



Published in final edited form as:

*Nat Protoc.* ; 7(8): 1534–1550. doi:10.1038/nprot.2012.086.

## The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments

Nicholas T. Ingolia<sup>1,2,\*</sup>, Gloria A. Brar<sup>1</sup>, Silvia Rouskin<sup>1</sup>, Anna M. McGeachy<sup>2,3</sup>, and Jonathan S. Weissman<sup>1,†</sup>

<sup>1</sup>Howard Hughes Medical Institute, Department of Cellular and Molecular Pharmacology, University of California, San Francisco, and California Institute for Quantitative Biosciences, San Francisco, CA 94158, USA

<sup>2</sup>Department of Embryology, Carnegie Institution for Science, Baltimore, MD, 21218

<sup>3</sup>Department of Biology, The Johns Hopkins University, Baltimore, Maryland 21218

### Abstract

Recent studies highlight the importance of translational control in determining protein abundance, underscoring the value of measuring gene expression at the level of translation. We present a protocol for genome-wide, quantitative analysis of in vivo translation by deep sequencing. This ribosome profiling approach maps the exact positions of ribosomes on transcripts by nuclease footprinting. The nuclease-protected mRNA fragments are converted into a DNA library suitable for deep sequencing using a strategy that minimizes bias. The abundance of different footprint fragments in deep sequencing data reports on the amount of translation of a gene. Additionally, footprints reveal the exact regions of the transcriptome that are translated. To better define translated reading frames, we describe an adaptation that reveals the sites of translation initiation by pre-treating cells with harringtonine to immobilize initiating ribosomes. The protocol we describe requires 5–7 days to generate a completed ribosome profiling sequencing library. Sequencing and data analysis requires a further 4 – 5 days.

### Keywords

translation; ribosome profiling; ribosome footprinting; deep sequencing

### INTRODUCTION

Gene expression profiling is a central tool for understanding cellular physiology and regulation. Historically, studies of gene expression have typically measured mRNA abundances rather than rates of protein synthesis, in large part because such data are much easier to obtain. The focus on overall mRNA levels increased with the emergence of

\*Correspondence: [ingolia@ciwemb.edu](mailto:ingolia@ciwemb.edu); tel: 410 246 3025; fax: 410 243 6311. †Correspondence: [weissman@cmp.ucsf.edu](mailto:weissman@cmp.ucsf.edu); tel: 415 502 7642; fax: 415 514 2073.

#### AUTHOR CONTRIBUTIONS

N.T.I. and J.S.W. designed the study. G.A.B. and J.S.W. developed the rRNA depletion protocol. S.R. and J.S.W. adapted the protocol to use preadenylylated linker ligation. N.T.I., S.R., G.A.B., and A.M.M. performed experiments. N.T.I. and A.M.M. analyzed the data. N.T.I. and J.S.W. wrote the manuscript.

#### COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests (see the HTML version of this article for details).

microarrays<sup>1</sup> and more recently RNA-Seq<sup>2,3</sup> as comprehensive and quantitative expression profiling techniques. These measurement techniques have revolutionized our ability to monitor the internal state of cells, but they have naturally led to a focus on transcriptional regulatory networks. However, mRNA and protein levels are imperfectly correlated in yeast and in mammalian cells<sup>4-7</sup>, and translational control can play a critical role in modulating gene expression<sup>7,8</sup>.

### Overview of Ribosome Profiling

We recently developed an approach, termed ribosome profiling, based on deep sequencing of ribosome-protected mRNA fragments, that now makes it possible to monitor translation directly<sup>9</sup>. The protein being synthesized by a ribosome is, of course, determined by the mRNA sequence it is decoding. A translating ribosome encloses a ~30 nt portion of this mRNA template and protects it from nuclease digestion. These ribosome-protected mRNA fragments have previously been used to map the positions of ribosomes in homogeneous *in vitro* translation reactions<sup>10,11</sup>. Dramatic advances in sequencing technology<sup>12</sup> now make it possible to characterize the complex pool of fragments produced by nuclease footprinting of ribosomes from living cells. Each ribosome produces a footprint fragment whose sequence indicates which mRNA it was translating, as well as its precise position on the transcript. Deep sequencing of ribosome footprints thus provides information about ribosome positions, which is inaccessible to existing polysome profiling approaches for measuring translation, as well as measuring expression quantitatively.

Here, we present a detailed experimental protocol for ribosome profiling in cultured mammalian cells (Fig. 1). This technique has been applied to study developmental changes in mouse ES cells<sup>13</sup> and to monitor the effects of drug therapies in human cancer cell models<sup>14</sup>, and should be applicable to many other biological questions. It begins with cell lysis and harvesting under conditions that should maintain *in vivo* ribosome positions on mRNAs. These lysates are treated by nuclease digestion to perform ribosome footprinting, and ribosomes are recovered by ultracentrifugation. Ribosome footprints are purified and ligated to a single-stranded linker that serves as a priming site for reverse transcription. The first-strand reverse transcription products are circularized, providing a second priming site flanking the captured footprint sequence, which is used for PCR amplification of a deep sequencing library.

### Applications of Ribosome Profiling

Ribosome footprint sequences indicate which portions of the genome are actually being translated into protein. These translated sequences include conventional protein-coding genes, but also reading frames that encode short peptides. A few short ORFs have been identified genetically<sup>15</sup>, and ribosome profiling data has revealed many more<sup>13,16,17</sup>. Thus, it is likely that the number of small peptides is much larger than currently known. In order to be translated, a sequence must first be transcribed. Recent studies have revealed great diversity in the mammalian transcriptome<sup>18</sup>, though many of these transcripts lack long open reading frames. Short ORFs on traditional and noncanonical messages can be difficult to identify reliably by computational approaches<sup>19</sup>, but ribosome profiling has proven to be a highly useful tool for exploring the peptide coding potential of these RNAs<sup>13</sup>.

In addition to discovering these novel ORFs, ribosome profiling data leads to revisions in the annotation of known genes. In many cases, footprinting data indicates the translation of extended or truncated forms of proteins. These alternate protein isoforms can have functions that are distinct from or antagonistic to the annotated form. They can arise from the translation of different mRNA isoforms or from the use of alternative initiation sites on the same transcript. When alternate isoforms are co-expressed, translation from upstream

initiation sites can obscure the presence of downstream initiation. These internal initiation sites are revealed by ribosome profiling after treatment with harringtonine, a drug that immobilizes ribosomes immediately after translation initiation and results in footprint accumulation at all initiation sites<sup>20,21</sup>. Thus, the presence or absence of ribosome footprints at downstream AUG codons in harringtonine-treated samples marks sites of potential internal initiation leading to shorter protein isoforms. More generally, initiation site profiling with harringtonine can be combined with ribosome profiling to detect elongating ribosomes over the entire reading frame to produce an experimentally-based annotation of the translated products from a genome<sup>13</sup>.

Still, the broadest application of ribosome profiling may be measurements of gene expression at the level of actual protein synthesis. Each ribosome footprint corresponds to a translating ribosome, so the number of footprints produced from a transcript should correspond to the number of ribosomes engaged in synthesizing the encoded protein. This is proportional to the amount of the protein being produced and to the time required to produce it. We have shown that the speed of protein synthesis is broadly consistent across different groups of genes<sup>13</sup>. Thus under a given condition, the translation time of an ORF is simply proportional to its length. One can therefore determine the rate at which a protein is being produced by measuring the density of ribosome footprints on its transcript. Ribosome footprint density thus can be used in place of mRNA abundance measurements to quantify gene expression. It can also be combined with mRNA abundance measurements to identify translational regulation as changes in protein expression that cannot be explained by transcript levels<sup>9,17</sup>. RNA-Seq measurements of transcript abundance measurements can be made by analysis of randomly fragmented complete mRNA in parallel with ribosome footprints<sup>9,16</sup> or by other standard approaches<sup>22</sup>.

