fastq.gz ./raw/SRR315613_1.fastq.gz ./
raw/SRR315614_1.fastq.gz ./raw/SRR315615_1.fastq.gz > ./
fastq/ES_90sHR_60sCH_ribo.fastq.gz;
 cat ./raw/SRR315604_1.fastq.gz ./raw/SRR315605_1.fastq.gz ./
raw/SRR315606_1.fastq.gz > ./fastq/ES_120sHR_60sCH_ribo.
```

```
fastq.gz;
 cat ./raw/SRR315607_1.fastq.gz ./raw/SRR315608_1.fastq.gz ./
raw/SRR315609_1.fastq.gz > ./fastq/ES_150sHR_60sCH_ribo.
fastq.gz;
 cat ./raw/SRR315610_1.fastq.gz ./raw/SRR315611_1.fastq.gz >
./fastq/ES_180sHR_60sCH_ribo.fastq.gz;
```

5. **Parallel and sequential execution of commands:** At the preprocessing stage (see Subheading 2.4), we used 32 computational threads by setting the respective command-line parameters (`--runThreadN 32` of *STAR* and `-j 32` of *cutadapt*). These parameters might require tweaking in accordance with the available computational resources. Many commands in the actual *papolarity* protocol were wrapped in a `for` loop with `parallel`. By default, *GNU parallel* utilizes all available CPU cores, but this behavior can be customized (see `man parallel`). Alternatively, the `| parallel` ending of the commands can be replaced by `| bash`, and the respective commands will run sequentially. In case neither `| parallel` nor `| bash` is present, the commands will be printed in the terminal but not executed, which might be useful for debugging purposes.

4.2 Notes on Ribo-Seq Data Analysis

6. **Tweaking data preprocessing strategy:** The data preprocessing strategy (see Subheading 2) was not specifically optimized and, with minor modifications, should be generally applicable to a wide range of Ribo-Seq data obtained with Illumina sequencing. Of note, our workflow requires transcriptomic alignments (i.e., those in the coordinates of the spliced transcripts). Here we use *STAR* and transcripts annotation to obtain such alignments, but other strategies, such as direct mapping to transcript sequences, should be also usable and might be less computationally demanding. However, the genome-level transcript annotation (GTF) will still be necessary for the following workflow to extract transcript structures. In our protocol, for simplicity, we filter uniquely mapped reads before starting the actual analysis. In some cases, especially when studying the expression of repetitive elements or highly homologous genes, this step might be non-necessary or even misleading.
7. **Using arbitrary GTF annotation:** The protocol requires the GTF file (see Subheading 3.1) to follow common conventions and will fail if `gene_id` and `transcript_id` attributes are not present. The transcript CDS annotation can be used for clipping, but it is an optional step. The gene and transcript types filtering in the protocol was performed using GENCODE `gene_type` and `transcript_type` attributes but

should be easily adaptable to Ensembl GTF files by changing the respective filters to use `gene_biotype` and `transcript_biotype` attributes. Currently, *papolarity* toolbox supports only basic filters for GTF annotation, so external tools are required for more complex queries, e.g., to consider particular subsets of transcripts.

8. **Strategy for aggregation of footprint coverage profiles:** In Ribo-Seq data analysis, there is a commonly accepted approach to improve the resolution of the read coverage profiles, the so-called sub-codon phasing. Triplet periodicity of ribosome movement allows identifying ribosome P-site position in footprints of different lengths and thus produces quantitative per-codon estimates of ribosome coverage. However, we do not recommend to perform the reads phasing for the polarity score estimation and, especially, for the relative slope analysis. First, the accented triplet periodicity of the phased profiles will introduce a regular bias the profile segmentation step. Second, a notable fraction of non-phased reads is lost in phased profiles. Third, in many cases, the reliability of P-site identification within footprints is not possible due to weak triplet periodicity when Ribo-Seq experiments are performed with non-optimal nucleases (e.g., with micrococcal nuclease). Thus, we believe that, in general, the aggregation of complete footprint coverage serves as a more universal and robust approach.
9. **Filtering transcripts by coverage:** As a primary measure of expression, here we use mean pooled coverage, i.e., the total transcript coverage normalized by the transcript or CDS length (depending on whether clipping to CDS was applied or not). Transcripts with low mean coverage are excluded at the filtering stage (see Subheading 3.2). However, analysis of positional profiles also is not meaningful for transcripts with mean coverage obtained only from huge "spikes" in the mostly zero profile. Such spikes might represent various technical artifacts (e.g., PCR overamplification or read mapping issues) and affect polarity and, more significantly, relative slope estimation. Thus, in the protocol, we selected transcripts having the mean pooled coverage of no less than 1 and the third quartile (75th percentile) of the coverage distribution greater than 0. These are empirically selected numbers that can be optimized for particular data sets under analysis depending on the sequencing coverage and number of samples. Ideally, the coverage threshold should retain several thousands of transcripts. In the test data, 5984 transcripts passed the filter (5918 if only the major isoforms are considered). Low coverage will force including noisy profiles with multiple coverage gaps (e.g., in extreme cases only a few covered positions). Furthermore, additional filtering might be necessary in some cases, e.g., if there are

major differences in sequencing library size between samples or dramatically different footprint coverage of many transcripts. In such a scenario, a more reliable filter would require certain coverage of a transcript to be reached in all samples under consideration or in a given pair of an experiment and control samples (for the relative slope estimation). Since the *papolarity* protocol mainly relies on tab-separated text files, such a filter can be applied on any stage and should be convenient to implement with external tools.

10. **Segmentation parameter selection:** The *pasio* segmentation algorithm (used in Subheading 3.3) has multiple parameters, particularly, those related to the properties of the prior distribution of the Poisson lambda, which are estimated and considered constant in each segment. In particular, as the prior distribution of Poisson lambda, the *pasio* software uses gamma distribution $\Gamma(\alpha,\beta)$, and $(\beta/\alpha)^{\alpha}$ can be considered as a penalty for creating an additional segment. Thus the lengths of the segments and their total count can be adaptively controlled by adjusting the α and β parameters. The default settings ($\alpha = 1$ and $\beta = 1$) work pretty well, but extremely low or high Ribo-Seq coverage might require adjusting the parameters to obtain a reasonable number of segments.

Acknowledgments

A part of this project was performed at the School of Molecular and Theoretical Biology (Barcelona, 2019). We wholeheartedly thank Fyodor Kondrashov for organizing this incredible event. We thank Dmitry Nesterov, Maria Gavryushina, and Sofya Belyaeva for their invaluable contribution to the prototype of the presented software. We also thank Pavel Baranov for fruitful discussions and Evgenia Serebrova for help in manuscript preparation. This study was supported by the Russian Science Foundation [grant 19-74-10079] (polarity estimation) and the Russian Foundation for Basic Research [grant 19-34-51047] (Ribo-Seq data preparation).

References

1. Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324:218–223. https://doi.org/10.1126/science.1168978
2. Andreev DE, O'Connor PBF, Loughran G et al (2017) Insights into the mechanisms of eukaryotic translation gained with ribosome profiling. Nucleic Acids Res 45:513–526. https://doi.org/10.1093/nar/gkw1190
3. Ingolia NT, Hussmann JA, Weissman JS (2019) Ribosome profiling: global views of translation. Cold Spring Harb Perspect Biol:11. https://doi.org/10.1101/cshperspect.a032698
4. Buskirk AR, Green R (2017) Ribosome pausing, arrest and rescue in bacteria and eukaryotes. Philos Trans R Soc Lond B Biol Sci:372. https://doi.org/10.1098/rstb.2016.0183

5. 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. https://doi.org/10.1016/j.cell.2011.10.002
6. Gerashchenko MV, Lobanov AV, Gladyshev VN (2012) Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress. Proc Natl Acad Sci U S A 109:17394–17399. https://doi.org/10.1073/pnas.1120799109
7. Gerashchenko MV, Gladyshev VN (2014) Translation inhibitors cause abnormalities in ribosome profiling experiments. Nucleic Acids Res 42:e134. https://doi.org/10.1093/nar/gku671
8. Liu B, Han Y, Qian S-B (2013) Cotranslational response to proteotoxic stress by elongation pausing of ribosomes. Mol Cell 49:453–463. https://doi.org/10.1016/j.molcel.2012.12.001
9. Young DJ, Guydosh NR, Zhang F et al (2015) Rli1/ABCE1 recycles terminating ribosomes and controls translation reinitiation in 3′UTRs In Vivo. Cell 162:872–884. https://doi.org/10.1016/j.cell.2015.07.041
10. Young DJ, Makeeva DS, Zhang F et al (2018) Tma64/eIF2D, Tma20/MCT-1, and Tma22/DENR recycle post-termination 40S subunits in vivo. Mol Cell 71:761–774. e5. https://doi.org/10.1016/j.molcel.2018.07.028
11. Young DJ, Guydosh NR (2019) Hcr1/eIF3j Is a 60S Ribosomal subunit recycling accessory factor in vivo. Cell Rep 28:39–50. e4. https://doi.org/10.1016/j.celrep.2019.05.111
12. Kasari V, Pochopien AA, Margus T et al (2019) A role for the *Saccharomyces cerevisiae* ABCF protein New1 in translation termination/recycling. Nucleic Acids Res 47:8807–8820. https://doi.org/10.1093/nar/gkz600
13. Tuller T, Carmi A, Vestsigian K et al (2010) An evolutionarily conserved mechanism for controlling the efficiency of protein translation. Cell 141:344–354. https://doi.org/10.1016/j.cell.2010.03.031
14. Shalgi R, Hurt JA, Krykbaeva I et al (2013) Widespread regulation of translation by elongation pausing in heat shock. Mol Cell 49:439–452. https://doi.org/10.1016/j.molcel.2012.11.028
15. Kasari V, Margus T, Atkinson GC et al (2019) Ribosome profiling analysis of eEF3-depleted *Saccharomyces cerevisiae*. Sci Rep 9:3037. https://doi.org/10.1038/s41598-019-39403-y
16. Schuller AP, Wu CC-C, Dever TE et al (2017) eIF5A Functions globally in translation elongation and termination. Mol Cell 66:194–205. e5. https://doi.org/10.1016/j.molcel.2017.03.003
17. Campos RK, Wijeratne S, Shah P et al (2019) Ribosomal stalk proteins RPLP1 and RPLP2 promote biogenesis of flaviviral and cellular multi-pass transmembrane proteins. bioRxiv:713016. https://doi.org/10.1101/713016
18. Das Sharma S, Metz JB, Li H et al (2019) Widespread alterations in translation elongation in the brain of juvenile Fmr1 knockout mice. Cell Rep 26:3313–3322 .e5. https://doi.org/10.1016/j.celrep.2019.02.086
19. Chen Y-X, Xu Z, Wang B-W et al (2019) Selective translation by alternative bacterial ribosomes. bioRxiv:605931. https://doi.org/10.1101/605931
20. Mohammad F, Green R, Buskirk AR (2019) A systematically-revised ribosome profiling method for bacteria reveals pauses at single-codon resolution. Elife:8. https://doi.org/10.7554/eLife.42591
21. Andreev DE, O'Connor PB, Fahey C et al (2015) Translation of 5′ leaders is pervasive in genes resistant to eIF2 repression. Elife:4. https://doi.org/10.7554/eLife.03971
22. Kiniry SJ, Michel AM, Baranov PV (2019) Computational methods for ribosome profiling data analysis. Wiley Interdiscip Rev RNA: e1577. https://doi.org/10.1002/wrna.1577
23. Dunn JG, Weissman JS (2016) Plastid: nucleotide-resolution analysis of next-generation sequencing and genomics data. BMC Genomics 17:958. https://doi.org/10.1186/s12864-016-3278-x
24. Leinonen R, Sugawara H, Shumway M, International Nucleotide Sequence Database Collaboration (2011) The sequence read archive. Nucleic Acids Res 39:D19–D21. https://doi.org/10.1093/nar/gkq1019
25. Egorov AA, Sakharova EA, Anisimova AS et al (2019) Svist4get: a simple visualization tool for genomic tracks from sequencing experiments. BMC Bioinformatics. https://doi.org/10.1186/s12859-019-2706-8

Chapter 14

Unsupervised Bayesian Prediction of RNA Translation from Ribosome Profiling Data

Etienne Boileau and Christoph Dieterich

Abstract

Ribosome profiling has been instrumental in leading to important discoveries in several fields of life sciences. Here we describe a computational approach that enables identification of translation events on a genome-wide scale from ribosome profiling data. Periodic fragment sizes indicative of active translation are selected without supervision for each library. Our workflow allows to map the whole translational landscape of a given cell, tissue, or organism, under varying conditions, and can be used to expand the search for novel, uncharacterized open reading frames, such as regulatory upstream translation events. Through a detailed workflow example, we show how to perform qualitative and quantitative analysis of translatomes.

Key words Ribosome profiling, Translation, Open reading frame, Bayesian

1 Introduction

Ribosome profiling, or Ribo-seq, has been used to discover novel open reading frames (ORFs) [1], to characterize the dynamics of translation under various conditions, or to assess translational efficiency [2–4]. The technique is based on next-generation sequencing of libraries obtained from ribosome-protected RNA fragments, known as ribosome footprints, which are generated by enzymatic digestion [5]. Conceptually, actively translated regions are predicted based on ribosome density. In general, however, ribosome occupancy itself is not sufficient to determine translation, and sources of noise can arise due to non-ribosome-mediated RNA protection, ribosome scanning, aborted translation events, or technical artifacts from the experimental procedure or the library preparation. Specially designed methodologies are required to recover the active translation events from the ribosome profiling data [6–10]. Here, we use Rp-Bp to predict translation events from ribosome profiling data [7]. Rp-Bp is an unsupervised Bayesian approach based on probabilistic graphical models that is used to (i) recover the precise location of the peptidyl-site (P-site) of the

Vyacheslav M. Labunskyy (ed.), *Ribosome Profiling: Methods and Protocols*, Methods in Molecular Biology, vol. 2252, https://doi.org/10.1007/978-1-0716-1150-0_14,

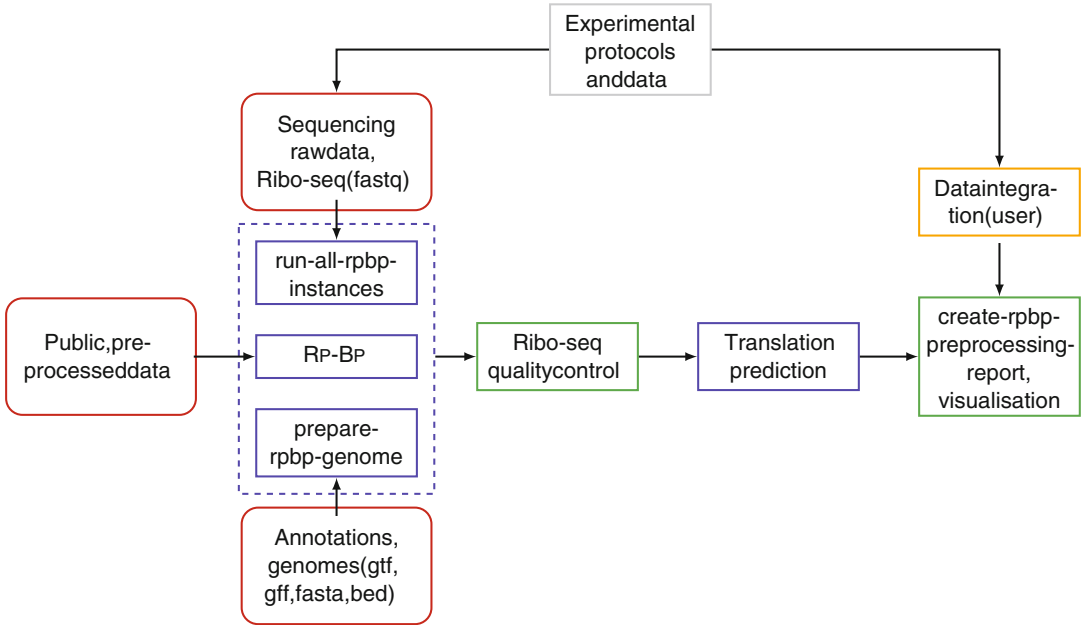

Fig. 1 Schematic overview of the Rp-Bp workflow: index creation (`prepare-rpbp-genome`), periodicity estimation and translation prediction (`run-all-rpbp-instances`), including quality control and visualization (`create-rpbp-preprocessing-report`). Results from Rp-Bp (in FASTA and BED format) can be used in combination with other software or external databases. All steps of the pipeline are described in this chapter

ribosome within each fragment and automatically select fragment lengths that exhibit a 3-nucleotide periodicity along the ORF and (ii) predict translated ORFs across the whole transcriptome, including de novo translation events. Through this chapter, we show how to install and use Rp-Bp and describe an example workflow to estimate footprint periodicity and predict translation (Fig. 1).

2 Materials

2.1 Hardware

For most practical applications (big data, large genome), a machine with a x86-64 architecture, either a Linux or macOS (previously Mac OS X) operating system, and at least 30GB of RAM is required. Our workflow was developed and tested on Debian GNU/Linux, including the Ubuntu distribution, and is integrated to run with the Slurm workload manager [11].

2.2 Software

The following software is required:

1. Python 3, including the third-party Python packages setuptools and pip. For Linux users, the procedure to install a specific version of Python depends on the distribution. Python can also

be obtained from the Python Software Foundation website at python.org. The best way to install Python 3 on macOS is through the Homebrew package manager.

2. Rp-Bp (ribosome profiling with Bayesian predictions), available from https://github.com/dieterich-lab/rp-bp. We recommend to install the package in a virtual environment. We suggest to use the installation script available online. To download it