Finally, ribosome profiling provides an approach for studying the mechanics of translation and of co-translational processes in vivo. Just as the overall number of ribosome footprints on a gene indicates how many ribosomes are typically translating it, the number of footprints centered on a codon should reflect how often a ribosome is found at that particular spot. If ribosomes stall at a specific point when translating a gene, then ribosomes will spend more time there than elsewhere, producing a corresponding excess of footprints. We have used this excess of ribosome footprints to detect peptide-mediated translational stalling in mammalian cells<sup>13</sup> and RNA-mediated stalling in bacteria<sup>23</sup>. Ribosome footprint density has also been applied to determine codon-specific elongation rates in bacteria<sup>23</sup> and yeast<sup>24</sup>, as well as *C. elegans* and human cultured cells<sup>25</sup>. It has also been applied to monitor co-translational processes in protein biogenesis, including chaperone association and protein secretion<sup>26,27</sup>.

### Convergence of Expression Profiling Techniques

Ribosome profiling bridges the gap between global measurements of steady-state mRNA and protein levels. As such, it will be particularly valuable to compare ribosome profiling and mass spectrometry measurements of protein expression levels. At present, sequencing technologies provide deeper measurement than mass spectrometry measurements in most circumstances. However, steady-state measurements by mass spectrometry are sensitive to protein degradation as well as synthesis. In fact, high-quality ribosome profiling and proteomic measurements may offer a new approach to determine the turnover rate of native proteins in unperturbed cells. Similar interpolation between RNA-Seq and mass spectrometry measurements recently quantified the large contribution of translation to steady-state protein levels<sup>6</sup>.

Until now, the translational status of mRNAs typically has been assessed by separating intact ribosome-mRNA complexes based on the total number of ribosomes bound to a

transcript. A genomic adaptation of this assay, called polysome profiling, measures the mRNA constituents of different ribosome number fractions using microarrays<sup>28,29</sup>. Ribosome profiling has technical advantages over polysome profiling for making routine expression measurements, but polysome profiling can complement ribosome footprinting experiments particularly for performing mechanistic studies of translational control. Ribosome profiling provides more precise expression measurement, because it avoids the difficulty in resolving the exact number of ribosomes bound to highly ribosome-loaded transcripts. Failure to separate these transcripts can obscure changes in the exact number of ribosomes bound to them and thereby compress the dynamic range of polysome profiling experiments. Ribosome profiling also avoids certain technical hurdles that arise in polysome profiling. While many skilled investigators reliably obtain high-quality, intact polysomes, RNA degradation remains a challenge. Ribosome profiling requires only the nuclease footprint from single ribosomes, so it is less sensitive to compromised RNA integrity. Finally, ribosome profiling can distinguish between ribosomes translating protein-coding genes and those translating regulatory upstream open reading frames.

Polysome profiling monitors the translational status of entire transcripts, which provides data that cannot be determined from footprint sequencing measurements that focus on the activities of individual ribosomes. Thus, polysome profiling can distinguish between a uniform decrease in the number of ribosomes on all copies of a transcript and a complete repression of a sub-population of mRNAs, a phenomenon that was revealed by polysome profiling of mouse ES cells<sup>29</sup>. By contrast, ribosome profiling would simply detect a quantitative decrease in the ensemble-averaged rate of protein synthesis in either case. Similarly, polysome profiling may have a greater ability to measure differences in the translation of alternate transcript isoforms, particularly when they differ in their 5' or 3' untranslated regions. These measurements could complement ribosome profiling data to provide insight into the molecular mechanism of translational regulation.

## Experimental Design

**Cell lysis**—Ribosome profiling begins with the preparation of cell lysates where ribosome-mRNA complexes accurately reflect *in vivo* translation. The best approach for lysate preparation will vary based on the sample being analyzed. Traditionally, polysomes have been “stabilized” by treating cells with translation elongation inhibitors before cell lysis. We found that in mammalian cells, brief treatment with such drugs causes an accumulation of ribosomes in the first five to ten codons of all genes. This may well reflect an artifactual accumulation of ribosomes that initiate during drug treatment and stall translation shortly thereafter. We therefore favor the *in situ* detergent lysis of adherent, cultured cells because it seems to produce the least opportunity for perturbation between normal growth and ribosome extraction. However, this approach is not suitable for all samples. In *Saccharomyces cerevisiae*, we found that elongation inhibitors suppressed changes in translation that occurred during cell harvesting<sup>9</sup>, and similar polysome stabilization may be necessary in other situations as well. Drug pre-treatment may provide other valuable information. For instance, brief pre-treatment of cells with harringtonine (Box 1) enriches ribosomes specifically on initiation sites, enhancing the detection and annotation of translated sequences.

### Box 1

#### Harringtonine Treatment

The following protocol describes the procedure for harringtonine pre-treatment, which immobilizes initiating ribosomes in order to profile translation start sites.

1. Add harringtonine to cell culture media to a final concentration of 2  $\mu\text{g/ml}$ . Mix quickly and return to incubator.
2. Incubate cells for 120 s with harringtonine. [CRITICAL STEP – Be prepared to proceed quickly to cycloheximide treatment and cell lysis after harringtonine addition.]
3. Add cycloheximide to cell culture dish to a final concentration of 100  $\mu\text{g/ml}$ .
4. Mix well and proceed immediately to cell lysis (Step 1, main Procedure).

We also found that cryogenic pulverization of frozen yeast produced effective lysis and homogenization under conditions that blocked biological responses. This technique is also applicable in mammalian cells and tissues that require physical disruption for ribosome extraction. Indeed, flash freezing of tissues followed by cryogenic pulverization and thawing in the presence of translation inhibitors provides a particularly robust and simple approach for the analysis of animal-derived samples <sup>30</sup>.

**Nuclease footprinting**—Nuclease footprinting converts ribosome positions into RNA sequence tags that can be analyzed by deep sequencing. We have found that similar digestion conditions can be used to footprint ribosomes in yeast, HeLa, and mouse ES cell lysates, suggesting that extensive optimization is not needed for profiling in various eukaryotic systems. However, we have identified an effect of lysis buffer conditions on the size and reading frame precision of ribosome footprints. Ribosome footprinting in mammalian cells resulted in longer ribosome footprints whose termini showed less specific positioning relative to the reading frame being decoded <sup>13</sup>. We noted that the polysome buffer used in our initial analysis of mammalian cells had higher ionic strength and higher magnesium concentration than the yeast polysome buffer. Subsequently, we observed that reducing salt and magnesium improved the resolution of our ribosome footprints without substantially altering overall measurements of gene expression (Fig. 2 and Supplementary Figs. 1 and 2). High magnesium concentration inhibits spontaneous conformational changes in bacterial ribosomes <sup>31</sup>. If a similar effect occurs in eukaryotic ribosomes, then the increased interconversion in low-magnesium conditions could permit more complete and uniform nuclease digestion. Alternately, buffer conditions could affect the interactions between the mRNA and the ribosome. In either case, while these differences do not affect gene expression measurements (Fig. 2 and Supplementary Fig. 1), we favor the higher-precision footprinting seen in the buffer conditions presented here.

**Ribosome recovery**—Following nuclease digestion, we separate intact ribosome-footprint complexes from cell lysates prior to RNA extraction. We originally performed sucrose density gradient purification of 80S ribosome particles. However, sucrose density gradient fractionation is challenging, and in fact represents a substantial barrier to analyzing translation through traditional polysome approaches. More recently, we purified ribosomes by sedimenting them through a 1M sucrose cushion, which provides a more accessible density-based separation. We did not observe increased contamination with untranslated but protein-bound RNA sequences, such as 3'UTRs, in samples purified by sucrose cushion, though it is in principle less specific than sucrose gradient purification and it may be more important to verify that RNA fragments show the characteristic size and reading frame distribution of true ribosome footprints. Other approaches are possible as well—ribosomes can be purified by gel filtration <sup>32</sup>, and in certain systems genetic manipulation can be used to add epitope tags to ribosomes, enabling affinity purification <sup>33,34</sup>.

**Linker ligation**—Deep sequencing analysis typically requires libraries containing specific linker sequences; in the case of the Illumina sequencer employed here, the library is double-stranded DNA with defined sequences flanking the target fragment. In our previous work, we identified and worked to minimize significant biases in the conversion of RNA footprints into a sequencing library. These biases are present in all RNA-Seq libraries, but cause particular difficulties in analyzing ribosome footprinting data. While our approach achieved notably good uniformity, it involved the addition of a poly-(A) tail to each sequence. This degenerate sequence complicated bioinformatic analyses. We have subsequently shown that an optimized RNA ligation of a preadenylated linker<sup>35</sup> can achieve comparable results, and both the genetically modified RNA ligase and the chemically modified linker required for this approach are now commercially available. We have also altered the sequences of reverse transcription and PCR primers used in the protocol to allow sequencing with the standard Illumina primers. This includes the option of adding a 6 nt index that can be read in the same manner as the indices added to standard Illumina libraries.

**rRNA Depletion**—Ribosomal RNA contamination substantially decreases the amount of informative sequence data obtained in a ribosome profiling experiment. This is unsurprising, as there are several kilobases of rRNA in each ribosome-footprint complex, but only ~28 bases of footprint mRNA. We observed that a few specific rRNA fragments represented a large fraction of the overall contamination present in the 26 – 34 nt size window that we purified, presumably because nuclease digestion of intact ribosomes results in reproducible cleavage at a limited number of positions. We therefore depleted first-strand cDNAs derived from these high-abundance contaminants by hybridization to biotinylated sense strand oligonucleotides followed by removal of the duplexes through streptavidin affinity.

**Analysis and interpretation**—Ribosome footprint sequencing data can be preprocessed and then aligned to the genome using tools available for RNA-Seq analysis (Fig. 3). Most of the bioinformatics challenges, such as the alignment of reads across splice junctions and the possibility of multiple distinct genomic alignments, are similar in these data. Considerable specific rRNA contamination can remain even after depletion by subtractive hybridization. Thus, we also implement a bioinformatics filter to remove these sequences first. Examination of the positive alignments from this filter should point to specific contaminating rRNA sequences that can be targeted with additional biotinylated subtraction oligos. We also identify some contaminating sequences derived from other abundant non-coding RNAs (ncRNAs), such as tRNAs and snRNAs. These contaminants typically derive from a single specific position, whereas ribosome footprints will cover many positions along a reading frame, and ncRNA fragments will show an atypical length distribution. The extent of rRNA and ncRNA contamination can vary, particularly when global changes in protein synthesis alter the fraction of active ribosomes, and thus the number of ribosome-protected footprints relative to other RNAs.

Many applications of ribosome profiling, including expression measurements, depend on comparing the numbers of aligned sequencing reads between genes, samples, or specific codons. Quantitative analysis of ribosome profiling data, like RNA-Seq data<sup>2</sup>, is powerful, but this analysis must account for limitations that arise in this sort of data. Two major concerns that have been studied in RNA-Seq data are systematic sequence-dependent biases and stochastic sampling errors<sup>22,36-41</sup>. RNAs will be captured during library generation with differing efficiencies, perhaps due to sequence or structural preferences of the enzymes used in library generation<sup>42-44</sup>. While we strove to minimize these biases, they are present in all sequencing samples<sup>22</sup> and it is important to avoid confusing library generation biases with differences in the underlying abundance of different footprints. These sequence biases are minimized in expression measurements due to averaging across the entire sequence of the mRNA; comparison of the same gene across different samples, one of the most frequent



uses of profiling data, provides further protection from these effects. Stochastic error also arises, and is most serious when comparing small absolute numbers of reads. Several statistical approaches have been developed to estimate and model this error in RNA-Seq data, which can exceed the expectation derived from Poisson statistics<sup>37,38,41,45</sup>. Many of these techniques and tools should be directly applicable to ribosome profiling expression measurements.

## MATERIALS

### Reagents

- HEK293 cells (ATCC, cat. no. CRL-1573) or other cultured mammalian cells including HeLa cells, mouse neutrophils<sup>46</sup>, and PC3 cells<sup>14</sup>.
- 100 mg/ml cycloheximide (Sigma/Aldrich, cat. no. C4859-1ML) [CAUTION – Cycloheximide is very toxic and harmful to the environment. Handle solutions containing cycloheximide with care and decontaminate and dispose of waste in accordance with institutional regulations]
- Harringtonine (LKT Laboratories, cat. no. H0169) for optional harringtonine treatment (Box 1) [CAUTION – Harringtonine is very toxic. Handle solutions containing harringtonine with care and decontaminate and dispose of waste in accordance with institutional regulations]
- Dimethyl sulfoxide, cell culture grade (Sigma/Aldrich, cat. no. D2650) for optional harringtonine treatment.
- PBS, pH 7.2 (Invitrogen, cat. no. 20012-027)
- RNase-free water (Invitrogen, cat. no. AM9930)
- 1 M Tris\*Cl pH 8, RNase-free (Invitrogen, cat. no. AM9855G)
- 1 M Tris\*Cl pH 7, RNase-free (Invitrogen, cat. no. AM9850G)
- 5 M NaCl, RNase-free (Invitrogen, cat. no. AM9760G)
- 1 M MgCl<sub>2</sub>, RNase-free (Invitrogen, cat. no. AM9530G)
- 1 M DTT, BioUltra (Sigma/Aldrich, cat. no. 43816-10ML)
- Turbo DNase, 2 U/μl (Invitrogen, cat. no. AM2238)
- Triton X-100, molecular biology grade (Calbiochem, cat. no. 648466)
- SUPERase\*In, 20 U/μl (Invitrogen, cat. no. AM2694)
- Sucrose, molecular biology grade (VWR, cat. no. IB37160)
- 3 M NaOAc pH 5.5, RNase-free (Invitrogen, cat. no. AM9740)
- RNase I, 100 U/μl (Invitrogen, cat. no. AM2294) [CRITICAL – Substantial changes in the RNase activity during nuclease footprinting could compromise the experiment]
- 13 × 51 mm polycarbonate ultracentrifuge tube (Beckman, cat. no. 349622)
- miRNeasy RNA isolation kit (Qiagen, cat. no. 217004)
- Chloroform, molecular biology grade (VWR, cat. no. IB05040) [CAUTION – Chloroform is harmful and volatile. Use proper protection when using chloroform and dispose of waste in accordance with institutional regulations]

- Ethanol, molecular biology grade (Sigma/Aldrich, cat. no. E7023—500ML) [CAUTION – Ethanol is highly flammable and volatile]
- Isopropanol, molecular biology grade (VWR, cat. no. 87000-048) [CAUTION – Isopropanol is highly flammable and volatile and is an irritant]
- GlycoBlue, 15 mg/ml (Invitrogen, cat. no. AM9515)
- 0.5 M EDTA, RNase-free (Invitrogen, cat. no. AM9260G)
- Bromophenol blue (Bio-Rad, cat. no. 161-0404)
- Formamide, molecular biology grade (Promega, cat. no. H5051) [CAUTION – Formamide is a reproductive toxin]
- Denaturing 15% polyacrylamide TBE-urea gel, 12 wells (Invitrogen, cat. no. EC68852BOX) [CAUTION – Acrylamide is a neurotoxin. Use proper protection when handling polyacrylamide gels.]
- 10 bp ladder, 1 µg/µl (Invitrogen, cat. no. 10821015)
- Upper size marker oligoribonucleotide NI-NI-19, 5′-AUGUACACGGAGUCGAGCUCAACCCGCAACGCGA-(Phos). The designation (Phos) indicates 3′ phosphorylation. Note that all residues are ribonucleotides.
- Lower size marker oligoribonucleotide NI-NI-20, 5′-AUGUACACGGAGUCGACCCAACGCGA-(Phos). The designation (Phos) indicates 3′ phosphorylation. Note that all residues are ribonucleotides.
- 10x TBE, RNase-free (Promega, cat. no. V4251)
- 10,000x SYBR Gold (Invitrogen, cat. no. S11494) [CAUTION – Nucleic acid stains are typically mutagenic. Use personal protection when handling gel staining solution and dispose of waste in accordance with regulations.]
- 10% SDS, molecular biology grade (Promega, cat. no. V6551) [CAUTION – SDS is an irritant]
- Non-denaturing 8% polyacrylamide TBE gel, 12 wells (Invitrogen, cat. no. EC62162BOX) [CAUTION – Acrylamide is a neurotoxin. Use proper protection when handling polyacrylamide gels.]
- T4 polynucleotide kinase (New England Biolabs, cat. no. M0201S). Supplied with 10x T4 polynucleotide kinase buffer. [CRITICAL -- Avoid the 3′ phosphatase minus mutant, cat. no. M0236S.]
- T4 RNA ligase 2, truncated (New England Biolabs, cat. no. M0242S). Supplied with PEG 8000 50% w/v and 10x T4 Rnl2 buffer
- Preadenylated and 3′-blocked linker: Any of miRNA Cloning Linker 1/5rApp/CTGTAGGCACCATCAAT/3ddC/ (IDT), Universal miRNA Cloning Linker 5′ rAppCTGTAGGCACCATCAAT–NH2 3′ (New England Biolabs, cat. no. S1315S), or AIR adenylated linker A (bioo, cat. no. 510205).
- 10 mM dNTP mix (Invitrogen, cat. no. 18427-013)
- SuperScript III (Invitrogen, cat. no. 18080-093). Supplied with 5x first-strand buffer and 0.1 M DTT
- Reverse transcription primer, 5′-(Phos)-AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCCG



C-(SpC18)-CACTCA-(SpC18)-TTCAGACGTGTGCTCTTCCGATCTATTGATGGTGCCTACAG. The designation (Phos) indicates 5'-phosphorylation and-(SpC18)- indicates a hexa-ethyleneglycol spacer.