```
wget https://data.dieterichlab.org/s/rp-bp-mmb/download \
-O rp-bp-mmb.zip
```

Extract the compressed zipped folder, and then install the package

```
cd rp-bp-mmb
chmod +x setup; ./setup
```

Additional files to run the example workflow are also included. The script will create a pre-defined directory structure and prepare the configuration file necessary to run the workflow (macOS systems do not come with wget; install using `brew install wget`).

3. Flexbar (Flexible Barcode Adapter and Removal), available at https://github.com/seqan/flexbar. The package is available as pre-compiled executables for Linux and macOS and via package managers on Debian systems, in Homebrew, and in Bioconda.
4. Bowtie2, available from https://github.com/BenLangmead/bowtie2 or http://bowtie-bio.sourceforge.net/bowtie2. The package is available as pre-compiled executables for Linux and macOS and via package managers, notably Bioconda.
5. STAR (Spliced Transcripts Alignment to a Reference), available from https://github.com/alexdobin/STAR. The package is available as pre-compiled executables for Linux and macOS.
6. Samtools, available from http://www.htslib.org or https://github.com/samtools/samtools. The package is also available via package managers, notably Bioconda.
7. (*Optional*) FastQC, available from https://www.bioinformatics.babraham.ac.uk/projects/download.html#fastqc

Supported versions of the different software and packages are specified in each Rp-Bp release. The current workflow is based on Rp-Bp version 2.0.0 and is compatible with Python version >=3.6, <3.7.0a0, Flexbar version 3.5.0, Bowtie2 version 2.3.0, STAR version 2.6.1d, and Samtools version 1.7 (*see* **Notes 1** and **2**).

2.3 Data Sources

To run the example workflow, we use a subset of the samples from Chothani et al. [12]. The data consists of genome-wide measurements of translation during human cardiac fibroblast activation monitored with ribosome profiling.

1. Ribosome profiling raw FASTQ files used in this example workflow, via the NCBI SRA https://www.ncbi.nlm.nih.gov/sra, accession PRJNA542670 (BioProject). We use the following SRA accessions: SRR9049055, SRR9049060, and SRR9049071. A summary run table of the SRA is given in the supplementary material, along with additional files to run the workflow (https://data.dieterichlab.org/s/rp-bp-mmb/download). To use the SRA toolkit, see https://www.ncbi.nlm.nih.gov/sra/docs/toolkitsoft. After downloading the files, put them in the destination specified in the configuration file (**rp-bp-mmb/riboseq-analysis/raw-data**).
2. Genome sequences in FASTA format and annotations in GTF format. The annotations must match the version of the reference genome, and both exon and CDS features must be present (*see* **Note 3**). To download the human genome and annotations (Ensembl release 96):

```
wget ftp://ftp.ensembl.org/pub/release-96/fasta/homo_
sapiens/dna/Homo_sapiens.GRCh38.dna.primary_assembly.fa.gz

wget ftp://ftp.ensembl.org/pub/release-96/gtf/homo_
sapiens/Homo_sapiens.GRCh38.96.gtf.gz
```

   Decompress the gzipped files prior to running the workflow, and put them in the destination specified in the configuration file (**rp-bp-mmb/riboseq-analysis/genome**). The ribosomal DNA sequences in FASTA format are included in the supplementary material and will be located under **rp-bp-mmb/riboseq-analysis/genome/rRNA** after running the installation script.

3 Methods

If Rp-Bp has been installed in a Python 3 virtual environment (using the installation script provided in the supplementary material), this environment needs to be activated to run this example workflow. In the following, paths are relative to the current working directory (**rp-bp-mmb**).

```
source envs/rpbpenv/bin/activate
# go to the config directory
cd riboseq-analysis/config
```

All subsequent commands have to be run from within the **config** directory (*see* **Note 4**).

3.1 Quality Control

Pre-processing analysis and ribosome profiling-specific QC are described in the Section ***Quality control and periodicity estimation***. FastQC reports can be generated automatically on all datasets (raw, trimmed/filtered, rRNA-depleted, and mapped reads).

3.2 Creating Reference Genome Indices

This section describes how to prepare genome indices and annotation files used by Rp-Bp. This step must be run once for each reference genome and annotations, before estimating periodicity and/or predicting RNA translation events. To run the pre-processing phase

```
prepare-rpbp-genome rpbp-pipeline.yaml [logging options] \
[processing options]
```

In the following, we are not specifying what are the logging and parallel processing options. For more information, the reader is referred to the online documentation at https://rp-bp.readthedocs.io/en/latest/usage-instructions.html. The two most useful processing options are `--num-cpus` (number of processes to spawn; for STAR, number of threads passed via `--runThreadN`) and `--mem` (the amount of RAM to request for STAR genome indexing via `--limitGenomeGenerateRAM`).

3.2.1 Output Files

The following output files will be created under the parent directory specified by `genome_base_path` (in the configuration file, i.e., **rp-bp-mmb/riboseq-analysis/genome**):

- The Bowtie2 and STAR index files, under the directories **rRNA** and **star**, respectively (or otherwise depending on the configuration file)
- A file in BED format containing transcript information used by Rp-Bp, with a name specified by `genome_name` (in the configuration file)
- **transcript-index**: a directory where annotated transcript sequences, ORF coordinates, and labels are written to files in FASTA and BED format, respectively

3.2.2 A Note on ORF Labels

Rp-Bp identifies all ORFs (annotated and unannotated) based on their transcript exon structure, using the spliced transcript sequences extracted from the reference genome. We understand an ORF as a potentially translatable sequence that consists of a series of codons beginning with a start codon and ending with a stop codon. Translatable ORFs can be found anywhere: in the 5′ untranslated region (5′UTR), in the 3′ untranslated region

(3′UTR), within or overlapping with annotated coding sequences (CDSs), in transcripts that were previously thought to be noncoding (lincRNAs, antisense, pseudogene, or other processed transcripts), or in novel transcripts (intra-/intergenic; *see* **Note 5**). Unless specified, Rp-Bp only identifies ORFs starting with an AUG codon. It is possible to use non-AUG start codons; however, this option has not been thoroughly validated.

Coordinates and unique identifiers are assigned to the ORFs using the following convention, *trx_chrom:start-stop:strand*, where *trx* is an associated transcript identifier, *chrom* is the chromosome or contig, *start*, *end* are genomic start and end coordinates in BED style (first nucleotide of start codon included, but first nucleotide of stop codon excluded), and *strand* is the frame orientation. The ORFs are also labeled according to their position relative to the annotated coding sequences. A schematic of the ORF labels is shown in Fig. 2. It is important to note that label assignment is independent of RNA translation prediction, i.e., Rp-Bp uses the ORF coordinates and the ribosome profiling data to predict translation, regardless of the labels (*see* **Note 6**).

3.3 Running the Rp-Bp Pipeline

The workflow is split in two major steps: First, Rp-Bp selects periodic ribosome profiling read lengths and their P-site offsets,

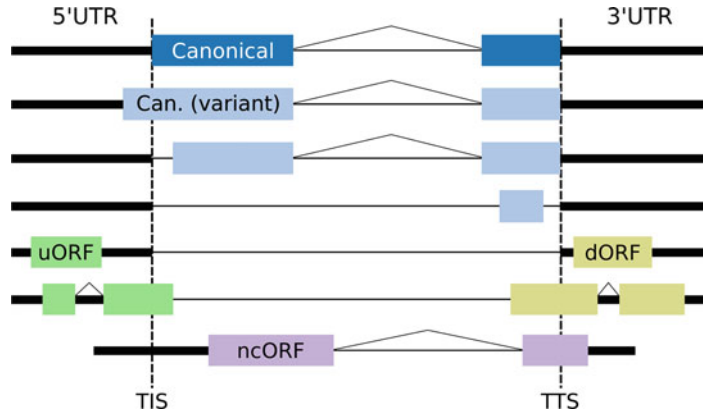

Fig. 2 Schematic of the label assignment. Canonical ORFs correspond to the annotated coding sequences, and their start and stop positions coincide with the annotated translation initiation and termination sites, respectively. Canonical variants include "N-terminus"-extended and "N-terminus"-truncated sequences and ORFs that are found inside canonical ORFs (typically out-of-frame). Upstream (uORF) and downstream (dORF) ORFs are found in the untranslated regions but can also overlap the primary coding sequence. Noncoding ORFs (ncORF) are found in transcripts annotated as noncoding, including biotypes such as processed transcripts, small noncoding RNAs, or pseudogenes. *5′UTR* 5′ untranslated region, *3′UTR* 3′ untranslated region, *TIS* annotated translation initiation site, *TTS* annotated translation termination site

and the ORF profiles are created using the P-site-shifted 5′ end of each read; second, Rp-Bp predicts translation from the profiles, for all the ORFs identified above, for each sample or replicate. All steps are unsupervised and do not require parameter tuning or model training. A Bayesian model selection approach is used to incorporate and propagate uncertainty in the inference process. To run the full pipeline

```
run-all-rpbp-instances rpbp-pipeline.yaml \
--merge-replicates --run-replicates \
--keep-intermediate-files \
[logging options] [processing options]
```

The flag `--merge-replicates` is used to indicate that biological replicates from the same condition, specified in the configuration file, must be considered together for predicting translation events. Rp-Bp handles replicates by adding their ORF profiles, allowing selection of different read lengths from different replicates and/or different P-site offsets for reads of the same length across replicates. If the `--merge-replicates` flag is given, then predictions will not be made for the individual datasets, unless the `--run-replicates` flag is also given, in which case predictions will be made for both the merged replicates and the individual samples. By default, if none of these flags are provided, predictions are made for the individual samples only. We use the `--keep-intermediate-files` flag to keep track of all the intermediate steps, which will allow us to quantify the proportion of reads filtered out for the creation of the ORF profiles.

If we are not interested in ORF discovery, we can run the first part of the pipeline only using the `--profiles-only` flag. This option overrides the `--merge-replicates`. This can be useful for quality control, before further processing, or if a different workflow is intended, e.g., differential expression using periodic-only ribosome footprint lengths.

3.4 Creating the ORF Profiles

To estimate periodicity, Rp-Bp first constructs a base genome profile, as follows:

1. Remove sequencing adapters and low-quality reads using Flexbar. We use a generic list of adapters, included in the configuration file using the key `adapter_file:`, followed by the path to the FASTA file. Ideally, this list contains only adapters that were used for the library preparation. If a single adapter sequence has been used, it can be included directly in the configuration file following the key `adapter_sequence:`. We use these options:

```
--max-uncalled 1 \
--pre-trim-left 0 \
--qtrim-format sanger \
--qtrim TAIL \
--qtrim-threshold 10
```

   To override default settings, or to pass additional parameters to Flexbar, we can use `--flexbar-options`. The Flexbar options and their values are read in by Rp-Bp as a space-delimited list, with each option quoted separately, such as `--flexbar-options "--qtrim-threshold 25"`, and passed to Flexbar. The long parameter name must be used, i.e., `--qtrim-threshold`, instead of `-qt`. By default, output files are compressed (gz). Parallel processing options (number of threads to employ) are passed to Flexbar via the usual option `--num-cpus`.

2. Remove reads mapping to ribosomal sequences. Although the input material is generally treated to remove as much ribosomal RNA (rRNA) as possible, experience has shown that contamination can still represent a large proportion of any given library after sequencing. This contamination can be an issue, so a subtracting strategy is usually necessary. We employ Bowtie2 to map adapter-free reads to a reference made up of ribosomal RNA sequences, using the option `--very-fast`. The reference FASTA file was used to generate the Bowtie2 index when we created the reference genome indices. At this stage, we use the created index to filter out rRNA reads. We use for our analysis those reads that are not mapping onto this reference. By default, output files are compressed (gz). Parallel processing options (number of threads to employ) are passed to Bowtie2 via the usual option `--num-cpus`. Currently, it is not possible to override Bowtie2 settings.

3. Align reads to the genome using a splice-aware aligner. We use the STAR index created above to map all our trimmed, rRNA-free reads on the genome. Since we did not use the annotations at the index creation step, they are used on the fly for mapping and passed via `--sjdbGTFfile`. We use these options:

```
--outFilterMismatchNmax 1 \
--outFilterType BySJout \
--outFilterIntronMotifs RemoveNoncanonicalUnannotated \
--outSAMattributes AS NH HI nM MD \
--outSAMtype BAM SortedByCoordinate \
--sjdbOverhang 33 \
--seedSearchStartLmaxOverLread 0.5 \
--winAnchorMultimapNmax 100
```

   To override default settings, or to pass additional parameters to STAR, we can use `--star-options`. The STAR options and their values are read by Rp-Bp as a space-delimited list, with each option quoted separately, such as `--star-options "alignIntronMax 10000"`, and passed to STAR. Results are written as sorted, compressed BAM files using `--outSAMtype BAM SortedByCoordinate`. The additional SAM attribute MD tag is added, but it is not used in Rp-Bp. We adjust a number of parameters for short read mapping: `--sjdbOverhang 33`, or set it to the 90-percentile of all fragment lengths; to seed from the middle of the read, we use `--seedSearchStartLmaxOverLread 0.5`; and `--winAnchorMultimapNmax 100` to increase the number of loci anchors are allowed to map to. Unspecified options are default STAR options. The maximum available RAM for sorting needs to be set via the usual `--mem` option. Parallel processing options (number of threads to employ) are passed to STAR using `--num-cpus`.

4. Remove reads with multiple alignments. This is done after mapping, using the NH SAM tag (a value of 1 corresponds to unique mappers). Rp-Bp removes all multi-mappers by default and uses uniquely mapped reads to estimate periodicity and construct the ORF profiles (*see* **Note 7**).

   Using uniquely mapped reads, a metagene profile is constructed for each read length by counting the 5′ ends of all reads aligned at each position. Using probabilistic graphical models, Rp-Bp then estimates the periodicity of the metagene profiles of each read length starting at the observed peak. We only keep read lengths that are periodic, according to the models, and the location of the peak gives the P-site offset for reads of that length. This technique is referred to as *Bayesian*

Periodic fragment length and P-site offset Selection (BPPS). It is also possible to skip the BPPS and use fixed values (*see* **Notes 8** and **9**). Default Rp-Bp settings for the periodicity estimation and Markov chain Monte Carlo (MCMC) simulations can be specified via the configuration file. The reader is referred to the online documentation at https://rp-bp.readthedocs.io/en/latest/usage-instructions.html#running-the-rp-bp-pipeline, in particular *metagene periodicity*, *periodicity and offset*, *smoothing*, and *shared MCMC* options.

3.4.1 Output Files

After running the first part, the following output directories will be created under the parent directory specified by `riboseq_data:` (in the configuration file):

- **without-adapters**: contains trimmed and filtered reads in compressed FASTQ format (output from Flexbar)
- **with-rrna**: contains discarded reads aligning to the ribosomal index in compressed FASTQ format (output from Bowtie2)
- **without-rrna**: contains rRNA-free reads in compressed FASTQ format (output from Bowtie2)
- **without-rrna-mapping**: contains sorted reads aligned to the genome, including unique mappers, in BAM format (output from STAR)
- **metagene-profiles**: contains
  - The metagene profiles (*metagene-profile*) for all read lengths in gzipped CSV format.
  - The periodicity estimates (*metagene-periodicity-bayes-factors*) for all P-site offsets in gzipped CSV format.
  - The selected P-site offsets (*periodic-offsets*) for each read length in gzipped CSV format. This file includes all read lengths; filtering is done on the fly by Rp-Bp.
- **orf-profiles**: contains the unsmoothed ORF profiles (*profiles*) in compressed sparse matrix market format (*see* **Note 10**)

The file names follow the convention *<sample-name>[.<note>]*, where *sample-name* is a key from the `riboseq_samples:` entry in the configuration file and where *note* is optionally given by `note: note`. Unless `keep_riboseq_multimappers:` is used, all files after mapping will additionally contain the string *-unique*. For further details on the naming conventions, the reader is referred to the online documentation.

3.5 Predicting Translated ORFs

This part of the workflow is performed automatically, unless the `--profiles-only` flag has been passed (*see* **Note 11**). Before proceeding to the Bayesian model selection and inference, unlikely ORFs are filtered out, based on a minimum number of reads,

which can be specified in the configuration file using `orf_min_profile_count_pre:` (default is 5). Internally, we ensure that the number of reads mapped to the first reading frame always exceed the number of reads mapped in either of the other two reading frames, individually. The final prediction set is further filtered such that the longest ORF predicted is selected for each stop codon, and among each group of overlapping ORFs, the one with the highest expected Bayes factor is chosen. The minimum ORF length can be specified in the configuration file using `orf_min_length:` (default is >8 nucleotides or 3 codons). The reader is referred to the online documentation, in particular *Bayes factor estimation*, *selecting predicted ORFs*, and *shared MCMC* options.

3.5.1 Output Files

After running the second part, the following output directories will be created under the parent directory specified by `riboseq_data:` (in the configuration file):

- **orf-predictions**: contains
  - The Bayes factor estimates for all ORFs in gzipped BED format (*bayes-factors*)
  - The ORFs predicted as translated in gzipped BED format (*filtered.predicted-orfs*)
  - The DNA sequence for the ORFs predicted as translated in FASTA format (*filtered.predicted-orfs.dna*), the fasta header matching the id column in the BED file
  - The protein sequence for the ORFs predicted as translated in FASTA format (*filtered.predicted-orfs.protein*), the fasta header matching the id column in the BED file

  If replicates are merged, then these files will be created for each condition. Otherwise, they will be created for each sample (or both if the appropriate options are given; *see* **Note 12**).

3.6 Analysis

Rp-Bp offers an extended quality control workflow, with post-processing scripts to facilitate the analysis (*see* **Note 13**). This section describes the final output from Rp-Bp and how these results can be used for downstream analysis.

3.6.1 Quality Control and Periodicity Estimation

To generate plots that summarize the pre-processing and ORF profile construction