- Sodium hydroxide (EMD Chemicals, cat. no. SX0590-1) [CAUTION – Sodium hydroxide is highly corrosive]
- CircLigase (Epicentre, cat. no. CL4111K). Supplied with 10x CircLigase buffer, 1 mM ATP, and 50 mM MnCl<sub>2</sub> [CRITICAL – CircLigase II cannot be substituted.]
- Biotinylated subtraction oligonucleotides. A set of fourteen sequences suitable for subtractive hybridization in mouse and human samples is given in Table 1. Oligonucleotides should be modified by the addition of the 5'-biotin-TEG and purified by HPLC to eliminate unbiotinylated products that could compete with effective, biotinylated molecules during subtraction.
- 20x SSC, RNase-free (Invitrogen, cat. no. AM9763)
- MyOne Streptavidin C1 DynaBeads (Invitrogen, cat. no. 65001)
- Forward library PCR primer, 5'-AATGATACGGCGACCACCGAGATCTACAC
- Indexed reverse library PCR primers, 5'-CAAGCAGAAGACGGCATAACGAGATNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCG. The underlined NNNNNN indicates the reverse complement of the index sequence discovered during Illumina sequencing. The six forward-strand barcode sequences in Table 2 are separated by at least three mismatches from each other and from the index sequence on the multiplexed PhiX control used on Illumina sequencers (e.g. HiSeq2000).
- Phusion polymerase (New England Biolabs, cat. no. M0530S). Supplied with 5x HF buffer.
- Ficoll 400, BioXtra for molecular biology (Sigma/Aldrich, cat. no. F2637)
- High sensitivity DNA kit (Agilent Technologies, cat. no. 5067-4626)
- TruSeq SBS v3 kit, 50 cycles (Illumina, cat. no. FC-401-3002)
- TruSeq SR Cluster Kit v3, cBot, HS (Illumina, cat. no. GD-401-3001)

## Equipment

- Cell lifter (VWR, cat. no. 29442-200)
- Non-stick RNase-free microfuge tubes (Invitrogen, cat. no. AM12450)
- 26-gauge needle (VWR, cat. no. BD305111)
- 1ml syringe (VWR, cat. no. BD309659)
- Pre-sterilized, RNase-free filter pipette tips (Rainin, cat. nos. RT-10F, RT-20F, RT-200F, and RT-1000F)
- Pre-sterilized, RNase-free gel loading pipette tips (National Scientific, cat. no. MN520R-LRS)
- Refrigerated microcentrifuge 5430R (VWR, cat. no. 97027-866)
- Optima TLX Ultracentrifuge (Beckman, cat. no. 361545)
- TLA 100.3 rotor (Beckman, cat. no. 349481)

- Dry block heater (VWR, cat. no. 12621-104)
- Dry block for microfuge tubes (VWR, cat. no. 13259-002)
- Mini-Cell polyacrylamide gel box (Invitrogen, cat. no. EI0001)
- Electrophoresis power supply (VWR, cat. no. 27370-265)
- DarkReader (Clare Chemical Research, cat. no. DR46B). A standard UV transilluminator can be used instead.
- Razors (VWR, cat. no. 55411-050)
- 21-gauge needle (VWR, cat. no. BD 305165) for optional rapid gel extraction.
- Non-stick RNase-free 0.5 ml microfuge tubes (Invitrogen, cat. no. AM12350) for optional rapid gel extraction
- Microfuge tube spin filter (VWR, cat. no. 29442-752) for optional rapid gel extraction.
- Thermal cycler (Bio-Rad, cat. no. 170-9713)
- DynaMag-2 separation rack (Invitrogen, cat. no. 12321D)
- ThermoMixer (VWR, cat. no. 21516-170)
- 2100 BioAnalyzer (Agilent Technologies, cat. no. G2940CA)
- Genome Analyzer II or HiSeq 2000 (Illumina)
- Computer hardware – a 64-bit computer running Linux with at least 4 GB of RAM<sup>47</sup> for the manual analysis options.
- Access to an instance of the Galaxy platform (<http://galaxyproject.org/>) for the Galaxy analysis option<sup>48</sup>.
- FastX-toolkit ([http://hannonlab.cshl.edu/fastx\\_toolkit/index.html](http://hannonlab.cshl.edu/fastx_toolkit/index.html)) and Illumina quality filter software ([http://cancan.cshl.edu/labmembers/gordon/fastq\\_illumina\\_filter/](http://cancan.cshl.edu/labmembers/gordon/fastq_illumina_filter/)) installed locally for the manual analysis options, or through Galaxy for the Galaxy analysis option.
- Bowtie software (<http://bowtie-bio.sourceforge.net/index.shtml>) installed locally for the manual analysis options or through Galaxy for the Galaxy analysis option<sup>49</sup>.
- TopHat software (<http://tophat.ccb.umd.edu/>) installed locally for the manual analysis options or through Galaxy for the Galaxy analysis option<sup>41,47</sup>.
- SAMtools software (<http://samtools.sourceforge.net/>) installed locally for hte manual analysis options or through Galaxy for the Galaxy analysis option<sup>50</sup>.

## Reagent Setup

**Harringtonine**—For optional harringtonine treatment. Dissolve 5 mg harringtonine in 5 ml DMSO to produce a 1 mg/ml solution. Prepare in advance, dispense into 0.5 ml aliquots and store indefinitely at –20°C in the dark. [CAUTION – Harringtonine is highly toxic]

**Polysome buffer**—20 mM Tris\*Cl pH 7.4, 150 mM NaCl, 5 mM MgCl<sub>2</sub>, 1 mM DTT, 100 µg/ml cycloheximide. Prepare fresh with RNase-free reagents and keep on ice. [CAUTION—Cycloheximide is highly toxic and harmful to the environment.]

**Lysis buffer**—Polysome buffer plus 1% Triton X-100 and Turbo DNase I 25 U/ml. Prepare fresh with RNase-free reagents, from a 20% Triton X-100 dilution in RNase-free water.

**Sucrose cushion**—Polysome buffer plus 1 M sucrose (~34% w/v sucrose, 10 ml is 3.4 g sucrose dissolved in 7.8 ml polysome buffer) and 20 U/ml SUPERase\*In. Prepare fresh with RNase-free reagents

**10 mM Tris pH 8**—Prepare in advance with RNase-free reagents and store indefinitely at room temperature (22–25 °C)

**2x Denaturing loading buffer**—Formamide 98% v/v with 10 mM EDTA and 300 µg/ml bromophenol blue. Prepare in advance by dissolving 15 mg bromophenol blue in 1.0 ml 0.5M EDTA and adding 200 µl to 9.8 ml formamide and store indefinitely at room temperature. Other denaturing nucleic acid loading buffers can be substituted, but avoid the dye xylene cyanol, which interferes with visualizing the ligation product band. [CAUTION—Formamide is a reproductive toxin]

**RNA gel extraction buffer**—300 mM NaOAc pH 5.5, 1.0 mM EDTA, 0.25% SDS. Prepare in advance and store indefinitely at room temperature.

**DNA gel extraction buffer**—300 mM NaCl, 10 mM Tris pH 8, 1 mM EDTA. Prepare in advance and store indefinitely at room temperature.

**1 N NaOH**—Prepare in advance and store indefinitely at room temperature [CAUTION—Sodium hydroxide is highly corrosive]

**Subtraction oligo mix**—Oligos at 10 µM each, up to a total oligo concentration of 200 µM, prepared in 10 mM Tris pH 8. Prepare in advance and store at –20°C indefinitely.

**2x Bind/wash buffer**—2 M NaCl, 1 mM EDTA, 5 mM Tris pH 7.5, 0.2% Triton X-100. Prepare in advance and store at room temperature indefinitely.

**6x Non-denaturing loading buffer**—10 mM Tris pH 8, 1 mM EDTA, 15% w/v Ficoll 400, 0.25% bromophenol blue. Prepare in advance and store at room temperature. Other standard DNA non-denaturing electrophoresis loading buffers can be substituted.

## Equipment Setup

**rRNA sequence index**—Start with a Fasta-format sequence file called `rrna_seqs.fa` containing rRNA sequences (e.g. NR\_003285.2, NR\_003286.1, NR\_003287.1, and NR\_023363.1 for human cells). Index this sequence file using `bowtie-build`.

```
bowtie-build rrna_seqs.fa rrna_seqs
```

**Genomic sequence**—Download the genome reference as a Fasta file. The current human genome reference, GRCh37, is available at [ftp://ftp.ncbi.nlm.nih.gov/genbank/genomes/Eukaryotes/vertebrates\\_mammals/Homo\\_sapiens/GRCh37/Primary\\_Assembly/assembled\\_chromosomes/FASTA/](ftp://ftp.ncbi.nlm.nih.gov/genbank/genomes/Eukaryotes/vertebrates_mammals/Homo_sapiens/GRCh37/Primary_Assembly/assembled_chromosomes/FASTA/) with a single compressed Fasta-format file per chromosome. The download can be automated using the following command:

```

for CHR in 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y
do
curl -O
ftp://ftp.ncbi.nlm.nih.gov/genbank/genomes/Eukaryotes/vertebrates_mammals/
Homo_sapiens/GRCh37/Primary_Assembly/assembled_chromosomes/FASTA/chr$
{CHR}.fa.gz
done

```

**Genomic sequence index**—Index the genomic reference Fasta files using bowtie-build. The GRCh37 genome downloaded as a collection of Fasta files, described above, can be indexed with the following command:

```

bowtie-build
chr1.fa.gz,chr2.fa.gz,chr3.fa.gz,chr4.fa.gz,chr5.fa.gz,chr6.fa.gz,chr7.
fa.gz,chr8.fa.gz,chr9.fa.gz,chr10.fa.gz,chr11.fa.gz,chr12.fa.gz,chr13.f
a.gz,chr14.fa.gz,chr15.fa.gz,chr16.fa.gz,chr17.fa.gz,chr18.fa.gz,chr19.
fa.gz,chr20.fa.gz,chr21.fa.gz,chr22.fa.gz,chrX.fa.gz,chrY.fa.gz hg19

```

**Genomic annotation**—Obtain a GTF-format annotation for the exact genome reference sequence downloaded and indexed above. For the GRCh37 human genome sequence, the UCSC genome browser annotations can be downloaded from their table browser (<http://genome.ucsc.edu/cgi-bin/hgTables?org=Human&db=hg19>). Select the group “Genes and Gene Prediction Tracks”, track “UCSC Genes”, table “knownGene”, output format “GTF”, and use “get output” to retrieve the file. Rename it as hg19.gtf.

## PROCEDURE

### Cell lysis (Timing: 30 min)

Note that, in order to carry out initiation profiling, harringtonine treatment must be performed immediately before lysis as described in Box 1.

- 1 Aspirate media from one 10 cm dish of adherent cells. Place the dish on ice, gently wash with 5 ml ice-cold PBS, and aspirate the PBS thoroughly.
- 2 Perform in-dish lysis either without freezing (option A) or with flash-freezing (option B). Flash-freezing may help to maintain in vivo ribosome positions if cell physiology might otherwise change during cell harvesting, but does not affect expression measurements or ribosome density profiles in standard cultured mammalian cells.
  - a. Lysis without freezing
    - i. Drip 400  $\mu$ l ice-cold lysis buffer onto cells, taking care to cover the entire surface of the dish.
  - b. Lysis with flash-freezing
    - i. Quickly immerse the plate in a shallow reservoir of liquid nitrogen
    - ii. Move the dish rapidly to dry ice
    - iii. Drip 400  $\mu$ l ice-cold lysis buffer onto the frozen dish

- iv.** Transfer the dish to wet ice and thaw in the presence of lysis buffer
- 3** Tip the dish and scrape cells down the slope into the lysis buffer pooled in the lower corner. Pipette lysis buffer from this pool back towards the top of the dish and scrape again down the slope of the dish.
- 4** Pipette cells in lysis buffer and withdraw the entire contents of the dish to a microfuge tube on ice. Pipette several times to disperse cell clumps and incubate 10 minutes on ice.
- 5** Triturate cells ten times through a 26 gauge needle.
- 6** Clarify the lysate by centrifugation for 10 minutes at  $20,000 \times g$ ,  $4^{\circ}\text{C}$  and recover the soluble supernatant.

### **Nuclease footprinting and ribosome recovery (Timing: 6 h)**

- 7** Take 300  $\mu\text{l}$  lysate from step 6 and add 7.5  $\mu\text{l}$  RNase I (100 U/ $\mu\text{l}$ ). Incubate 45 min at room temperature with gentle mixing, e.g. on a nutator.
- 8** Add 10.0  $\mu\text{l}$  SUPERase\*In RNase inhibitor to stop nuclease digestion.
- 9** Transfer digestion to a 13 mm  $\times$  51 mm polycarbonate ultracentrifuge tube and underlay 0.90 ml 1M sucrose cushion by carefully positioning a pipette tip (or a cannula or similar tool) at the very bottom of the tube and slowly dispensing the sucrose solution. The lysate should float on top of the sucrose, leaving a visible interface between the layers.
- 10** Pellet ribosomes by centrifugation in a TLA100.3 rotor at 70,000 rpm,  $4^{\circ}\text{C}$  for 4 hr.
- 11** Mark the outside edge of the ultracentrifuge tube, where the ribosome pellet will be found, before removing the tube from the rotor. Gently pipette the supernatant out of the tube. The ribosomal pellet is glassy and translucent, and may not be visible until the supernatant is removed
- 12** Resuspend the ribosomal pellet in 700 $\mu\text{l}$  Qiazol reagent from the miRNeasy kit.
- 13** Purify RNA from the resuspended ribosomal pellet using the miRNeasy kit according to the manufacturer's instructions for purifying total RNA including small RNA. Collect the eluate in a non-stick RNase-free tube. [CRITICAL STEP—From this point through the end of the reverse transcription reaction in Step 39, proper techniques must be employed to avoid RNase contamination. This includes the rigorous use of gloves and RNase-free reagents and consumables].
- 14** Precipitate RNA from the elution by adding 38.5  $\mu\text{l}$  water, 1.5 $\mu\text{l}$  GlycoBlue, and 10.0 $\mu\text{l}$  3M NaOAc pH 5.5, followed by 150 $\mu\text{l}$  isopropanol.
- 15** Chill precipitation for at least 30 minutes on dry ice. [PAUSE POINT—Precipitations may be left overnight on dry ice or at  $-80^{\circ}\text{C}$ ].
- 16** Pellet RNA by centrifugation for 30 min at  $20,000 \times g$ ,  $4^{\circ}\text{C}$  in a tabletop microfuge. Carefully pipette all liquid from the tube, place it in a sideways microfuge tube rack, and allow it to air-dry for 10 min.
- 17** Resuspend the RNA in 5.0  $\mu\text{l}$  10 mM Tris pH 8. [PAUSE POINT – RNA may be stored overnight at  $-20^{\circ}\text{C}$  or for months at  $-80^{\circ}\text{C}$ ].