```
create-rpbp-preprocessing-report rpbp-pipeline.yaml \
../analysis [logging options] [processing options]
```

It is required to specify the output directory `../analysis`. If this directory does not exist, it will be created. We use a relative path from the **config** directory. It is possible to specify the absolute path

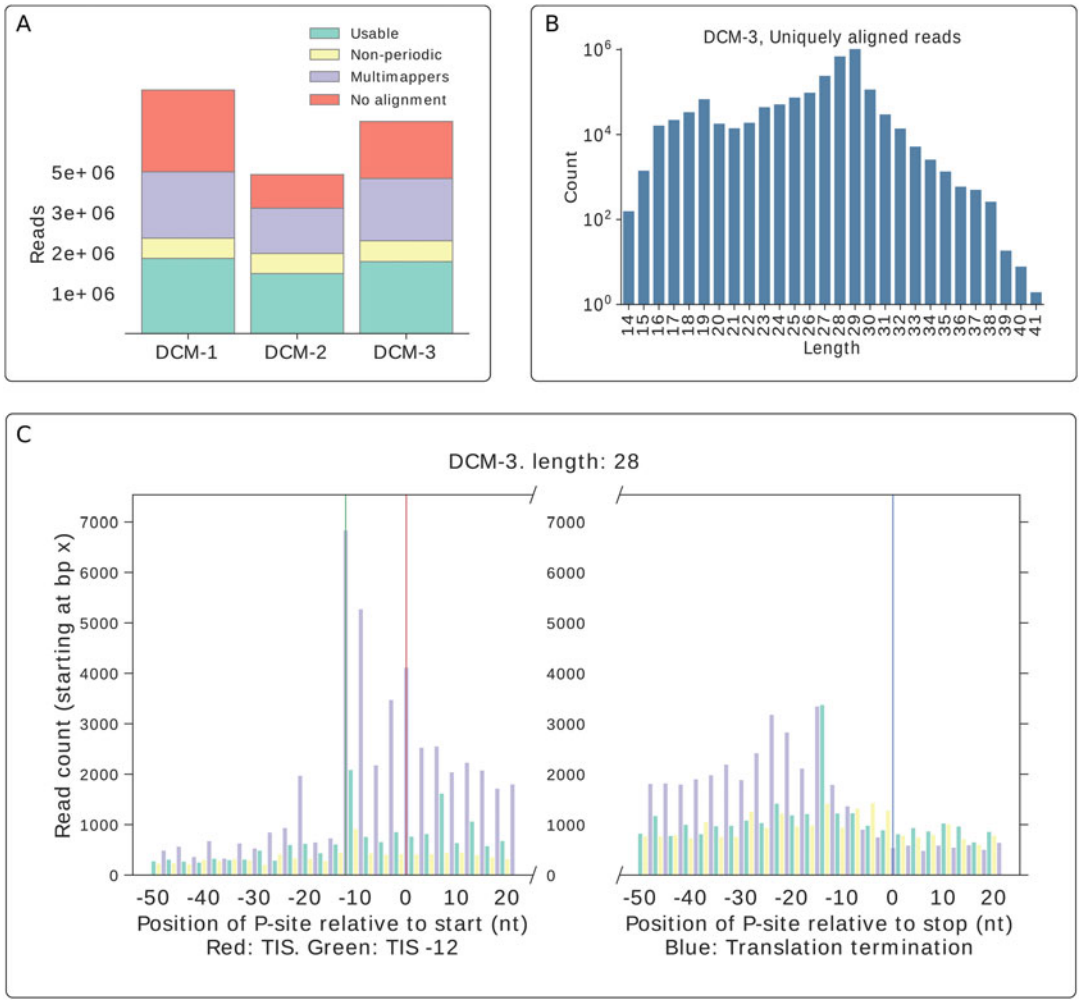

Fig. 3 Selected figures from the post-processing script. (**a**) The reads filtered at each step of the pipeline (excluding rRNA and poor-quality reads), (**b**) the distribution of uniquely mapped reads, and (**c**) the metagene profile for one fragment length, for one given replicate

using **path/to/rp-bp-mmb/riboseq-analysis/analysis**, where the first part **path/to** depends on where this workflow example has been downloaded and installed. To generate FastQC reports for all datasets (raw, trimmed/filtered, rRNA-depleted, and mapped reads), we can use the flag `--create-fastqc-reports`. The program FastQC needs to be installed and callable. Additional options (`--show-orf-periodicity --show-read-length-bfs`) can be passed to this script to generate metagene periodicity plots and Bayes factor for each possible P-site offset for each read length. We do not recommend to use these options, as these plots are time-consuming to create and not immediately informative and/or easy to interpret.

Table 1
Number of ORFs predicted on positive and negative strand, for the merged replicates (DCM)

ORF category	+	−
Canonical	5837	5789
Canonical (variant)	225	236
Canonical (within)	27	35
Five prime (uORF)	789	767
Five prime overlap (uORF)	27	25
Three prime (dORF)	88	77
Three prime overlap (dORF)	19	19
Noncoding (ncORF)	398	297

The most useful output files from this script are the counts and figures to visualize the reads filtered at each step of the pipeline, the figures to visualize the read-length distributions, and the metagene profiles (Fig. 3). A PDF report is also generated, which integrates all figures into a single document.

3.6.2 Output Files

If using `--create-fastqc-reports`, a subdirectory **fastqc** is generated under each of the following directories: **without-adapters**, **with-rrna**, **without-rrna**, and **without-rrna-mapping** (see the Subheading 3.4).

- **riboseq-results**: contains the data (*read-filtering-counts.csv.gz*) and the figures showing the reads filtered at each step of the pipeline
- **without-rrna-mapping**: contains the data files (*length-distribution.csv.gz*) and, under **plots**, figures for the read-length distributions (*see* **Notes 14–16**)
- **metagene-profiles**: contains the metagene profile figures for each fragment length

Additionally, the PDF report is found under the output directory specified when calling `create-rpbp-preprocessing-report`.

3.6.3 ORF Predictions

There is a command that can be used to generate a report (similar to the pre-processing analysis), called `create-rpbp-predictions-report`. We are currently working on a "next-generation" solution to the PDF reports and figures. Using the ORF predictions under **orf-predictions**, it is possible to get an overview of the translational landscape (Table 1; *see* **Note 17**). Beyond this introduction, the ORF predictions can be used as a starting point for an in-depth analysis, e.g., by performing conservation, or sequence/motif analyses, using the DNA or protein sequences. In particular,

the non-canonical ORFs can be characterized using InterPro [13], SignalP [14], or TMHMM [15], providing further functional insights.

4 Notes

1. To run the example workflow, we recommend to download pre-compiled executables, whenever possible, and avoid installing these via a Bioconda environment, since they will be invisible to the Rp-Bp installation. Flexbar, Bowtie2, STAR, and Samtools executables must be available on your `$PATH`, otherwise they will not be callable from Rp-Bp. The executables can be downloaded to a directory such as `$HOME/.local/bin`, which is usually available on the `$PATH`. If you are using Bash, you can set the $PATH variable in the *~/.bashrc* file by adding the following line

```
export PATH=$PATH:$HOME/.local/bin
```

   and sourcing the file `source ~/.bashrc`. To confirm that the directory was successfully added, print the value of your `$PATH` by typing `echo $PATH`.

2. Rp-Bp is designed to handle all steps of the prediction pipeline starting from the raw FASTQ files. Residual adapter removal with Flexbar and subsequent steps of the read mapping workflow up to and including read alignment using STAR are incorporated in Rp-Bp. It is however possible to predict RNA translation from existing alignments in the BAM file format, e.g., if FASTQ files were mapped using a different strategy. In the latter case, it is not required to install Flexbar, Bowtie2, and STAR. To use Rp-Bp on these BAM files, however, they must be placed at the appropriate location and follow the naming convention used in Rp-Bp. The reader is referred to the online documentation at https://rp-bp.readthedocs.io/en/latest/custom-alignment-files.html#using-custom-alignment-files.
3. While we advise to use the most recent version available of the reference genome, we do not recommend to use the "top level" assembly, which includes haplotype information. This could significantly increase the running time allocated for mapping and may result in many reads being discarded as multi-mappers, where sequence overlap is non-negligible. Here, we use the "primary" assembly from Ensembl.
4. If changing the directory structure of the example, or when running your own workflow, the configuration file needs to be adjusted so that all paths point to the correct destinations.

5. If matching RNA-seq is available, it is possible to use a de novo assembly with Rp-Bp to identify novel ORFs (i.e., ORFs in novel transcripts, and not only unannotated ORFs in known exonic sequences, which can be identified using a standard annotation). The information from a de novo assembly must follow the same format specifications as the main GTF file. It is specified in the configuration file using the key `de_novo_gtf`.
6. For complex genomes and annotations, the transcript identifier of the ORF may not "match" the assigned label, in cases where multiple isoforms contain the same ORF, despite the label being correctly assigned. In such cases, it is advised to look at the gene structure, e.g., using the Integrative Genomics Viewer (IGV), and rely uniquely on genomic coordinates.
7. If we want to keep multi-mapping ribosome footprints, we can add the key:value `keep_riboseq_multimappers: TRUE` in the configuration file. If the key is present in the configuration file with any value, including `FALSE`, multi-mapping reads will be kept for analysis. Unless there are good reasons to do so, we do not recommend to keep multi-mappers.
8. Rp-Bp uses the length of the alignment query sequence from the BAM file, i.e., it excludes soft-clipped bases, which does not correspond to the actual fragment length in the FASTQ file.
9. To skip the BPPS, the key `use_fixed_lengths:` with any value must be given in the configuration file, together with `lengths:` and `offsets:`, each consisting of a list of fragment lengths and P-site offsets, respectively, that will be used for creating the profiles. Unless there are reasons to do so, we do not advise to use fixed lengths, as Rp-Bp is designed to discover without supervision which fragment lengths should be used.
10. The matrix market format (mtx) uses base-1 indices, while Rp-Bp numbers the ORFs using base-0 indices.
11. To run the prediction pipeline, in the case where only the first part is available (e.g., after using the `--profiles-only` flag), it suffices to call again the main script `run-all-rpbp-instances`, as described above, and to omit the `--profiles-only` flag. It is important not to use the `--overwrite` flag, such that steps for which the output files already exist will be skipped and files will not be overwritten.
12. There are *unfiltered* and *filtered* versions of the files. The *filtered* version results from performing the filtering described in the text (taking the longest predicted ORF for each stop codon and then selecting the ORF with the highest expected Bayes factor among each group of overlapping ORFs). We always recommend to use the *filtered* predictions.

13. Several recent analysis methods have been described, including RiboTaper [6], Ribo-TISH [8], RiboCode [9], and PRICE [10]. The RiboTaper method is well established, but not yet available on software development version control platforms. Its documentation is limited, and it requires matched RNA-seq data. Methods such as RiboCode, Ribo-TISH, and PRICE are all available on GitHub and provide minimal documentation. Not all methods offer extended analysis scripts or quality control assessment. Ribo-TISH is specifically designed to handle TI-seq/QTI-seq data. PRICE is the most recent method and is able to handle overlapping ORFs and non-canonical translation initiation. While Rp-Bp is unable to distinguish overlapping ORFs and is based on unnormalized parameter estimates, it propagates uncertainty by maintaining distributions over quantities of interest through the entire prediction process. Rp-Bp is also available on GitHub, maintained, and well-documented. It is modular, so that only parts of the pipeline can be used, and it can be run in parallel on computing clusters integrated with the Slurm workload manager.
14. A characteristic feature of a high-quality Ribo-seq library is its read-length distribution, which typically peaks around 29 nucleotides in eukaryotic organisms; however, broader distributions can be observed under different protocols, depending on the nuclease treatment, the drugs/inhibitors used, etc. It is also known that different ribosomal conformations correspond to distinct read-length distributions and that these can also be affected by ribosomes belonging to different pools (mitochondrial ribosomes were shown previously to display a bimodal distribution, compared to cytosolic-derived fragments). All these considerations must be taken into account when analyzing the distribution of read lengths.
15. To determine globally the proportion of reads mapped to each frame, for each replicate, we can use the ORF profiles (in sparse mtx format, under **orf-profiles**) and sum all entries across each frame. The profiles for a given replicate can be loaded in Python using `numpy.loadtxt('path/to/mtx', skiprows=3)`.
16. The selected fragment lengths and their P-site offsets can be collected from selected P-site offsets file (*periodic-offsets*). However, instead of using this file and relying on the internal filtering criteria to find the periodic fragment lengths, we can collect them "manually" from the file names, for each replicate, under **orf-profiles**. The naming convention is as follows: *<sample-name>[.<note>]-unique.length-<lengths>.offset-<offsets>.profiles.mtx.gz*. This can be useful if performing translation efficiency analysis or other downstream analysis using

periodic Ribo-seq reads only. To extract periodic fragments from a BAM file