**Footprint Fragment Purification (Timing: 2.5 h, overnight, 4 h)**

- 18** Pre-run a 15% polyacrylamide TBE-Urea gel at 200 V for 15 min. in 1 × TBE [CRITICAL STEP—The electrophoresis apparatus used for this and subsequent preparative RNA gels must be maintained free of RNase contamination. Decontaminate the tank and electrodes if the equipment has been used for other purposes. Molecular biology grade water obtained directly from the purifier can be tested for nuclease contamination and used to prepare running buffer due to the large volume of nuclease-free water needed for this purpose.]
- 19** Add 5.0 µl 2x denaturing sample buffer to each RNA sample. Prepare a control oligo sample for two lanes with 1.0 ul lower marker oligo 10 uM, 1.0 ul upper marker oligo 10 µM, 8.0 ul 10 mM Tris pH 8, and 10.0 µl 2x denaturing sample buffer. Prepare a ladder sample with 0.5 µl 10 bp ladder 1 µg/µl, 4.5 µl 10 mM Tris pH 8, and 5.0 ul 2x denaturing sample buffer.
- 20** Denature samples for 90 seconds at 80°C.
- 21** Load samples on the polyacrylamide gel with control oligo sample (mixed upper and lower markers) on either side of the RNA samples.
- 22** Separate by electrophoresis for 65 minutes at 200 V.
- 23** Stain the gel for 3 minutes with 1x SYBR Gold in 1x TBE running buffer on a gentle shaker
- 24** Visualize the gel and excise the 26 nt – 34 nt region demarcated by the marker oligos NI-NI-19 and NI-NI-20 from each footprinting sample. Place each excised gel slice in a clean non-stick RNase-free microfuge tube. Similarly excise the marker oligo bands from the gel and place them all in a microfuge tube as well. These oligos will be processed the same way as the samples for the remainder of the procedure, as an internal control.
- 25** Extract RNA from the polyacrylamide gel slices using either of the gel extraction protocols described in Option A or Option B. Rapid gel extraction (option A) is faster, while overnight gel extraction (option B) may provide more reproducibly high yields.
  - a.** Rapid Gel Extraction
    - i.** Pierce the bottom of an 0.5 ml RNase-free microfuge tube with a 21 gauge needle and cut off the cap.
    - ii.** Nest the pierced small tube inside a 1.5 ml RNase-free microfuge tube and place the gel slice in the inner tube.
    - iii.** Spin the tube 2 minutes at full speed in a tabletop microcentrifuge to force the gel slice through the needle hole.
    - iv.** Transfer any remaining gel debris from the pierced 0.5 ml microfuge tube and discard the pierced tube.
    - v.** Add 360 µl RNase-free water to the gel debris.
    - vi.** Incubate 10 min. at 70°C.
    - vii.** Cut the tip off of a 1000 µl pipet tip.
    - viii.** Transfer all liquid and gel slurry into a microfuge tube spin filter.



- ix. Spin 2 min. at full speed in a tabletop microcentrifuge to recover all liquid from the gel slurry.
- x. Transfer the filtrate to a fresh RNase-free non-stick microfuge tube and add 40  $\mu$ l 3M NaOAc.

**b. Overnight Gel Extraction**

- i. Add 400  $\mu$ l RNA gel extraction buffer and freeze samples for 30 min on dry ice.
  - ii. Leave samples overnight at room temperature with gentle mixing, e.g. on a nutator.
  - iii. Briefly centrifuge gel extractions to collect the liquid at the bottom of the tube. Transfer 400  $\mu$ l eluate into a clean non-stick RNase-free microfuge tube.
- 26** Precipitate RNA by adding 1.5  $\mu$ l GlycoBlue, mixing, and then adding 500  $\mu$ l isopropanol. Recover RNA as described in Steps 15 and 16 [PAUSE POINT—Precipitations may be left on dry ice or at  $-80^{\circ}\text{C}$  overnight as described in Step 15.
- 27** Resuspend size-selected RNA in 10.0  $\mu$ l 10 mM Tris pH 8 and transfer to a clean non-stick RNase-free microfuge tube. [PAUSE POINT—RNA may be stored overnight at  $-20^{\circ}\text{C}$  or indefinitely at  $-80^{\circ}\text{C}$ ]
- 28** Prepare dephosphorylation reaction by adding 33  $\mu$ l RNase-free water to the samples from step 27 and denaturing 90 seconds at  $80^{\circ}\text{C}$ . Equilibrate to  $37^{\circ}\text{C}$ , set up the reaction tabulated below and incubate 1 hour at  $37^{\circ}\text{C}$ . Then heat-inactivate the enzyme for 10 min at  $70^{\circ}\text{C}$ .

Component	Amount per reaction ( $\mu$ l)	Final
RNA sample	43.	
T4 PNK buffer (10x)	5.0	1x
SUPERase*In (20 U/ $\mu$ l)	1.0	20 U
T4 PNK (10 U/ $\mu$ l)	1.0	10 U

- 29** Precipitate RNA by adding 39  $\mu$ l water, 1.0  $\mu$ l GlycoBlue, 10.0  $\mu$ l 3M NaOAc, mixing, and adding 150  $\mu$ l isopropanol. Recover RNA as described in Steps 15 and 16. [PAUSE POINT—Precipitations may be left on dry ice or at  $-80^{\circ}\text{C}$  overnight as described in Step 15]

**Linker Ligation (Timing: ~ 6 h, overnight, ~1.5 h)**

- 30** Resuspend dephosphorylated RNA in 8.5  $\mu$ l 10 mM Tris pH 8 and transfer to a clean non-stick RNase-free microfuge tube. [PAUSE POINT—RNA may be stored overnight at  $-20^{\circ}\text{C}$  or indefinitely at  $-80^{\circ}\text{C}$ ].
- 31** Add 1.5  $\mu$ l preadenylated linker (0.5  $\mu\text{g}/\mu\text{l}$ ) and denature 90 seconds at  $80^{\circ}\text{C}$ , then cool to room temperature.
- 32** Set up the ligation reaction below and incubate 2.5 h at room temperature:

Component	Amount per reaction (μl)	Final
RNA and linker	10.0	
T4 Rnl2 buffer (10x)	2.0	1x
PEG 8000 (50% w/v)	6.0	15% w/v
SUPERase*In (20 U/μl)	1.0	20 U
T4 PNK (10 U/μl)	1.0	10 U

- 33** Add 338 μl water, 40 μl 3M NaOAc pH 5.5, and 1.5 μl GlycoBlue to each reaction, followed by 500 μl isopropanol. Recover RNA as described in Steps 15 and 16.
- 34** Separate the ligation reactions by polyacrylamide gel electrophoresis as described in Steps 18 – 23.
- 35** Excise the ligation product bands, including the marker oligo ligation, and place each gel slice in a clean non-stick RNase-free microfuge tube. Recover RNA from these samples as described in Steps 25 – 26. [PAUSE POINT—Precipitations may be left on dry ice or at –80°C overnight]

### Reverse Transcription (Timing: ~ 5 h, overnight, ~1.5 h)

- 36** Resuspend the ligation product in 10.0 μl 10 mM Tris pH 8 and transfer to a clean PCR tube. [PAUSE POINT—RNA may be stored at –20°C overnight or indefinitely at –80°C].
- 37** Add 2.0 μl reverse transcription primer at 1.25 μM. Denature 2 min at 80°C in a thermal cycler and then place on ice. Cool the thermal cycler to 48°C.
- 38** Set up the reverse transcription reaction tabulated below and incubate 30 min at 48°C in the thermal cycler:

Component	Amount per reaction (μl)	Final
Ligation and primer	12.0	
First-strand buffer (5x)	4.0	1x
dNTPs (10 mM)	1.0	0.5 mM
DTT (0.1M)	1.0	5 mM
SUPERase*In (20 U/μl)	1.0	20 U
SuperScript III (200 U/μl)	1.0	200 U

- 39** Hydrolyze RNA by adding 2.2 μl 1N NaOH to each reaction and incubate 20 min at 98°C. The GlycoBlue dye will turn pink.
- 40** Add 20 μl 3M NaOAc pH 5.5, 2.0 μl GlycoBlue, and 156 ul water to each reverse transcription reaction, followed by 300 ul isopropanol. Recover RNA from the precipitation as described in Steps 15 and 16.
- 41** Separate the reverse transcription products from the unextended primer by polyacrylamide gel electrophoresis as described in Steps 18 – 23. Omit the preparation of marker oligo samples and instead prepare one sample with 2.0 μl reverse transcription primer (1.25 μM), 3.0 μl 10 mM Tris pH 8, and 5.0 μl 2x denaturing sample buffer.

- 42 Excise the reverse transcription product bands from the gel and place each in a clean non-stick RNase-free microfuge tube.
- 43 Extract DNA from the polyacrylamide gel using either of the gel extraction protocols described in Steps 25A or 25B, except that it is no longer necessary to use RNase-free reagents, though non-stick tubes are still required, and overnight extraction should be performed in the DNA gel extraction buffer rather than the RNA gel extraction buffer.
- 44 Precipitate DNA by adding 1.5  $\mu$ l GlycoBlue, mixing, and then adding 500  $\mu$ l isopropanol. Recover DNA as described in Steps 15 and 16. [PAUSE POINT—Precipitations may be left on dry ice or at  $-80^{\circ}\text{C}$  indefinitely]

#### Circularization (Timing: 1.5 h)

- 45 Resuspend reverse transcription products in 15.0  $\mu$ l 10 mM Tris pH 8 and transfer to a PCR tube. [PAUSE POINT—DNA may be stored indefinitely at  $-20^{\circ}\text{C}$ ]
- 46 Prepare the circularization reaction tabulated below and incubate for 1 hr at  $60^{\circ}\text{C}$  and then heat-inactivate for 10 min at  $80^{\circ}\text{C}$  in a thermal cycler:

Component	Amount per reaction ( $\mu$ l)	Final
First-strand cDNA	15.0	
CircLigase buffer (10x)	2.0	1x
ATP (1 mM)	1.0	50 mM
MnCl <sub>2</sub> (50 mM)	1.0	2.5 mM

[PAUSE POINT—Circularized DNA may be stored in the circularization reaction buffer indefinitely at  $-20^{\circ}\text{C}$ ]

#### rRNA Depletion (Timing: 2.5 h)

- 47 Combine 5.0  $\mu$ l circularization reaction with 1.0  $\mu$ l subtraction oligo pool, 1.0  $\mu$ l 20x SSC, and 3.0  $\mu$ l water in a PCR tube.
- 48 Place PCR tubes in a thermal cycler and denature 90 sec at  $100^{\circ}\text{C}$ , and then anneal at  $0.1^{\circ}\text{C}/\text{second}$  to  $37^{\circ}\text{C}$ . Incubate 15 min at  $37^{\circ}\text{C}$ . Warm a ThermoMixer to  $37^{\circ}\text{C}$ .
- 49 Vortex MyOne Streptavidin C1 DynaBeads (10 mg/ml) vigorously to resuspend beads. Use 25.0  $\mu$ l beads per subtraction reaction, plus an additional 12.5  $\mu$ l. Transfer beads to a clean non-stick microfuge tube and place the tube on a magnetic rack for 1 minute to isolate beads. Gently withdraw all liquid from the tube, remove the tube from the rack, and resuspend in 1 volume (i.e., 25.0  $\mu$ l per subtraction reaction, plus an additional 12.5  $\mu$ l) 1x bind/wash buffer. Repeat this procedure twice more.
- 50 Place the beads on a magnetic rack for 1 min to isolate beads, withdraw the final wash, and resuspend in 0.4 volumes (i.e., 10.0  $\mu$ l per subtraction reaction, plus an additional 5.0  $\mu$ l) 2x bind/wash buffer. Take one 10.0  $\mu$ l aliquot of beads per subtraction reaction into a clean non-stick microfuge tube. Place bead aliquots in the ThermoMixer at  $37^{\circ}\text{C}$  and equilibrate.

- 51 Transfer 10.0  $\mu\text{l}$  subtraction reaction directly from the PCR tube in the thermal cycler (from step 46) to a bead aliquot in the ThermoMixer. Incubate 15 min at 37°C with mixing at 1000 rpm.
- 52 Transfer tubes directly from the ThermoMixer to a magnetic rack and isolate beads for one minute. Recover 17.5  $\mu\text{l}$  eluate from the depletion and transfer it to a new non-stick microfuge tube.
- 53 Add 2.0  $\mu\text{l}$  GlycoBlue, 6.0  $\mu\text{l}$  5M NaCl, and 74.  $\mu\text{l}$  water to each depletion, followed by 150  $\mu\text{l}$  isopropanol. Recover DNA as described in Steps 15 and 16. [PAUSE POINT—Precipitations may be left indefinitely on dry ice or at –80°C]
- 54 Resuspend depleted DNA in 5.0  $\mu\text{l}$  10 mM Tris pH 8. [PAUSE POINT—DNA may be stored indefinitely at –20°C]

### PCR Amplification and Barcode Addition (Timing: ~2 h, overnight, ~ 2 h)

- 55 Prepare a 100  $\mu\text{l}$  PCR mixture for each sample, according to the table below. Use a different indexing primer for each sample.

Component	Amount per reaction ( $\mu\text{l}$ )	Final
Phusion HF Buffer (5x)	20	1x
dNTPs (10 mM)	2.0	0.2 mM
Forward Library Primer (100 $\mu\text{M}$ )	0.5	0.5 $\mu\text{M}$
Reverse Indexed Primer (100 $\mu\text{M}$ )	0.5	0.5 $\mu\text{M}$
Circularized DNA template (from step 54)	5.0	
Nuclease-free water	71.0	
Phusion polymerase (2 U/ $\mu\text{l}$ )	1.0	2 U

- 56 Set up 5 PCR tube strips and transfer a 16.7  $\mu\text{l}$  aliquot of the PCR mixture into one tube in each strip.
- 57 Perform PCR amplification with varying numbers of cycles by placing all strip tubes in the thermal cycler and starting a program with the conditions below. 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 thermal cycler until the end of cycle 14.

Cycle number	Denature	Anneal	Extend
1	98°C, 30 s		
2–15	98°C, 10 s	65°C, 10 s	72°C, 5 s

- 58 Add 3.3  $\mu\text{l}$  6x non-denaturing loading dye to each reaction. Prepare a ladder sample with 1.0  $\mu\text{l}$  10 bp ladder, 15.7  $\mu\text{l}$  10 mM Tris pH 8, and 3.3  $\mu\text{l}$  6x non-denaturing loading dye.
- 59 Set up one or two 8% polyacrylamide non-denaturing gels. Load amplification reactions for the same sample in adjacent wells to facilitate direct comparison.

- 60** Separate by electrophoresis for 40 minute at 180 V. Stain the gel for 3 min in 1x SYBR Gold in 1x TBE gel running buffer.
- 61** Visualize the gel and excise the amplified PCR product. Select one or two reactions for each cycle with a prominent product band but little accumulation of re-annealed partial duplex library products (see Fig 4d). Avoid any lower product band derived from unextended reverse transcription primer. Place excised gel slices in clean, non-stick microfuge tubes.
- 62** Recover DNA from the gel slices as described in Steps 42–44, using the overnight gel extraction option. [CRITICAL STEP -- It is particularly important that the gel extractions remain at 25°C or below to avoid the formation of re-annealed partial duplexes. Such duplexes will complicate the quantitation of the library.] [PAUSE POINT—Precipitations may be stored overnight on dry ice or indefinitely at –80°C].
- 63** Resuspend library DNA in 15.0 µl 10 mM Tris pH 8. [PAUSE POINT—Double-stranded DNA may be stored indefinitely at 4°C or at –20°C].
- 64** Quantify and characterize the library by preparing 1.5 µl library with 6.0 µl water and using the High Sensitivity DNA chip on the Agilent BioAnalyzer according to the manufacturer’s protocol.