```
import pysam

bam = pysam.AlignmentFile('path/to/original.bam')
alignments = bam.fetch()
out_bam = pysam.AlignmentFile('path/to/periodic.bam', "wb",
template=bam)

lengths = [26, 28, 29, 30]

for a in alignments:
 if a.qlen in lengths:
  out_bam.write(a)
 else:
  pass
out_bam.close()
```

17. In a high-quality Ribo-seq library, reads mostly map to coding sequences (CDS or Canonical) (typically >85%) and to the 5'UTR (up to 10%). A smaller proportion map to the 3'UTR. The amount of reads mapping to noncoding regions can vary, but in general the signal is not very strong. These numbers are generally reflected in the ORF predictions. It is possible to use other tools such as `bedtools coverage`, to explore the distribution of the reads, in combination with the ORF predictions.

Acknowledgments

The authors acknowledge funding by the Klaus Tschira Foundation. The Rp-Bp package was developed by Brandon Malone. Computing infrastructures in the Dieterich Lab are maintained by Tobias Jakobi.

References

1. Ruiz-Orera J, Alba MM (2019) Translation of small open reading frames: roles in regulation and evolutionary innovation. Trends Genet 35 (3):186–198. https://doi.org/10.1016/j.tig.2018.12.003
2. Kmietczyk V, Riechert E, Kalinski L, Boileau E, Malovrh E, Malone B, Gorska A, Hofmann C, Varma E, Jurgensen L, Kamuf-Schenk V, Altmuller J, Tappu R, Busch M, Most P, Katus HA, Dieterich C, Volkers M (2019) m (6)A-mRNA methylation regulates cardiac gene expression and cellular growth. Life Sci Alliance 2:2. https://doi.org/10.26508/lsa.201800233
3. Doroudgar S, Hofmann C, Boileau E, Malone B, Riechert E, Gorska AA, Jakobi T, Sandmann C, Jurgensen L, Kmietczyk V, Malovrh E, Burghaus J, Rettel M, Stein F, Younesi F, Friedrich UA, Mauz V, Backs J, Kramer G, Katus HA, Dieterich C, Volkers M

(2019) Monitoring cell-type-specific gene expression using ribosome profiling in vivo during cardiac hemodynamic stress. Circ Res 125(4):431–448. https://doi.org/10.1161/CIRCRESAHA.119.314817

4. van Heesch S, Witte F, Schneider-Lunitz V, Schulz JF, Adami E, Faber AB, Kirchner M, Maatz H, Blachut S, Sandmann CL, Kanda M, Worth CL, Schafer S, Calviello L, Merriott R, Patone G, Hummel O, Wyler E, Obermayer B, Mucke MB, Lindberg EL, Trnka F, Memczak S, Schilling M, Felkin LE, Barton PJR, Quaife NM, Vanezis K, Diecke S, Mukai M, Mah N, Oh SJ, Kurtz A, Schramm C, Schwinge D, Sebode M, Harakalova M, Asselbergs FW, Vink A, de Weger RA, Viswanathan S, Widjaja AA, Gartner-Rommel A, Milting H, Dos Remedios C, Knosalla C, Mertins P, Landthaler M, Vingron M, Linke WA, Seidman JG, Seidman CE, Rajewsky N, Ohler U, Cook SA, Hubner N (2019) The translational landscape of the human heart. Cell 178 (1):242–260. e229. https://doi.org/10.1016/j.cell.2019.05.010
5. Aeschimann F, Xiong J, Arnold A, Dieterich C, Grosshans H (2015) Transcriptome-wide measurement of ribosomal occupancy by ribosome profiling. Methods 85:75–89. https://doi.org/10.1016/j.ymeth.2015.06.013
6. Calviello L, Mukherjee N, Wyler E, Zauber H, Hirsekorn A, Selbach M, Landthaler M, Obermayer B, Ohler U (2016) Detecting actively translated open reading frames in ribosome profiling data. Nat Methods 13 (2):165–170. https://doi.org/10.1038/nmeth.3688
7. Malone B, Atanassov I, Aeschimann F, Li X, Grosshans H, Dieterich C (2017) Bayesian prediction of RNA translation from ribosome profiling. Nucleic Acids Res 45 (6):2960–2972. https://doi.org/10.1093/nar/gkw1350
8. Zhang P, He D, Xu Y, Hou J, Pan BF, Wang Y, Liu T, Davis CM, Ehli EA, Tan L, Zhou F, Hu J, Yu Y, Chen X, Nguyen TM, Rosen JM, Hawke DH, Ji Z, Chen Y (2017) Genome-wide identification and differential analysis of translational initiation. Nat Commun 8 (1):1749. https://doi.org/10.1038/s41467-017-01981-8
9. Xiao Z, Huang R, Xing X, Chen Y, Deng H, Yang X (2018) De novo annotation and characterization of the translatome with ribosome profiling data. Nucleic Acids Res 46(10):e61. https://doi.org/10.1093/nar/gky179
10. Erhard F, Halenius A, Zimmermann C, L'Hernault A, Kowalewski DJ, Weekes MP, Stevanovic S, Zimmer R, Dolken L (2018) Improved Ribo-seq enables identification of cryptic translation events. Nat Methods 15 (5):363–366. https://doi.org/10.1038/nmeth.4631
11. Yoo AB, Jette MA, Grondona M (2003) SLURM: Simple linux utility for resource management. In: Job scheduling strategies for parallel processing. Springer, Berlin, Heidelberg, pp 44–60
12. Chothani S, Schafer S, Adami E, Viswanathan S, Widjaja AA, Langley SR, Tan J, Wang M, Quaife NM, Jian Pua C, D'Agostino G, Guna Shekeran S, George BL, Lim S, Yiqun Cao E, van Heesch S, Witte F, Felkin LE, Christodoulou EG, Dong J, Blachut S, Patone G, Barton PJR, Hubner N, Cook SA, Rackham OJL (2019) widespread translational control of fibrosis in the human heart by RNA-binding proteins. Circulation 140(11):937–951. https://doi.org/10.1161/CIRCULATIONAHA.119.039596
13. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S (2014) InterProScan 5: genome-scale protein function classification. Bioinformatics 30(9):1236–1240. https://doi.org/10.1093/bioinformatics/btu031
14. Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8(10):785–786. https://doi.org/10.1038/nmeth.1701
15. Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305(3):567–580. https://doi.org/10.1006/jmbi.2000.4315

Chapter 15

uORF-seqr: A Machine Learning-Based Approach to the Identification of Upstream Open Reading Frames in Yeast

Pieter Spealman, Armaghan Naik, and Joel McManus

Abstract

The identification of upstream open reading frames (uORFs) using ribosome profiling data is complicated by several factors such as the noise inherent to the procedure, the substantial increase in potential translation initiation sites (and false positives) when one includes non-canonical start codons, and the paucity of molecularly validated uORFs. Here we present uORF-seqr, a novel machine learning algorithm that uses ribosome profiling data, in conjunction with RNA-seq data, as well as transcript aware genome annotation files to identify statistically significant AUG and near-cognate codon uORFs.

Key words uORF, Upstream open reading frame, Translational regulation, Near-cognate codon, Non-canonical start codon, Alternative translation initiation site

1 Introduction

Upstream open reading frames (uORFs) are *cis*-regulatory elements located within the transcript leader (previously referred to as the 5′ untranslated region or 5′ UTR) that modulate translation of downstream ORFs. uORFs are composed of three components—a translation initiation site (TIS), the uORF body, and an in-frame stop codon [1]. Traditionally, uORFs have been characterized as translational repressors, wherein the ribosome initiates at the uORF start codon, the body is translated, and finally translation terminates at the stop codon and the ribosome dissociates from the transcript. Not only does this prevent the translation of the downstream protein coding region, it can also induce other regulatory pathways such as nonsense-mediated decay (NMD) [1, 2]. However, uORFs can also perform more complex regulation, as seen in the yeast *GCN4* gene. Here multiple aspects of translation such as the phosphorylation of eIF2α, the rate of ribosome initiation, and the distance to the downstream protein coding region all

Vyacheslav M. Labunskyy (ed.), *Ribosome Profiling: Methods and Protocols*, Methods in Molecular Biology, vol. 2252,
https://doi.org/10.1007/978-1-0716-1150-0_15,

contribute to *GCN4* translation under stress [3, 4]. While substantial work has established uORF position, relative to the downstream ORF, as contributing to the regulatory effect [4], more recently, research has begun to identify the roles of uORF coding regions in translation regulation [5–7].

It has long been known that ribosomes can initiate on codons other than the canonical start triplet AUG, albeit with lower efficiency [8–10]. The most well studied of these non-AUG codons are the near-cognate codons (NCCs) which differ from AUG by one nucleotide. Generally, NCCs rely on the same initiator tRNA (Met-tRNAi) used by AUG and therefore produce peptide sequences with methionine substitutions instead of the endogenous encoded amino acid [10, 11] although this is not always the case [10–12]. Importantly, NCC usage increases dramatically under stress, concomitant with increased translation of upstream open reading frames (uORFs) [13, 14] and N-terminal extensions (NTEs) [15]. This increase in NCC utilization may derive from stress-induced post-translational modifications (PTMs) of initiation factors, such as eIF1, eIF2α, eIF4F, and eIF4G [16–20]. Additionally, some environmental conditions may alter NCC utilization by direct physicochemical effect [21].

Using ribosome profiling data to identify potential uORFs dates back to the very introduction of ribosome profiling [13] where observations of ribosomes within transcript leaders (TLs) and downstream of start codons were interpreted as evidence of translation. These identifications were largely based on a case-by-case manual analysis of the data and often required the presence of an AUG start codon. The inclusion of NCC as start codons greatly expands the number of potential start codons; however it also greatly increases the likelihood of false identification, which is already a significant challenge, particularly for sparsely ribosome-occupied transcripts. Indeed, analyses of ribosome profiling data have resulted in thousands of NCC uORF predictions [13, 14, 22–24]. However, without statistical control, the functional significance of these predicted uORFs is difficult to assess.

It is worth noting that library preparation performed with cycloheximide (CHX) can increase the number of ribosomes recovered from transcript leaders, depending on species [25]. While this can create artifacts [26], because CHX prevents elongation but not initiation, it also amplifies the signal of ribosomes present at translation initiation sites [23]. We have previously shown increased accuracy in uORF identification in CHX-treated data [27], as exemplified by the detection of only 2 of 5 of *GCN4*s molecularly validated uORFs without CHX treatment, while all 5 uORFs were identified when CHX-treated samples were used.

Numerous programs exist that make use of ribosome profiling data to identify actively translated ORFs, such as ORF-RATER [28], RiboTaper [29], and RiboCode [30]. These programs often

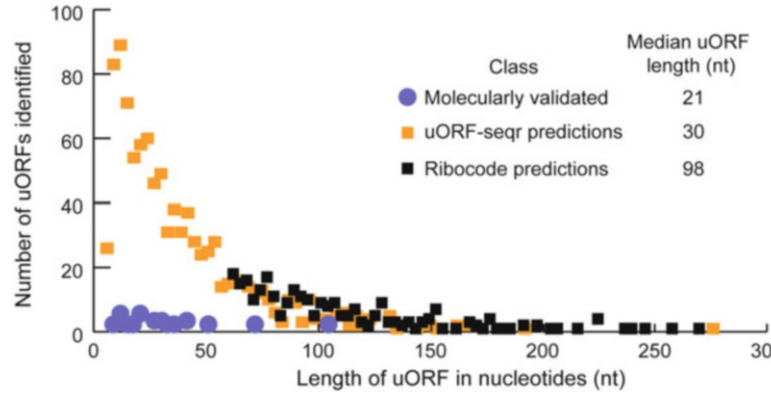

Fig. 1 Distribution of uORF sizes identified using a general ORF identifier (RiboCode) versus uORF-seqr. Here, we show three distributions of uORF sizes from *Saccharomyces cerevisiae*. These include 22 molecularly validated uORFs (blue dots) [27], with a median length of 21nt. Similarly, predicted uORFs from RiboCode (black squares, median length 98nt) and uORF-seqr (orange squares, median length 30nt) are also plotted

rely on features derived from protein coding ORFs, which may have unintended consequences in the identification of shorter ORFs such as uORFs. Indeed, a comparison of RiboCode to uORF-seqr using our published 2018 data found that RiboCode, which relies on 3nt periodicity, was unable to resolve uORFs shorter than 50nt in length (Fig. 1). This is especially relevant given that the median length of molecularly validated uORFs in yeast is 21nt, while the median length of RiboCode-identified uORFs was 90nt.

uORF-seqr is distinct from these general ORF identification methods as it does not assume that the distribution of ribosomes across a uORF will be similar to that observed on a long protein coding ORF. Indeed, we feel that this assumption is unfounded, given what is known about the difference between ribosome-associated proteins during early elongation phase (mostly initiation factors) and late elongation phase (elongation factors) and the time-dependent manner by which initiation factors dissociate and elongation factors associate [31, 32]. Notably, the rate of initiation factor dissociation is not equal for all factors involved [33]. Therefore, early phase ribosomes have a heterogeneous assembly of factors associated that could alter rates of elongation and distributions of ribosomes. Because of this, uORF-seqr only uses features derived from uORFs.

Here, we present uORF-seqr, a novel machine learning algorithm that uses ribosome profiling data, in conjunction with RNA-seq data, as well as transcript aware genome annotation files, to identify statistically significant AUG and NCC-uORFs [27].

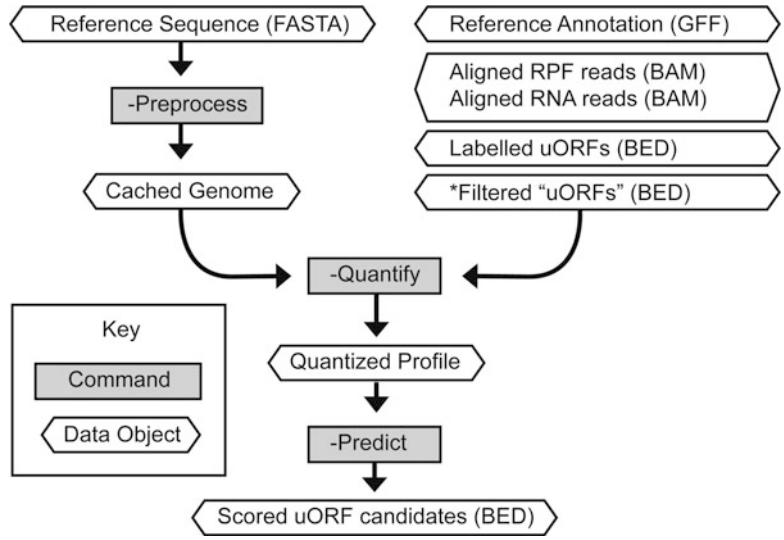

Fig. 2 Overview of uORF-seqr workflow. This schematic represents the standard uORF-seqr workflow, starting with preprocessing the genome, followed by quantification of expression data into a quantized profile used in the prediction step to generate a scored set of uORF candidates

uORF-seqr was originally designed for the identification of uORFs in the budding yeast clade *Saccharomyces* and has been used to identify hundreds of high-confidence uORFs across *S. cerevisiae*, *S. paradoxus*, and *S. eubayanus* [27]. This included the molecular validation of six out seven novel predicted uORFs, including four NCC-uORFs.

1.1 Overview of Workflow

As shown in Fig. 2, uORF-seqr requires a FASTA reference sequence to build a map of the chromosomes in a cached file using the "preprocess" command. GFF reference annotations, paired RPF and RNA bam files (from triplicate experiments), and a set of known good uORFs are required to quantify occupancy and derive feature weights using the "quantify" command. Known false positives or confounding elements, such as those produced by N-terminal extensions [34], can also be filtered at this step. The quantify command produces numerous quantized profiles that are then used to identify high-confidence candidate uORFs using the "predict" command.

1.2 P-site Fractionation for Reading Frame Determination

Because the reading frame of a ribosome dictates what is being translated, the determination of the reading frame is of utmost importance [13, 23, 35]. To determine the reading frame, one intuitive choice is to estimate the P-site of the translating ribosome, as it holds the tRNA being base-paired with a given anti-codon. While methods of P-site estimation vary, most rely on a given offset relative to the most upstream (5′) mapped nucleotide of a read

[13]. Given that perfectly sized fragments (i.e., 28nt) may make a minority of total fragment lengths (possibly in a protocol- and expression-dependent manner), additional adjustments are often made based on the fragment sizes in a given dataset [36–38].

Here, we use P-site fractionation for reading frame determination, such that the fraction of a P-site is allotted to nucleotides given the size of the aligned length (i.e., "soft match" nucleotides are not counted towards the fragment length). For example, for a read with a 28nt aligned length (a "28mer"), the entirety of the P-site fraction is placed 12nt downstream from the 5′-end. A 29nt read could include anomalous bases on either 5′ or 3′ ends; these are not easily distinguished, and so we mark both possible P-sites with a half probability of a read. With similar reasoning, longer reads are fractionally assigned to more than one p-site.

1.3 Regression Feature Definition

All candidate uORFs (cuORFs) require an AUG or NCC start codon located within a transcript leader and are evaluated and scored using 18 features (Table 1). These features are used in the identification of candidate uORFs (cuORFs) as described below. These include 12 "expression" features that are derived from the ribosome profiling and RNA-seq data and 6 "positional" features that take into account the local genomic context of the cuORF (Fig. 3). For each uORF the values of features are calculated. Values for these features are calculated per replicate within "newattempt.py" and "newcompute_features.py" files and include mRNA and RPF read counts, as well as P-site fractions.

Because translating ribosomes have several features similar to waves (e.g., amplitude, frequency, periodicity), several of our features make use of a discrete Fourier transform (DFT) of the ribosome footprint profile taken over the cuORF. This is calculated using the real fast Fourier transform in NumPy's fast Fourier transform package (numpy.fft.rfft). These include amplitude (in units of P-site fractional ribosome occupancy): (a) the maximum amplitude of any nontrivial frequency (i.e., the maximum proportion of P-site ribosome occupancies associated to a particular nucleotide step size), (b) the corresponding phase (in radians; capturing the relative offset from the start codon of the cuORF), (c) the maximum amplitude and (d) phase of the within-frame frequency (as for (a, b), but for a given frame), and (e) the nucleotide spacing of the frequency of (a,b). To normalize for abundance differences among replicate libraries, copies of cuORF features (a,c) were divided by the fraction of total P-site coverage of the corresponding TL out of the total read depth deriving the unitless features (f) and (g), respectively. Using (f) as an example of this, we first calculate (a) or "within_power_of_max_power_freq," as the power spectra of the maximum power spectra of the p-site frequency. We then normalize this respective to the sample by calculating the fraction of p-sites within the TL relative to the total p-sites within the sample,

Table 1
List of expression features (a-k) and positional features (m-r) represented in Fig. 3

Fig. 3 code	Name
a	Maximum amplitude
b	Phase of maximum amplitude frequency
c	In-frame to uORF maximum amplitude
d	In-frame to uORF phase of frequency
e	Nucleotide spacing of frequency with maximum amplitude
f	Normalized maximum amplitude (a)
g	Normalized in-frame to uORF maximum amplitude (c)
h	Normalized median amplitude
i	Normalized upstream abundance
j	Normalized downstream abundance
k	Maximum normalized amplitude anywhere with the transcript leader
l	Relative start magnitude
m	Length of uORF
n	Distance to transcription start site (TSS)
o	Distance to main ORF translation initiation site (TIS)
p	Starts within an N-terminal extension
q	Stops within an N-terminal extension
r	Percent overlap with main ORF

such that $(f) = (a)/(\text{number\_of\_p-sites\_in\_TLS}/\text{number\_of\_p-sites\_in\_sample})$. The median amplitude across nontrivial frequencies was also normalized (as above), generating (h) the normalized median amplitude. Features (i,j) are the normalized (as above) sum of P-site occupancy upstream and downstream of the cuORF, respectively. A similar DFT was performed on the whole of the TL, allowing for the calculation of (k) the maximum amplitude of any nontrivial frequency (analogous to feature (f)). For each cuORF, we also calculated (I) the "relative start magnitude" [14]: the normalized number of P-sites from the first nucleotide of the start codon minus the normalized average P-site coverage of all nucleotides two codons upstream, normalized as above. "Positional" features were (m) the cuORF length, (n) distance from the TSS to the cuORF start, (o) distance between the cuORF end and main ORF start, whether (p) the cuORF starts within a possible N-terminal extension region of the main ORF, whether (q) the possible N-terminal extension region spans the closest in-frame

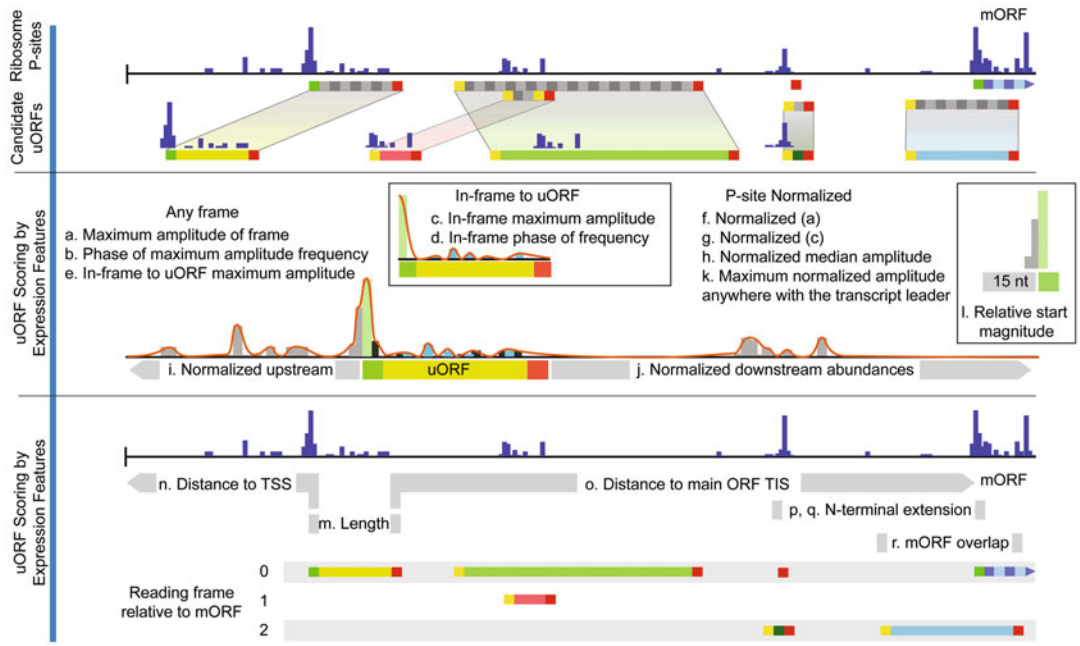

Fig. 3 Overview of uORF scoring and resolution using uORF-seqr. Representative figure of candidate uORF scoring using expression and positional features (Table 1)

(relative to the main ORF TIS) upstream stop codon to the main ORF TIS, and finally, (r) the fraction of the length of the cuORF that does not overlap the main ORF.

1.4 uORF-seqr Regression and Statistical Control

For each experiment, a model was trained to predict the fraction of (biological) replicates within which each cuORF was detected. Because ribosome profiling and RNA-seq data can greatly vary between labs, protocols, and conditions, the particular weights of the regression features are calculated for each sample. We assumed that the most valuable features are those that are both predictive and perform consistently between replicates. As such features are weighted so to maximize predictive performance across replicates.

To avoid the confounding potential of repeat-counting of the same ribosomes in different reading frames/cuORFs, we limit our training set to the subset of non-overlapping cuORFs. An

L1-penalized sparse binomial regression (generalized linear model with a log-linkage; R glmpath package) was fit by tenfold stratified cross-validation over the set of non-overlapping cuORFs to determine the L1 penalization parameter by minimizing holdout squared error. Stratification was performed per gene such that all cuORFs of the same gene were either held-out or held-in for any fold. The final regression used the minimizing penalization parameter and was fit to all of the cuORFs in the non-overlapping set. Each cuORF, including those with potential overlaps, was assessed with this regression in turn.

Given the intrinsic noise of ribosome profiling and RNA-seq, the mere identification of ribosome protected fragments is not indicative of translation. To account for this, we apply a statistical control to every cuORF wherein we compare it to regressions trained with a null model consisting of randomly permuted ribosome protection data and to a reference distribution made by random permutations of the features.

2 Materials

2.1 Data

1. Reference genome: This should include both FASTA format genome sequence file and matching GFF3 genome annotation file. Genome annotations should include transcript leaders (aka 5′ UTRs) and 3′ UTRs (*see* **Notes 1–3**).
2. Aligned reads: One should have paired ribosome profiling and RNA-seq reads for each replicate. These need to be aligned using STAR [39]. It is strongly advised that a minimum of three replicates should be used. Using fewer replicates is known to decrease the accuracy of predictions (*see* **Note 4**).
3. uORF training set: uORF-seqr uses features and values derived from a small set of molecularly validated uORFs to train the identification model. uORF-seqr comes with training sets for *S. cerevisiae*, *S. paradoxus*, *and S. eubayanus*. However, work outside of these species will require a user-supplied 6-column bed file. This uORF training set would ideally be a list of molecularly validated uORFs of various lengths and start codon types which are occupied by ribosomes within your dataset. Manually selected uORFs should be avoided as they often fit a preconception of what a uORF looks like and can directly introduce this bias during training. In our experience a minimal training set should be at least ten uORFs.

2.2 Computational Resources

Linux OS (CentOS 7+ or Ubuntu 16+) with:

Python (2.7.x) [40], R (3.2.3+) [41], Samtools (1.6+) [42], Git (2.7+) [43], Pip (9.0.1+ for python 2.7) [44], TkInter [45], NumPy [46].

Installing on Ubuntu 16+:

```
sudo apt-get update
sudo apt-get upgrade
sudo apt-get install python
sudo apt-get install r-base
sudo apt-get install samtools
sudo apt-get install git
sudo apt-get install python-tk
sudo apt-get install python-numpy
sudo apt-get install python-pip
```

2.3 R Requirements

Glmpath [47]

2.4 Python Package Requirements

Rpy2 [48], Pysam [42], Statsmodels [49], Scikit-Learn [50], GffUtils [51], Matplotlib [52], SciPy [53]

```
pip install rpy2==2.8.6
pip install pysam
pip install -U statsmodels
pip install scikit-learn==0.17.1
pip install gffutils
pip install matplotlib
pip install scipy
```

3 Methods

3.1 Installation of uORF-seqr

1. Install uORF-seqr using GitHub:

```
git clone https://github.com/pspealman/uorfseqr.git
```

2. Verify that the host machine is set up correctly and meets the necessary requirements using the "test" option.

```
cd uorfseqr
python uorfseqr.py -test
```

3. To run a demonstration of the most common commands and appropriate syntax, use the "demo" option.

```
cd to uorfseqr.py
python uorfseqr.py -demo
```

3.2 Preprocess a New Reference Genome

Preprocessing a reference genome creates a directory (as defined by genome_name) and creates a per-chromosome cache of sequences in a binary format. This only needs to occur once per reference genome.

Usage:

```
python uorfseqr.py -preprocess -genome_name <str> -fa <path to reference fasta file>
```

Example:

```
python uorfseqr.py - preprocess -genome_name s_cerervisiae -fa data/reference_genomes/Scer_SacCer3.fa
```

3.3 Quantify Features for a New Experiment

The quantify command scores the features of known uORFs across three replicates as well as null models. It relies on the cached processed genome from the preprocess task. It will create the output directory, several subdirectories, and a command file called "analysis.json."

Numerous optional commands are available for the end-user to alter the performance of the analysis.

Usage:

```
python uorfseqr.py -quantify -genome_name <str> -samples <sample_name> <path to sample RPF.bam> <path to sample mRNA.mRNA> -o<output_dir>
```

Example:

```
python uorfseqr.py -quantify -genome_name s_cerervisiae -samples Scer_A ../data/bam/Scer_A_RPF.bam ../data/bam/Scer_A_mRNA.bam Scer_B ../data/bam/Scer_B_RPF.bam ../data/bam/Scer_B_mRNA.bam Scer_C ../data/bam/Scer_C_RPF.bam ../data/bam/Scer_C_mRNA.bam -o scer.demo
```

3.3.1 Short Descriptions of Arguments

GFF files. General feature format (gff) files are used to identify coordinates for transcript leaders (aka 5′ UTRs), 3'UTRs, and genes (*see* **Notes 1 and 2**). By default we use the GFF from Spealman and Naik [27] for *Saccharomyces cerevisiae* (*see* **Note 3**), located in 'data/reference_genomes/saccharomyces_cerevisiae.gff'.

Users can set their own gff file by using the -gff option:

```
-gff data/reference_genomes/saccharomyces_cerevisiae.gff
```

Filter regions. Genomic regions can be filtered from feature scoring to prevent counterproductive learning. Users can set their own genes or regions to be filtered by loading one or more bed files containing the coordinates of regions to be filtered. If no file is

defined by the end-user, the scer_baduorf file can be loaded with numerous false positives identified using *S. cerevisiae* S288C grown under rich media conditions; this file is located at "data/labelled_uorfs/scer_baduorf.bed."

N-terminal extensions (NTEs), given that they are very similar in regard to uORFs, identical with the exception that they are both in-frame and lack an intervening stop codon with the main ORF, are common sources for false positives. To aid in the identification of uORFs, we also include "nte-seqr" an optional tool for the identification of potential NTEs, so that they may be filtered (*see* **Note 5**).

```
-filter data/labelled_uorfs/scer_baduorf.bed
```

Known uORFs. A list of positively labelled uORFs is required to train and evaluate feature scoring. This list is provided as a BED format file containing the uORF from start codon to stop codon and strand. By default, the file loaded contains 16 molecularly validated uORFs identified in *S. cerevisiae* before 2017 (located at "data/labelled_uorfs/STANDARD-golden.bed"). Alternatively, we provide a file that also includes the 6 uORFs predicted by uORF-seqr and subsequently verified [27] (located at "data/labelled_uorfs/*EXPANDED-golden.bed*") as well as a file that contains all 432 predicted *S. cerevisiae* uORFs from Spealman and Naik [27] (located at "data/labelled_uorfs/Spealman_Naik_2017.bed"). Users can set the known uORFs using the -known command.

```
-known data/labelled_uorfs/STANDARD-golden.bed
```

Ribosome footprint sizes. Users can set minimum (default 27) and maximum (default 33) sizes using the -min_rpf and -max_rpf commands, respectively.

```
-max_rpf 33
-min_rpf 22
```

Non-start near-cognate codons (NCCs). Near-cognate codons can act as translation initiation sites. NCCs differ from the canonical start codon "AUG" by one nucleotide. All NCCs have been observed functioning as start codons in vivo and vitro experiments, with the exception of "AGG" and "AAG." By default all NCCs are considered as potential start codons except "AGG" and "AAG." Users can alter this exception by using the -non_start option.

```
-non_start AGG AAG
```

Kozak consensus sequence. The sequence around potential start codons is scored based on similarity to the provided Kozak consensus sequence. By default this is "AAAAAAATGT" where the underlined sequence is the start codon [54].

```
-kozak_seq "AAAAAAATGT"
```

Main ORF mask. The start codons of some main ORFs exhibit ribosome pileups as ribosomes "queue" in preparation of initiation at the main ORF start codon. As these queues can contain many more ribosomes than are present within the rest of the transcript leader, they can present problems for scoring and identification. The main ORF mask filters a given number of nucleotides upstream of the main ORF start codon. By default this is set to 15 nt.

```
-morf_mask 15
```

Minimum 5′ UTR length. This allows for the filtering of any gene with an annotated 5′ UTR shorter than the given number of nucleotides. By default this is set to 15nt.

```
-min_utr 15
```

Minimum supporting RPF. This defines the minimum number of RPFs required to be considered. A potential uORF must still outperform the expectation at random in order to be predicted as a uORF. By default, this is set to 3.

```
-min_rpf_ct 3
```

3.4 Predict Candidate uORFs Based on Calculated Features

With the -predict command, we load the calculated features from the –quantify command and begin the construction of the regression and scoring of the potential uORFs. Those potential uORFs, as well as their respective qval scores, are reported in a bed file format in the "results" directory.

Usage:

```
python uorfseqr.py -predict -i <previous output_dir>
```

Example:

```
python uorfseqr.py -predict -i scer.demo
```

3.5 Interpreting Results of uORF Prediction

Numerous bed files are generated in the results directory by the -predict command. While each file contains information, the *-candidate_uORF.bed file is the strictest and highest confidence of the predictions. For your convenience this is copied into the user

provided –input_directory. This file is a renamed copy of the *-significant-resolved-qvals.bed file in the results' directory.

The *-candidate_uORF.bed file generated by uORF-seqr is a standard six-column bed format that contains tab-delimited lines such as:

```
chrV 140112 140124 YEL009C.140124 0 -
```

These columns are chromosome, region start, region stop, uORF_ID, Qval_score, and strand. Note that the lower the Qval score, the higher the confidence in the prediction.

4 Notes

1. uORF-seqr requires each gene that will be evaluated to have at least two entries in the GFF file. One entry is for the main ORF protein coding region or gene; we denote this by using the "gene" label for the feature type. The other required entry is the transcript leader or 5′ UTR; this is denoted using the "five_prime_UTR" label for the feature type. These are associated with each other by ID and PARENT Attributes, such that the "five_prime_UTR" entry must have a PARENT field whose value matches the ID field of the "gene" entry. Genes without paired "five_prime_UTR" entries are not evaluated for uORFs. 3′ UTRs can also be defined and will aid in the assignment of intergenic regions to specific genes. These can be denoted using the "three_prime_UTR" label for the feature type.
2. uORF-seqr does not currently support splice-aware transcript annotations. All transcripts are assumed to be contiguous without splicing.
3. The current default genome feature file (.gff) contains annotations not present in the genome annotations available from either NCBI or ENSEMBL. These annotations include particular features, such as "transcription_start_site" and "polyA_site," as well as certain genes and non-coding RNA.
4. Fewer replicates severely hinder the accuracy of predictions and should only be performed for the purpose of data exploration. Anyone wanting to explore their data can use the "pseudo_triplicate" command to generate three pseudo-replicates from one or two replicates.

   Usage:

```
python uorfseqr.py -pseudo -rep1 <replicate_1.bam> -rep2 <replicate_2.bam> -o
<output_filename_prefix>
```

Example:

```
python uorfseqr.py -pseudo -rep1 ../data/bam/Scer_A_RPF.bam -rep2 ../data/bam/
Scer_B_RPF.bam -o ../data/bam/Scer_RPF
python uorfseqr.py -pseudo -rep1 ../data/bam/Scer_A_mRNA.bam -rep2 ../data/bam/
Scer_B_mRNA.bam -o ../data/bam/Scer_mRNA
```

5. N-terminal extensions are similar to uORFs in that they are upstream alternative translation initiation sites only they are both in-frame to the main ORF and lack an intervening stop codon [55–58]. The NTE-seqr tool can be used to identify potential N-terminal extension events before running uORF-seqr; the generated bed file can be used here in conjunction with uORF-seqr for additional filtering.

   Briefly, NTE-seqr attempts to identify N-terminal extensions by first finding all regions upstream of main ORF start codons and the nearest in-frame upstream stop codon. These search regions are then scanned to identify genes with large numbers of in-frame ribosomes. Search regions are also scanned for AUG and NCC start codons. We presume that the start codon most likely to function as the initiation site will have a confluence of features: higher relative start magnitude [14], higher relative translational efficiency, and a significant fraction of total in-frame ribosomes.

   Usage:

```
python uorfseqr.py -nte -samples <sample_name> <path to sample RPF.bam> <path to
sample mRNA.mRNA> -o <output_filename_prefix>
```

Example:

```
python uorfseqr.py -nte -samples Scer_A ../data/bam/Scer_A_RPF.bam ../data/bam/
Scer_A_mRNA.bam Scer_B ../data/bam/Scer_A_RPF.bam ../data/bam/Scer_A_mRNA.bam Scer_C
../data/bam/Scer_C_RPF.bam ../data/bam/Scer_C_mRNA.bam -o ../data/bam/Scer_NTE
```

   The resulting output is a BED format file containing potential NTE events. These can be filtered by including this file in the *-filter_uorfs* option during the *-quantify* step.

   Genome reference FASTA and gff files function similarly to uORF-seqr; they can be defined using the *-fa* and *-gff* options, respectively. The gff assumes "gene" mark protein coding regions, while transcript leaders are assigned "five_prime_utr." These can be changed using the *-gene_tag* and *-transcript_leader_tag*, respectively.

References

1. Hinnebusch AG, Ivanov IP, Sonenberg N (2016) Translational control by 5-′-untranslated regions of eukaryotic mRNAs. Science 352(6292):1413–1416. https://doi.org/10.1126/science.aad9868
2. Kozak M (1989) The scanning model for translation: an update. J Cell Biol 108 (2):229–241. https://doi.org/10.1083/jcb.108.2.229
3. Hinnebusch AG, Jackson BM, Mueller PP (1988) Evidence for regulation of reinitiation in translational control of GCN4 mRNA. Proc Natl Acad Sci USA 85(19):7279–7283. https://doi.org/10.1073/pnas.85.19.7279
4. Hinnebusch AG (2005) Translational regulation of GCN4 and the general amino acid control of yeast. Annu Rev Microbiol 59:407–450. https://doi.org/10.1146/annurev.micro.59.031805.133833
5. Gunišová S, Beznosková P, Mohammad MP, Vlčková V, Valášek LS (2016) In-depth analysis of cis-determinants that either promote or inhibit reinitiation on GCN4 mRNA after translation of its four short uORFs. RNA 22 (4):542–558. https://doi.org/10.1261/rna.055046.115
6. Mohammad MP, Munzarová Pondelícková V, Zeman J, Gunišová S, Valášek LS (2017) In vivo evidence that eIF3 stays bound to ribosomes elongating and terminating on short upstream ORFs to promote reinitiation. Nucleic Acids Res 45(5):2658–2674. https://doi.org/10.1093/nar/gkx049
7. Lin Y, May GE, Kready H, Nazzaro L, Mao M, Spealman P et al (2019) Impacts of uORF codon identity and position on translation regulation. Nucleic Acids Res 47(17):9358–9367. https://doi.org/10.1093/nar/gkz681
8. Zitomer RS, Walthall DA, Rymond BC, Hollenberg CP (1984) *Saccharomyces cerevisiae* ribosomes recognize non-AUG initiation codons. Mol Cell Biol 4(7):1191–1197. https://doi.org/10.1128/mcb.4.7.1191
9. Clements JM, Laz TM, Sherman F (1988) Efficiency of translation initiation by non-AUG codons in *Saccharomyces cerevisiae*. Mol Cell Biol 8(10):4533–4536. https://doi.org/10.1128/mcb.8.10.4533
10. Peabody DS (1989) Translation initiation at non-AUG triplets in mammalian cells. J Biol Chem 264(9):5031–5035. PMID: 2538469
11. Schwab SR, Shugart JA, Horng T, Malarkannan S, Shastri N (2004) Unanticipated antigens: translation initiation at CUG with leucine. PLoS Biol 2(11):e366. https://doi.org/10.1371/journal.pbio.0020366
12. Starck SR, Jiang V, Pavon-Eternod M, Prasad S, McCarthy B, Pan T et al (2012) Leucine-tRNA initiates at CUG start codons for protein synthesis and presentation by MHC class I. Science 336(6089):1719–1723. https://doi.org/10.1126/science.1220270
13. Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324 (5924):218–223. https://doi.org/10.1126/science.1168978
14. Brar GA, Yassour M, Friedman N, Regev A, Ingolia NT, Weissman JS (2012) High-resolution view of the yeast meiotic program revealed by ribosome profiling. Science 335 (6068):552–557. https://doi.org/10.1126/science.1215110
15. Gerashchenko MV, Lobanov AV, Gladyshev VN (2012) Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress. Proc Natl Acad Sci U S A 109(43):17394–17399. https://doi.org/10.1073/pnas.1120799109
16. Sheikh MS, Fornace AJ Jr (1999) Regulation of translation initiation following stress. Oncogene 18(45):6121–6128. doi:https://doi.org/10.1038/sj.onc.1203131
17. Hinnebusch AG, Lorsch JR (2012) The mechanism of eukaryotic translation initiation: new insights and challenges. Cold Spring Harb Perspect Biol 4(10). https://doi.org/10.1101/cshperspect.a011544
18. Liu B, Qian S-B (2014) Translational reprogramming in cellular stress response. Wiley Interdiscip Rev RNA 5:301–305. https://doi.org/10.1002/wrna.1212
19. Vaidyanathan PP, Zinshteyn B, Thompson MK, Gilbert WV (2014) Protein kinase A regulates gene-specific translational adaptation in differentiating yeast. RNA 20(6):912–922. https://doi.org/10.1261/rna.044552.114
20. Kearse MG, Wilusz JE (2017) Non-AUG translation: a new start for protein synthesis in eukaryotes. Genes Dev 31(17):1717–1731. https://doi.org/10.1101/gad.305250.117
21. Chengguang H, Sabatini P, Brandi L, Giuliodori AM, Pon CL, Gualerzi CO (2017) Ribosomal selection of mRNAs with degenerate initiation triplets. Nucleic Acids Res 45 (12):7309–7325. https://doi.org/10.1093/nar/gkx472

22. Ingolia NT, Lareau LF, Weissman JS (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell 147(4):789–802. https://doi.org/10.1016/j.cell.2011.10.002
23. Andreev DE, O'Connor PBF, Loughran G, Dmitriev SE, Baranov PV, Shatsky IN (2017) Insights into the mechanisms of eukaryotic translation gained with ribosome profiling. Nucleic Acids Res 45(2):513–526. https://doi.org/10.1093/nar/gkw1190
24. Lee S, Liu B, Lee S, Huang S-X, Shen B, Qian S-B (2012) Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. Proc Natl Acad Sci USA 109((37):E2424–E2432. https://doi.org/10.1073/pnas.1207846109
25. Sharma P, Nilges BS, Wu J, Leidel SA (2019) The translation inhibitor cycloheximide affects ribosome profiling data in a species-specific manner. bioRxiv:746255. https://doi.org/10.1101/746255
26. Hussmann JA, Patchett S, Johnson A, Sawyer S, Press WH (2015) Understanding biases in ribosome profiling experiments reveals signatures of translation dynamics in yeast. PLoS Genet 11(12):e1005732. https://doi.org/10.1371/journal.pgen.1005732
27. Spealman P, Naik AW, May GE, Kuersten S, Freeberg L, Murphy RF et al (2018) Conserved non-AUG uORFs revealed by a novel regression analysis of ribosome profiling data. Genome Res 28(2):214–222. https://doi.org/10.1101/gr.221507.117
28. Fields AP, Rodriguez EH, Jovanovic M, Stern-Ginossar N, Haas BJ, Mertins P, Raychowdhury R, Hacohen N, Carr SA, Ingolia NT, Regev A, Weissman JS (2015) A regression-based analysis of ribosome-profiling data reveals a conserved complexity to mammalian translation. Mol Cell 60:816–827. https://doi.org/10.1016/j.molcel.2015.11.013
29. Calviello L, Mukherjee N, Wyler E, Zauber H, Hirsekorn A, Selbach M, Landthaler M, Obermayer B, Ohler U (2016) Detecting actively translated open reading frames in ribosome profiling data. Nat Methods 13:165–170. https://doi.org/10.1038/nmeth.3688
30. Xiao Z, Huang R, Xing X, Chen Y, Deng H, Yang X (2018) De novo annotation and characterization of the translatome with ribosome profiling data. Nucleic Acids Res 46:e61. https://doi.org/10.1093/nar/gky179
31. Panse VG, Johnson AW (2010) Maturation of eukaryotic ribosomes: acquisition of functionality. Trends Biochem Sci 35:260–266. https://doi.org/10.1016/j.tibs.2010.01.001
32. Karbstein K (2013) Quality control mechanisms during ribosome maturation. Trends Cell Biol 23:242–250. https://doi.org/10.1016/j.tcb.2013.01.004
33. Hershey JWB, Mathews MB, Sonenberg N (1996) Translational control. Cold Spring Harbor Laboratory
34. Wethmar K (2014) The regulatory potential of upstream open reading frames in eukaryotic gene expression. Wiley Interdiscip Rev RNA 5 (6):765–778. https://doi.org/10.1002/wrna.1245
35. Spealman P, Wang H, May G, Kingsford C, McManus CJ (2016) Exploring ribosome positioning on translating transcripts with ribosome profiling. Methods Mol Biol 1358:71–97. https://doi.org/10.1007/978-1-4939-3067-8_5
36. Legendre R, Baudin-Baillieu A, Hatin I, Namy O (2015) RiboTools: a Galaxy toolbox for qualitative ribosome profiling analysis. Bioinformatics 31(15):2586–2588. https://doi.org/10.1093/bioinformatics/btv174
37. Wang H, Kingsford C, McManus CJ (2018) Using the Ribodeblur pipeline to recover A-sites from yeast ribosome profiling data. Methods 137:67–70. https://doi.org/10.1016/j.ymeth.2018.01.002
38. Lauria F, Tebaldi T, Bernabò P, Groen EJN, Gillingwater TH, Viero G (2018) riboWaltz: Optimization of ribosome P-site positioning in ribosome profiling data. PLoS Comput Biol 14(8):e1006169. https://doi.org/10.1371/journal.pcbi.1006169
39. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S et al (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 29:15–21. https://doi.org/10.1093/bioinformatics/bts635
40. Python 2.7 (2010). https://www.python.org/downloads/. Accessed 9 April 2020
41. R 3.2.3 (2015). https://cran.r-project.org/. Accessed 9 April 2020
42. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352
43. Git 2.7 (2016). https://git-scm.com/downloads. Accessed 9 April 2020
44. Pip 9 (2016). https://pip.pypa.io/en/stable/. Accessed 9 April 2020

45. TkInter (2016). https://wiki.python.org/moin/TkInter. Accessed 9 April 2020
46. NumPy (2016). https://numpy.org/. Accessed 9 April 2020
47. Mee Young Park, Trevor Hastie (2018). Package 'glmpath'. https://cran.r-project.org/web/packages/glmpath/glmpath.pdf. Accessed 9 April 2020
48. Rpy2 (2016). https://rpy2.github.io/. Accessed 9 April 2020
49. Seabold S, Perktold J (2010) Statsmodels: econometric and statistical modeling with Python. In: Proceedings of the 9th Python in Science Conference. SciPy, pp 92–96. https://doi.org/10.25080/Majora-92bf1922-011
50. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V et al (2011) Scikit-learn: Machine learning in Python. J Mach Learn Res 12:2825–2830
51. Dale R. (2013). https://pythonhosted.org/gffutils/. Accessed 9 April 2020
52. Hunter JD (2007) Matplotlib: a 2D graphics environment. Comput Sci Eng 9:90–95. https://doi.org/10.1109/MCSE.2007.55
53. Jones, E., Oliphant, T., Peterson, P., et al. (2001) SciPy: open source scientific tools for Python. http://www.scipy.org/. Accessed 9 April 2020
54. Hamilton R, Watanabe CK, de Boer HA (1987) Compilation and comparison of the sequence context around the AUG start codons in *Saccharomyces cerevisiae* mRNAs. Nucleic Acids Res 15(8):3581–3593. https://doi.org/10.1093/nar/15.8.3581
55. Wethmar K, Smink JJ, Leutz A (2010) Upstream open reading frames: molecular switches in (patho)physiology. Bioessays 32 (10):885–893. https://doi.org/10.1002/bies.201000037
56. Strubin M, Long EO, Mach B (1986) Two forms of the Ia antigen-associated invariant chain result from alternative initiations at two in-phase AUGs. Cell 47(4):619–625. https://doi.org/10.1016/0092-8674(86)90626-4
57. Tang H-L, Yeh L-S, Chen N-K, Ripmaster T, Schimmel P, Wang C-C (2004) Translation of a yeast mitochondrial tRNA synthetase initiated at redundant non-AUG codons. J Biol Chem 279(48):49656–49663. https://doi.org/10.1074/jbc.M408081200
58. Chang K-J, Wang C-C (2004) Translation initiation from a naturally occurring non-AUG codon in *Saccharomyces cerevisiae*. J Biol Chem 279(14):13778–13785. https://doi.org/10.1074/jbc.M311269200

Chapter 16

Genome-Wide Analysis of Actively Translated Open Reading Frames Using RiboTaper/ORFquant

Dermot Harnett, Eelco Meerdink, Lorenzo Calviello, Dominique Sydow, and Uwe Ohler

Abstract

Ribosome profiling, or Ribo-seq, provides precise information about the position of actively translating ribosomes. It can be used to identify open reading frames (ORFs) that are translated in a given sample. The RiboTaper pipeline, and the ORFquant R package, leverages the periodic distribution of such ribosomes along the ORF to perform a statistically robust test for translation which is insensitive to aperiodic noise and provides a statistically robust measure of translation. In addition to accounting for complex loci with overlapping ORFs, ORFquant is also able to use Ribo-seq as a tool for distinguishing actively translated transcripts from non-translated ones, within a given gene locus.

Key words Ribo-seq, Translation, Genomics, Open reading frame, Periodicity, Sequencing, Ribosome

1 Introduction

By measuring the density of translating ribosomes, Ribo-seq data provides an overview of the translational activity within a biological sample. As discussed in previous chapters, Ribo-seq provides single nucleotide resolution information on the position of active ribosomes. This results in a periodic distribution of Ribo-seq reads over translated open reading frames (ORFs), since translation progresses in discrete 3-nucleotide, codon-by-codon steps. As this periodicity is uniquely associated with translation, it can be exploited to identify translated ORFs in the presence of noise [1].

Examining the profiles of ribosomal footprints (RPFs) of different lengths reveals that they are shifted relative to one another—e.g., pile-ups of reads of 29nt might tend to occur 2nt upstream of reads of 27nt. To effectively integrate RPF data, an offset must be chosen for each read length such that their periodicity is brought into phase—often Ribo-seq data is summarized by a track of the presumptive first-nucleotide position of the P-site within each RPF

Vyacheslav M. Labunskyy (ed.), *Ribosome Profiling: Methods and Protocols*, Methods in Molecular Biology, vol. 2252, https://doi.org/10.1007/978-1-0716-1150-0_16,

(with the A-site then occurring 3bp downstream). With this accomplished, the simplest approach to analyze periodicity (exploited by, e.g. [2]) is to simply count reads that are "in frame." However, counts of "in-frame reads" are highly variable in the presence of noise [1]. Identifying periodicity in noisy data is a problem well studied outside of biology, and a reliable solution exists in the form of the "multitaper" method.

Defining translated open reading frames using this multitaper test allows for the identification of translated ORFs without the use of ad hoc cutoffs and is suited to applications where low-abundance translation events, such as those in upstream reading frames or those present in minority of cell in a bulk sample, are important (Fig. 1). The multitaper test for periodicity (Box 1) is implemented in the command line tool RiboTaper [1]. A limitation of applying the multitaper test to genomic RPF data is the inherent difficulty of attributing short read-derived RPF signal to distinct isoforms – when one isoform largely overlaps another, RiboTaper will simply annotate both. Many genes in higher eukaryotes possess multiple isoforms, although abundances are often unevenly distributed, with a single isoform accounting for the bulk of transcripts in most genes [3]. ORFquant [4] addresses this issue by quantifying the relative contribution of the various ORFs at a locus. This makes it a powerful tool to distinguish those transcripts actually producing protein from those which are degraded, confined to the nucleus, or otherwise not effectively translated. ORFquant can thus be used to reduce a complex transcriptome down to a comparatively simple active translatome, focusing analysis on those isoforms which are highly likely to generate functional proteins [4].

Box 1:
The periodicity of a discrete, finite signal such as that yielded by Ribo-seq can be extracted from its power spectrum, found by FFT (fast Fourier transform). This process, equivalent to counting in frame reads, has high variance if there is noise in the input signal. We might choose instead to take the power spectrum of discrete windows—say every 100 bp of our sequence. However, such a rectangular window function (or taper) would distort the Fourier transform due to "spectral leakage." To deal with this problem, a class of taper functions, used in fields such as astronomy and electrical engineering, known as "slepians" (or discrete prolate spheroidal sequences (DPSS)) are available. The multitaper method multiplies the signal by a set of orthogonal slepian functions to obtain independent estimates of the power spectrum. Furthermore, the variance of these estimates in the presence of uniform white noise is well defined, so that an *F*-test can be used to derive *P*-value for the signal's periodicity.

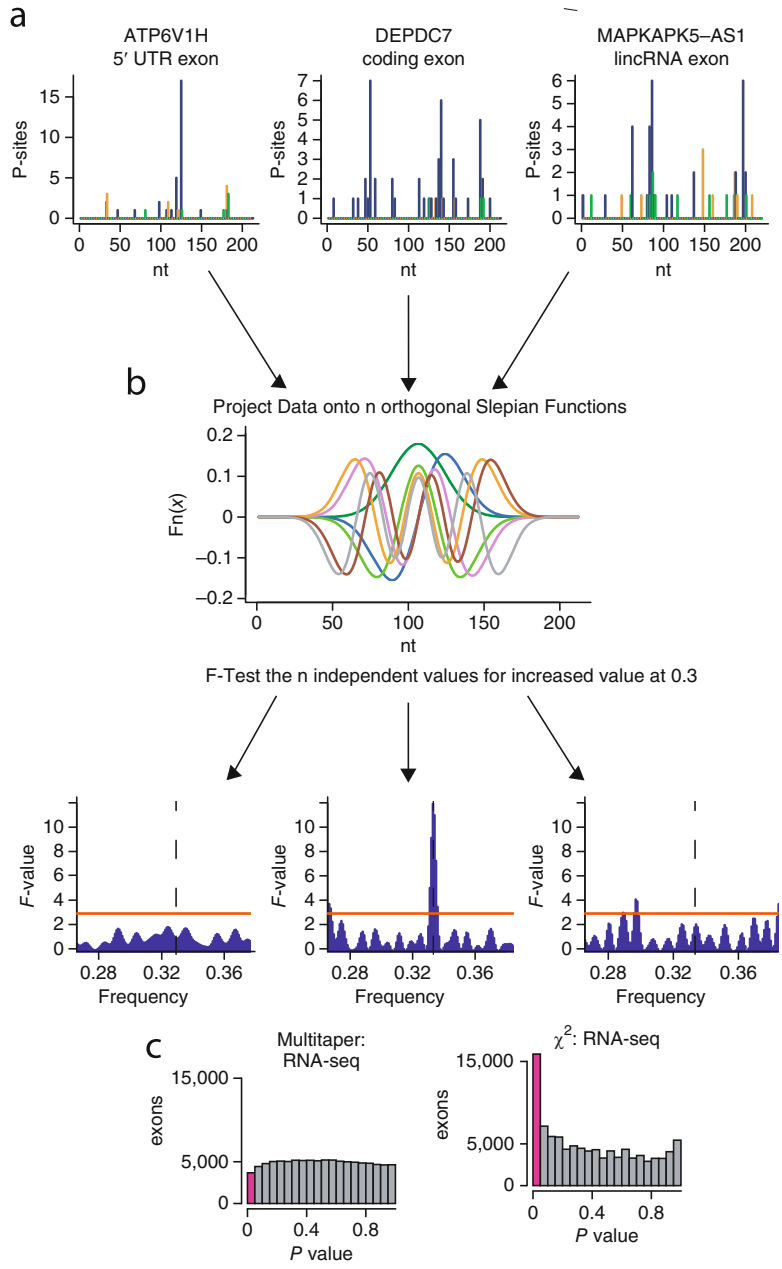

Fig. 1 Detecting periodic RPFs using a multitaper test. Ribosomal footprints over translated ORFs (**a**) show a characteristic periodic distribution, which reflects the codon-by-codon progress of translation. The various sources of noise mean that simply counting in-frame reads is an unreliable guide to which ORFs are translated. By multiplying the signal across the ORF by a set of orthogonal basis functions—slepians (**b**)—the multitaper test is able to derive power spectra which are sharply peaked at 0.33 nt for translated ORFs and close to uniform otherwise. This procedure yields a *P*-value distribution which is close to uniform for RNA-seq reads, as expected, unlike simple texts for in-frame bias (**c**). Taken from https:/doi.org/10.1101/031625

Here we will give instructions for the application of both RiboTaper and ORFquant, as a reimplementation of RiboTaper. Both RiboTaper and ORFquant perform multitaper tests on Ribo-seq data over candidate ORF regions. ORFquant's approach differs in that it attempts to identify a parsimonious set of ORFs which account for the Ribo-seq signal over a gene—it's results will be similar when a single isoform is present on a gene and more accurate where complex isoform structures exist. The procedure is described in detail in [4]; however in brief, ORFquant:

1. Collapses transcriptome variability independent of the translatome—e.g., transcripts differing only by their UTRs and transcripts fully nested within one another.
2. Collapses read counts to counts over features—exons and junctions—which provide information about isoform translation.
3. Eliminates isoforms using a greedy algorithm, for which all linked features are also explained by another isoform with more covered features, until a parsimonious set of ORFs is obtained.
4. Calculates scaling factors for each ORF, using coverage over features unique to that ORF where they exist and a prior assumption of uniform translation where they do not.
5. Applies cutoffs on scaling factors to identify (by default) ORFs accounting for 2% or more of the genes' translational activity

RiboTaper's output, detailed below, consists of a set .of files detailing the results of multitaper tests for ORFs within the provided CDS as well as within non-coding transcripts and for regions up- or downstream of annotated ORFs. ORFquant's output likewise consists of an annotation of the translatome, including information about canonical and novel ORFs, as well as nonstandard translation events such as upstream ORFs, downstream ORFs, translated retained introns, and read-through events.

2 Materials

1. A 64-bit computer running a Unix system with at least 8G of RAM (RAM requirements will depend on the size of the annotation, genome, and Ribo-seq data being processed—a machine with at least 8GB should be used for large genomes such as human or mouse. The `chunk_size` parameter can also be used to reduce memory footprint at the cost of computation time). RiboTaper requires at least two cores to run.
2. A working installation of R (version 3.5 or above). ORFquant is a package written in the R statistical programming language, and its API is simple enough that passing familiarity with the

language should be sufficient. A suitable development environment will also be required—we recommend RStudio, which can be downloaded, along with R, from www.rstudio.com.

3. To run ORFquant—a working installation of the ORFquant R package. This can be installed, along with dependencies, by opening an interactive R session and running using:

```
install.packages("devtools")
devtools::install_git("https://github.com/ohlerlab/
ORFquant")
```

4. To run RiboTaper—a working installation of RiboTaper. RiboTaper can be most easily installed on the Linux command line using the "conda" package manager (available at https://docs.conda.io/en/latest/miniconda.html). We recommend creating a dedicated conda environment for ORFquant and/or RiboTaper through the command line as follows:

```
conda create -c bioconda -n RiboTaper RiboTaper
conda activate RiboTaper
```

Alternatively, RiboTaper may also be downloaded from the following website: https://ohlerlab.mdc-berlin.de/software/RiboTaper_126/. Note that the following instructions will assume installation via conda.

Runtime: 30–60 min

5. A means of determining P-site offsets (i.e., the position of the P-site within each read) from RPF data. We recommend the RiboseQC package [5], which can be installed as follows in an interactive R session:

```
devtools::install_git("https://github.com/
ohlerlab/RiboseQC")
```

If another program is used to determine P-site offsets, the output should be processed into a tab-separated table with the headings "read_length," "cutoff," and "compartment," denoting the read length, the offset (distance between the 5′ footprint end and the inferred P-site), and the compartment (a string corresponding to a chromosome name from which the read originates, either "nucl" for a default applying to all reads or a chromosome such as "chrM" for which different offsets apply).

6. A GTF file containing valid gene annotation and an associated FASTA file containing gene sequences. ORFquant's use of GTF and not GFF files reflects the more restricted, and thus more easily parsed, nature of GTF files compared to GFF files. If only GFF files are available, the R packages GenomicFeatures and Rtracklayer can be used to import the GFF into R as a TxDb object and then export it as a GTF. The chromosome names in both genome and FASTA files need to be the same (we recommend removing, e.g., scaffolds with no annotation on them). The sequence in the FASTA file should be unmasked and contain only capital letters. Irregularities in gene annotations, particularly for non-model organisms, are by far the most frequent source of errors in the RiboTaper/ORFquant pipelines. The GTF file should follow the official GENCODE GTF standard format with metadata columns containing unique ids "gene_id" and "transcript_id" for genes and transcripts, as well as gene and transcript biotypes. For initial tests of the pipeline, particularly of large genomes, we recommend using a subset of the annotation, e.g., restricting analysis to chromosome 1.
7. Input Ribo-seq data. RiboTaper takes as input a BAM file, along with a set of offsets for each read length to be used (see below for instructions on determining P-sites). ORFquant can likewise either accept P-site data in the form of a table of offsets and a BAM file or use data in the form of processed P-site output from the RiboseQC package (which we recommend). Ribo-seq is tolerant to the presence of non-periodic noise in the input data. However, care should be taken to ensure that the alignments reflect the true position of ribosomal footprints—with adapters, etc. removed rather than soft clipped (see previous chapters).
8. (Optional—for RiboTaper)—Input RNA-seq data. RiboTaper additionally can make use of a BAM file containing RNA-seq data, which should be prepared in a similar manner to the Ribo-seq BAM file. In the absence of RNA-seq data, RiboTaper can be run on Ribo-seq data alone by simply passing the Ribo-seq BAM file again, in place of the RNA-seq file, when calling RiboTaper.

3 Methods

3.1 Preparing Annotation Files for RiboTaper

RiboTaper requires that annotation files be prepared in a dedicated directory before running the main program, which should be a subfolder of the directory into which RiboTaper's results will be written. This can be done by first opening the Unix command line (the "terminal" application on most distributions) and entering the following commands (or copying them into a shell script and executing this script):

```
mkdir my_RiboTaper_folder #create a directory for RiboTaper
cd my_RiboTaper_folder #step into this directory
mkdir RiboTaper_annot_folder #create the annotation subfolder

#create annotation files using the pre-existing files in the
#folder above
create_annotations_files.bash ../my_annotation.gtf \
../my_genome.fa true true  RiboTaper_annot_folder
```

The arguments to the above command represent, respectively, the call to the script itself (available on your path if you installed with conda and in your installation directory if you installed from source), the GTF file (see previous section), and the genome FASTA file (which should match the GTF file and be indexed using samtools faidx). The two logical values, which may be "true" or "false," are flags instructing RiboTaper to use CCDS annotation (present in GENCODE annotation for mouse and human—should be "false" for other annotations) and APPRIS annotation (which again may or may not be present in your annotation). The final argument represents the folder to which ORFquant's annotation files will be written.

The output of the command will be a set of some 35+ files containing various subsets of the annotation for use in later steps, as bed files. The command `wc -l RiboTaper_annot_folder/*` can be used to check that these files exist and contain data (note that the APPRIS and CCDS files may be empty depending on the parameters used above).

Estimated Time: 30 min

Estimated Disk Usage: 10–20Mb (in addition to that for input files)

Estimated Memory Requirements: 4GB

3.2 Preparing Annotation Files for ORFquant

The first step in running ORFquant is to open an interactive R session and prepare a dedicated directory for ORFquant's output, load our libraries, and create variables for the locations of our input files. We will assume the existence of (1) a pre-prepared Ribo-seq BAM file, (2) a valid annotation GTF, and (3) a genome sequence FASTA file. These files (or valid links to them) are stored in a directory called my_ORFquant_dir.

```
#load the RiboseQC (optional if another P-site determination
#algorithm is used) and ORFquant packages
library(RiboseQC)
library(ORFquant)
```

```
setwd("my_ORFquant_dir")#step into our directory
my_riboseq_bam <- 'my_sample.bam'
my_gtf_file <- 'my_annotation.gtf'
my_fasta_file <- 'my_genome.fa'
stopifnot(#verify these files exist in those locations
       file.exists(my_riboseq_bam)&
       file.exists(my_gtf_file)&
       file.exists(my_fasta_file)
)
```

To use a GTF file with ORFquant, it must first be converted into a serialized R object using the function prepare_annotation_file. This function requires a path describing the directory into which annotation files should be written, a GTF file (see above) with valid gene annotation, and a FASTA file to be used for sequence retrieval. Note that nonstandard circular chromosomes should also be passed as arguments, where present, if a FASTA file is used, to allow for sequence retrieval across the beginning of the coordinate system.

Irregularities in gene annotations, particularly for non-model organisms, are by far the most frequent source of errors in the ORFquant pipeline. The GTF file should follow the official GENCODE GTF standard format with metadata columns containing unique ids for genes, transcripts, and CDS, as well as gene and transcript biotypes.

The output of this step is an R object containing the necessary annotation for ORFquant. An example command is shown below:

```
orfquant_anno_file <-prepare_annotation_files(
 annotation_directory ="annotation_directory/" ,
 genome_seq = my_fasta_file,
 gtf_file =  my_gtf_file,
 scientific_name = "Human.test",
 annotation_name = "my_annotation",
 export_bed_tables_TxDb = F,
 create_TxDb = T
)
stopifnot(file.exists(orfquant_anno_file))
```

The resulting "orfquant_anno" object contains annotation data, as well as the file path for the genome FASTA file. Running the command help(prepare_annotation_file) will show documentation explaining the different elements of the object, which include GRanges objects describing exons and their locations in genomic and transcriptomic space and tables linking transcripts and gene loci.

Estimated Time: 30 min

Estimated Disk Usage: 10–20M (in addition to that for input files)

Estimated Memory Requirements: 4GB

3.3 Preparing P-Site Signal Data

The second step in running both RiboTaper and ORFquant is to load the input P-site data. Typically, P-site tracks are created by shifting each RPF start position by a read length-dependent amount, which corresponds to the most frequent location of the P-site in reads of that length (so that, e.g., the peak of 5′ RPF coordinates ~10bp before the start codon becomes a peak located directly over the start codon). P-site data will typically be more periodic than a simple track of 5′ RPF coordinates, but will also have less coverage, because some read lengths will be excluded due to the P-site will not be located at a single consistent distance from the 5′ end.

A large number of strategies for P-site inference have been described (see [6] for review), and the appropriate one will depend on factors including the species of origin, the desired downstream analysis, and the nature of the Ribo-Seq protocol, in particular the presence of sequence bias due to the digestion enzyme used, and whether or not cycloheximide has been used in the protocol. The quality of P-site inference influences the periodicity of the final ribosomal footprint track, and care should be taken with this step, since faulty P-site inference will reduce the sensitivity of both RiboTaper and ORFquant.

We recommend the package RiboseQC, which contains functions to assess the quality of P-site inference and which should reliably allow P-site inference for all libraries in which a distinct peak over the start codon is detectable—something easily verified in the metagene plots provided by RiboseQC. RiboseQC's output can be directly passed into ORFquant as a serialized R object. RiboseQC also allows for compartments, e.g., nuclear and mitochondrial genes, to be treated separately. This is a necessary step where genes of interest are translated by a different ribosome species—as is the case for both chloroplast and mitochondrial genes—since these organelle-specific ribosomes yield distinct footprints and necessitate distinct P-site offsets. An example of P-site data being loaded from the RiboseQC package is shown below:

```
#Run the RiboseQC pipeline
ribseqc_file <- RiboseQC_analysis(
  annotation_file = orfquant_anno_file,
  bam_files = my_riboseq_bam,
  fast_mode = T,
  create_report = T
)
#check that the file of p-site offsets has been created
my_offset_file <- paste0(
  my_riboseq_bam,
  "_P_sites_calcs"
)
#verify our psite offset file exists
file.exists(my_offset_file)
```

Here, we pass the annotation file created above using prepare_annotation_file, along with the BAM file, to RiboseQC. We use the fast_mode to tell RiboseQC to use only the top 500 most expressed genes (for final results, this flag should be set to false). The create_report parameter controls whether RiboseQC should create a html report, in addition to saved P-site data. RiboseQC has optional parameters which can be used to specify the name of the report file, P-site offset file, etc. Here we allow it to use the default for these values—which will output files prefixed with the name of the BAM file passed as input. RiboseQC's output is a html report showing various quality control metrics, various bedgraph files containing P-site tracks (see the package documentation), and a saved file "my_sample.bam_P_sites_calcs" (where "my_sample" is the sample_names argument) containing an optimal set of P-site offsets.

Good offsets should (a) place most P-sites within the coding sequence, (b) align the peak of P-sites at the start and/or end of the transcript for different read lengths, and (c) not differ dramatically (i.e., more than ~6nt) between adjacent read lengths. Note that some other package should be used to quality check the inferred locations of P-sites if RiboseQC is not used.

Estimated Time: 1 h with RiboseQC

Estimated Disk Usage: ~1Gb

Estimated Memory Requirements: 4GB

3.4 Running RiboTaper

With prepared annotation data, and the set of offsets required to map the RPF data to a single P-site track, and an RNA-seq BAM file, RiboTaper can be run on the command line with a command such as the following:

```
RiboTaper.sh ../my_sample.bam ../my_sample_rna.bam
RiboTaper_annot_folder/ 26,27,28,29,25,30,31 \
11,11,12,11,10,12,12   2
```

Note that here we supply the names of our Ribo-seq and RNA-seq BAMs as arguments, followed by a comma-separated list of read lengths, and offsets to be applied to them. See Fig. 2 for a schematic explanation of the internal steps carried out by RiboTaper.

Estimated Time: 3 h

Estimated Disk Usage: ~1Gb

Estimated Memory Requirements: 4GB

3.5 Running ORFquant

With objects containing P-site data and annotation, ORFquant can now be run. By default, ORFquant will run on all genes, which will take some time for large genomes. For instance, a machine with

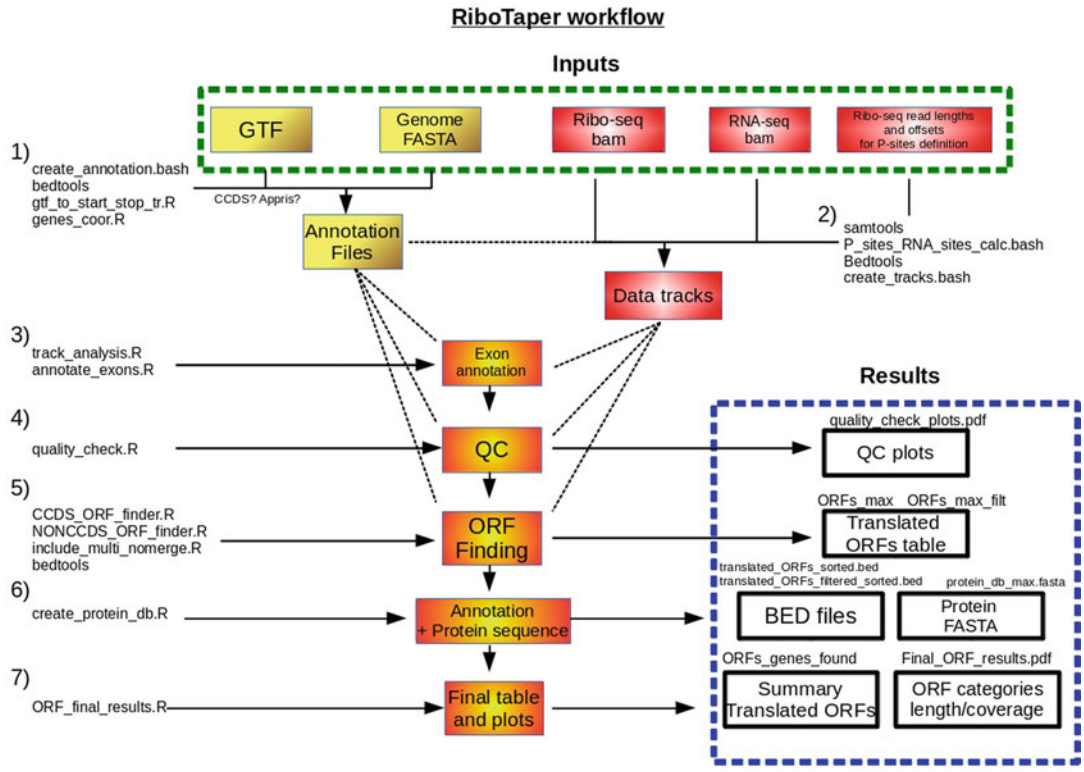

Fig. 2 The RiboTaper workflow. The RiboTaper.sh master script calls a series of subscripts (which can be accessed individually; run, e.g., create_protein_db.R for instructions on individual usage) to create P-site tracks, QC plots, tables of annotated ORFs, a protein FASTA file, and a final table of ORFs

8Gb of RAM and two cores might take 12 h to analyze the entire human transcriptome, using a Ribo-seq dataset with 3 million reads. We therefore recommend choosing some target genes (which should have a high expression—"GAPDH" is a suitable default choice for human data) to test run ORFquant before running on the entire genome, by setting the "gene_name" parameter. This feature can also be used to run ORFquant on subsets of the genome when using a HPC cluster. Care should be taken in this case to ensure that each node is given sufficient memory to load the input P-site data, which has a large memory footprint.

An example command for running ORFquant on the entire genome, using a text file of offsets (see Subheading 2), in this case generated above by RiboseQC, is shown below:

```
#prepare psite data from bam file and offset file
my_orfquant_psites_file <-prepare_for_ORFquant(
  orfquant_anno_file,my_riboseq_bam,
  path_to_rl_cutoff_file = my_offset_file
)

#run orfquant
ORFquant_results <- run_ORFquant(
 for_ORFquant_file = my_orfquant_psites_file,
 annotation_file = orfquant_anno_file,
 n_cores = 1)


orfquant_res_file <- paste0(
  my_riboseq_bam,
  "_for_ORFquant_final_ORFquant_results"
)
stopifnot(file.exists(orfquant_res_file))
```

Estimated Time: ~12 h (or 3 on a machine with 8 available cores)

Estimated Disk Usage: ~1.5 Gb

Estimated Memory Requirements: 4GB (or more for effective multicore processing)

3.6 Plotting, Interpreting, and Filtering Results

The master script RiboTaper.sh will call a series of scripts that create QC plots showing the ORFs identified and the biotype of the transcript and gene in which they are found.

The output of ORFquant is a number of output files along with an R object, now available in your main workspace called ORFquant_results. This object is a list containing various information about the feature selection and quantification by ORFquant. For a detailed account of each, see the help for the function run ORFquant.

The most important output of ORFquant is the set of identified ORFs listed under "ORFs_tx"—a GRanges object containing the coordinates of each identified ORF, in transcript space. The metadata of this object is accessible with the accessor function mcols. *P*-values for the periodicity tested are listed here, as well as a separate *P*-value for uniquely mapping reads only and columns linking each ORF to a particular transcript and "gene_id" value in the annotation. ORF_category_Tx classifies ORFs into categories such as ORFs present in the canonical annotation, novel ORFs not present in the annotation, and upstream or downstream ORFs.

Before interpreting ORFquant (or RiboTaper) results, assessing the coverage of P-sites over CDS (which may be less than that of RPFs, where not all read lengths have a reliably positioned P-site)

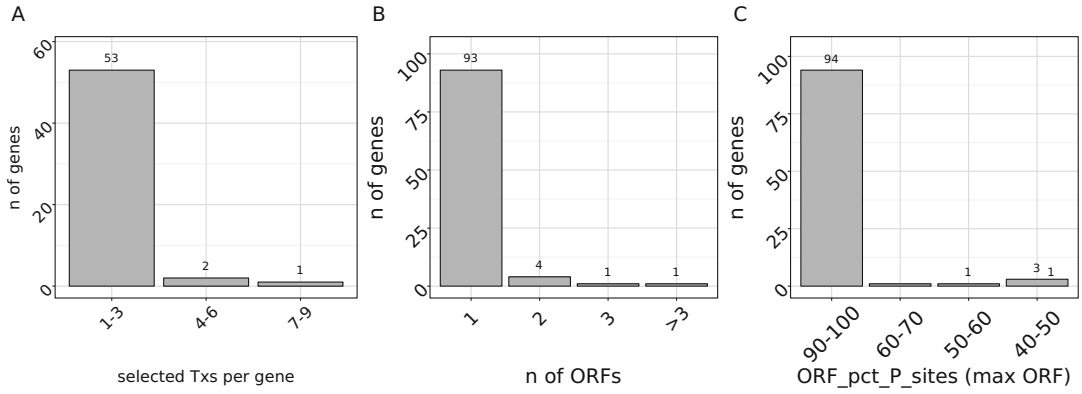

Fig. 3 Diagnostic plots generated by ORFquant. The function "plot_ORFquant_results" will create a folder containing 12 diagnostic plots which include (**a**) a histogram showing the number of transcripts selected for quantification by ORFquant; (**b**) a histogram showing the number of ORFs with evidence of periodic Ribo-seq, per gene; and (**c**) a histogram showing the relative proportion of P-sites (ORF_pct_P_sites) attributed to the most highly expressed ORF in each gene. In this small sample of *Arabidopsis* genes, all but 3 genes had between 1 and 3 transcripts with quantifiable translation, all but 5 of these had a single translated ORF, and in all but 5 of these cases, a single transcript could account for 90–100% of the translation

is a must. This information can be accessed via the `ORFquant_results` object:

```
hist(n=50,log10(1+ORFquant_results$ORFs_Tx$P_sites))
```

As a rule of thumb, most protein coding genes are well quantified when median log10 counts are at ~2 (i.e., 100 counts per gene). Higher coverage may be necessary to detect low expression ORFs, such as uORFs.

The function plot_ORFquant_results shows summary plots for the output (Fig. 3), including plots showing numbers of transcripts and ORFs selected per gene and total number, ORF length, periodicity, and proportion of gene's P-site attributed for ORFs of different biotypes (*see* **Note 1**). We can create summary plots (showing number of detected ORFs for different gene biotypes, statistics on ORF quantifications, etc.) in PDF format using:

```
plot_ORFquant_results(
 for_ORFquant_file = my_orfquant_psites_file,
 ORFquant_output_file =orfquant_res_file,
 annotation_file = orfquant_anno_file
)
plotfolder <- paste0 (orfquant_res_file,"_plots/")
stopifnot (file.exists(plotfolder))
```

These summary plots can also be embedded in an html report:

```
orfquantfile=paste0(
```

```
  "my_sample",
 "bam_for_ORFquant_final_ORFquant_results_plots/,
 "my_sample.bam_for_ORFquant_ORFquant_plots_RData"
)
create_ORFquant_html_report(
  input_files = orfquantfile,
  input_sample_names = "my_sample",
  output_file= "my_sample_ORFquant_report.html"
)
```

In addition, we provide functions for plotting specific loci of interest:

```
plot_orfquant_locus(locus= "AT3G15950",
  orfquant = orfquant_results,
  bam_files = ribo_bams,
  col = "green"
)
```

We recommend that loci of specific interest (e.g., for those intended for validation experiments) should be visualized individually (see Fig. 4).

3.7 Caveats to Interpretation

ORFquant's transcript quantification relies on transcripts having enough isoform-specific reads to differentiate their translation levels. As heuristic, if less than ten reads are uniquely attributable to a transcript, its expression relative to other isoforms is difficult to

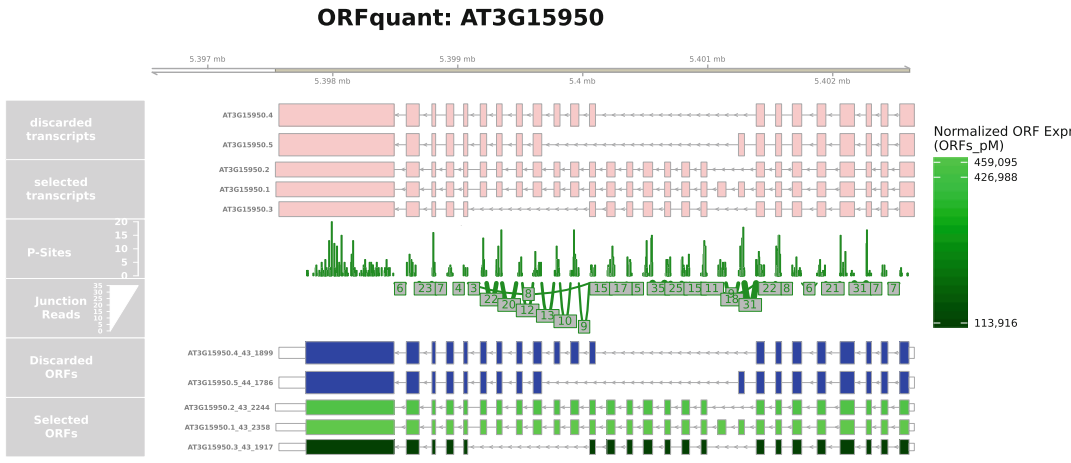

Fig. 4 Plotting ORFquant results at specific loci. The function "plot_orfquant_locus" plots Ribo-seq signal at a specific locus, using the Gviz package to visualize the local isoform structure; the Ribo-seq reads used by ORFquant (top two panels), including those that span junctions; the ORFs included (green) and discarded (blue) in the final results of ORFquant; and the scaled expression values attributed to each ORF (i.e., the ORFs_pM values) plotted using a color gradient. These values, analogous to TPMs, are normalized for library size and ORF length

quantify precisely. ORFquant's sensitivity will be influenced by the coverage achieved in the experiment, and the length of a given ORF, as well as experimental factors like the use of cycloheximide (*see* **Note 1**) and MNase (*see* **Note 2**). By default, ORFquant uses relatively permissive filtering to allow for downstream filtering and inspection. We recommend that ORFs used in further analyses, regardless of *P*-value, should contain at least ten uniquely mapping reads.

The multitaper tests used by ORFquant and RiboTaper fall prey to the uncertainty principle on very short sequences—the frequency of a signal cannot be accurately estimated over a very short time (or distance) window. Thus, ORFs less than ~50 bp may not be reliably detectable targets for ORFquant or RiboTaper.

An artifact seen in some Ribo-seq libraries is a depletion of signal toward the 5′ ends of genes (possibly still with a large peak at the start codon). The presence of this artifact can be easily ascertained by metagene plots of Ribo-seq signal, which should be performed for each library. In such libraries, spurious variant ORFs with 5′ truncations may appear in ORFquant's results.

ORFquant's results include a metadata column, "ORFs_pM"; ORFs_pM is a TPM-like value expressing the normalized translational output for the ORF in the sample. ORFs_pM is based on P_site counts, so normalization methods (next chapter) applied to most count data will be valid when applied to both expression values output by ORFquant.

4 Notes

1. The use of cycloheximide will tend to increase the periodicity of a Ribo-seq sample, by locking ribosomes in place and eliminating signal from ribosomes in intermediate stages of the translation cycle. This may however result in a loss of the ability to interpret the precise location of ribosomes, since a certain amount of "run-on" may occur before a ribosome locks, resulting in a loss of the correlation between the underlying codon and ribosomal occupancy. This should not however strongly influence the multitaper test's sensitivity [7].
2. A factor affecting the sensitivity of ORFquant will be the sequence bias of any enzyme used to create the ribosomal footprints. Enzymes such as MNase with strong sequence preferences will tend to "blur" the Ribo-seq signal and reduce periodicity—here, sequence-specific adjustments to P-site offsets may help to "deblur" the signal [8]. P-sites adjusted in this manner may be fed directly into ORFquant and RiboTaper. In RiboTaper's case, a bed file called should be created with the adjusted P-sites and placed in the working directory. For ORFquant, the psite_file argument should be used to pass a single

column text file containing the offset for each read, in the order in which they appear in the BAM file. Note that reads with non-numeric values such as NA in the offset column will be excluded.

Acknowledgments

This work supported by Deutsche Forschungsgemeinschaft (grant TR175) and the Federal Ministry for Education and Research (BMBF, grant de.NBI/RBC). The authors acknowledge helpful feedback and support from Antje Hirsekorn and Dilmurat Yusuf.

References

1. Calviello L, Mukherjee N, Wyler E, Zauber H, Hirsekorn A, Selbach M, Landthaler M, Obermayer B, Ohler U (2016) Detecting actively translated open reading frames in ribosome profiling data. Nat Methods 13 (2):165–170. https://doi.org/10.1038/nmeth.3688
2. Bazzini AA, Johnstone TG, Christiano R, Mackowiak SD, Obermayer B, Fleming ES, Vejnar CE, Lee MT, Rajewsky N, Walther TC, Giraldez AJ (2014) Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. EMBO J 33(9):981–993. https://doi.org/10.1002/embj.201488411
3. Gonzalez-Porta M, Frankish A, Rung J, Harrow J, Brazma A (2013) Transcriptome analysis of human tissues and cell lines reveals one dominant transcript per gene. Genome Biol 14 (7):R70. https://doi.org/10.1186/gb-2013-14-7-r70
4. Calviello L, Hirsekorn A, Ohler U (2019) Quantification of translation uncovers the functions of the alternative transcriptome. bioRxiv:608794. https://doi.org/10.1101/608794
5. Calviello L, Sydow D, Harnett D, Ohler U (2019) Ribo-seQC: comprehensive analysis of cytoplasmic and organellar ribosome profiling data. bioRxiv:601468. https://doi.org/10.1101/601468
6. Ahmed N, Sormanni P, Ciryam P, Vendruscolo M, Dobson CM, O'Brien EP (2019) Identifying A- and P-site locations on ribosome-protected mRNA fragments using integer programming. Sci Rep 9(1):6256. https://doi.org/10.1038/s41598-019-42348-x
7. Weinberg DE, Shah P, Eichhorn SW, Hussmann JA, Plotkin JB, Bartel DP (2016) Improved ribosome-footprint and mRNA measurements provide insights into dynamics and regulation of yeast translation. Cell Rep 14 (7):1787–1799. https://doi.org/10.1016/j.celrep.2016.01.043
8. Fang H, Huang YF, Radhakrishnan A, Siepel A, Lyon GJ, Schatz MC (2018) Scikit-ribo enables accurate estimation and robust modeling of translation dynamics at codon resolution. Cell Syst 6(2):180–191. e184. https://doi.org/10.1016/j.cels.2017.12.007

INDEX

Vyacheslav M. Labunskyy (ed.), *Ribosome Profiling: Methods and Protocols*, Methods in Molecular Biology, vol. 2252, https://doi.org/10.1007/978-1-0716-1150-0,

P

R

S

T

U