### Sequencing and Analysis

- 65** Sequence the library according on the Illumina GAI or HiSeq system according to the manufacturer’s protocol. The sequencing libraries use the standard Illumina genomic first-read sequencing primer for footprint sequencing and the standard Illumina indexing primer for index sequencing.
- 66** Preprocess and align the sequencing data (Fig. 3) produced by the CASAVA 1.8 pipeline using one of Options A–C, based on availability and familiarity of a local Linux computer for manual analysis or a suitable Galaxy server. Note that a history containing a sample analysis of one million footprints is available from the public Galaxy server (<https://main.g2.bx.psu.edu/u/ingolia/h/ribosome-footprint-alignment>).
- a.** Manual analysis
- i.** Preprocess the sequencing data produced by the CASAVA 1.8 pipeline by discarding low-quality reads, trimming the linker sequence from the 3’ end of each sequencing read, and removing the first nucleotide from the 5’ end of each read, as it frequently represents an untemplated addition during reverse transcription. The standard CASAVA 1.8 output is a collection of gzip-compressed FastQ files in a directory named Project\_YYY/Sample\_XXX. To perform all preprocessing steps in series, use the following command:

```
zcat/path/to/Project_YYY/Sample_XXX/*.fastq.gz
| \
fastq_illumina_filter --keep N-v | \
fastx_clipper -Q33 -a CTGTAGGCACCATCAAT -l 25 -
```

```
c -n-v | \
fastx_trimmer -Q33 -f 2 > XXX_trimmed.fq
```

- ii. Align trimmed sequencing reads to an rRNA reference using the Bowtie short-read alignment program, discard the rRNA alignments, and collect unaligned reads.

```
bowtie --seedlen=23 --un=XXX_norrna.fq
rrna_seqs >/dev/null
```

- iii. Align non-rRNA sequencing reads to a genomic reference using the TopHat splicing-aware short read alignment program.

```
tophat --no-novel-juncs --output-dir
XXX_vs_genome \
--GTF hg19.gtf hg19 XXX_norrna.fq
```

- iv. Extract perfect-match alignments from TopHat output.

```
samtools view -h XXX_vs_genome/
accepted_hits.bam | \
grep -E '(NM:i:0)|(^@)' | \
samtools view -S b - > XXX_vs_genome.bam
```

#### b. Semi-automated local analysis

- i. Download the pre-processing and alignment script, supplied as a Supplementary Note, and rename it Makefile.
- ii. Edit the four variables at the top of the script to contain filenames for the rRNA sequence index (RRNA\_EBWT), the genome sequence index (GENOME\_EBWT), the genome annotation (GENOME\_GTF), and the project directories generated by CASAVA 1.8 that contain sequencing data (PROJECT\_DIRS).
- iii. Run the analysis by typing 'make'.

#### c. Galaxy analysis.

- i. Upload the FastQ file containing footprint sequences with "Get Data/Upload File" as fastqsanger format.
- ii. Upload the Fasta file of rRNA sequences with "Get Data/Upload File".
- iii. Obtain genome annotations with "Get Data/UCSC Main", selecting the GTF output format and sending output to Galaxy.
- iv. Clip the adapter sequence using "NGS: QC and Manipulation/Clip" specifying a minimum length of 25 nucleotides, and enter a custom adapter sequence



CTGTAGGCACCATCAAT, do not discard sequences with unknown bases, and output only clipped sequences.

- v. Trim the adapter sequence using “NGS: QC and Manipulation/Trim sequences” specifying the first base to keep as 2 and the last base to keep as 50.
- vi. Map pre-processed reads to the rRNA database using “NGS: Mapping/Bowtie” selecting a reference genome from the history and then choosing the uploaded rRNA file. Select the full parameter list and select the option to write all reads that could not be aligned.
- vii. Map the unaligned reads to the genome using “NGS: RNA Analysis/Tophat for Illumina” using the appropriate built-in index (e.g., hg19 for human cultured cells). Select the full parameter list and specify a FR First Strand library, choose to “Use Own Junctions”, then “Use Gene Annotation Model” and select the GTF format genome annotation from (iii). Also choose to “Only look for supplied junctions”.

## TIMING

Steps 1 – 6, Cell lysis: ~30 min

Steps 7 – 17, Nuclease footprinting, and ribosome recovery: ~6 h

Steps 18 – 29, Footprint fragment purification: ~2.5 h, overnight gel extraction, ~ 4 h the following day

Steps 30 – 35, Linker ligation: ~ 6 h, overnight gel extraction, ~1.5 h the following day

Steps 36 – 44, Reverse transcription: ~ 5 h, overnight gel extraction, ~1.5 h the following day

Steps 45 – 46, Circularization: ~1.5 h

Steps 47–54, rRNA depletion: ~2.5 h

Steps 55–64, PCR amplification and barcode addition: ~2 h, overnight gel extraction, ~ 2 h the following day

Steps 65–66, Sequencing and analysis: ~4 h, followed by ~48 h of sequencing and indefinite analysis.

## TROUBLESHOOTING

**Step 34:** Low yield in linker ligation will result in more unligated RNA at ~30 nt and less ligated RNA product at ~50 nt (see Fig. 4b). One common cause is failure of the dephosphorylation reaction, which is sensitive to residual salt from precipitated RNA as well as to other contaminants. Note that the lower and upper size marker RNAs are chemically phosphorylated and serve as an internal control for both dephosphorylation and subsequent linker ligation. Take care to remove all liquid from the RNA pellet, dry it thoroughly, and resuspend the pellet in a small volume while avoiding residual salt on other parts of the precipitation tube before transferring it to a new, clean tube.

**Step 61:** A lower ~145 nt band from the PCR represents background derived from an unextended RT primer. When the amount of RT product is unusually low, this background

will comprise a greater fraction of the total DNA. To decrease this background, excise the RT product band precisely and avoid the background haze. In some cases, reducing the amount of RT primer may help as well.

**Step 61:** A broad, slower-migrating smear indicates excessive PCR amplification. When the PCR amplification consumes a large fraction of the total oligonucleotides present in the reaction mixture, re-annealing of library strands becomes kinetically competitive with primer annealing. Re-annealed library duplexes have long complementary sequences on each end, but typically contain non-complementary inserts, causing slow and heterogeneous migration relative to the fully complementary library duplex. Use product bands from reactions with fewer PCR cycles.

## ANTICIPATED RESULTS

The protocol typically produces 550–600  $\mu$ l lysate. RNA extraction from an aliquot of this lysate indicates a yield of 25–50  $\mu$ g total RNA from one 10 cm dish of 50%–80% confluent HEK293 cells. The RNA yield from the footprinting pellet is typically 40%–50% of the total RNA input, resulting in 6  $\mu$ g – 15  $\mu$ g of RNA from 300  $\mu$ l lysate. The lost RNA includes non-coding RNAs that do not enter the sucrose cushion, such as tRNAs, as well as mRNA and rRNA that is degraded during the footprinting digest. We have successfully prepared libraries from as little as 2  $\mu$ g of ribosomal pellet RNA.

Gel electrophoresis of the footprinting RNA will reveal a broad array of specific and fairly reproducible bands (Fig. 4a), most presumably derived from the rRNA. The marker oligos will guide the excision of the gel region that contains the footprint fragments, which may not be visible as a discrete band (Fig. 4a). They also provide a positive control through subsequent PAGE purification steps. The ligated control oligos indicate a specific region that should be excised in the linker ligation reaction (Fig. 4b). While the marker reverse transcription products do still produce a discernible doublet (Fig. 4c), the RT product in general forms a much tighter band because the relative length variation is lower, and it is not necessary to excise a broad region. The PCR products should produce a discrete band that is ~175 nt long (Fig. 4d).

Deep sequencing data from the HEK293 cell ribosome footprinting presented here are available for download from NCBI's Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo>) under accession number GSE37744.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank members of the Weissman and Ingolia labs as well as H. Guo, D. Bartel, S. Luo, and G. Schroth for advice in developing this protocol. This work was supported by the NIH through an NIH P01 (AG10770) (J.S.W.) and a Ruth L. Kirschstein NRSA (GM080853) (N.T.I.), an American Cancer Society Postdoctoral fellowship (117945-PF-09-136-01-RMC) (G.A.B.) and the Searle Scholars Program (N.T.I.)

## References

1. Brown PO, Botstein D. Exploring the new world of the genome with DNA microarrays. *Nat Genet.* 1999; 21:33–7. [PubMed: 9915498]
2. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods.* 2008; 5:621–8. [PubMed: 18516045]

3. Nagalakshmi U, et al. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*. 2008; 320:1344–9. [PubMed: 18451266]
4. Ghaemmaghami S, et al. Global analysis of protein expression in yeast. *Nature*. 2003; 425:737–41. [PubMed: 14562106]
5. Lu P, Vogel C, Wang R, Yao X, Marcotte EM. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol*. 2007; 25:117–24. [PubMed: 17187058]
6. Schwanhausser B, et al. Global quantification of mammalian gene expression control. *Nature*. 2011; 473:337–42. [PubMed: 21593866]
7. Vogel C. Translation's coming of age. *Mol Syst Biol*. 2011; 7:498. [PubMed: 21613985]
8. Sonenberg N, Hinnebusch AG. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell*. 2009; 136:731–45. [PubMed: 19239892]
9. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*. 2009; 324:218–23. [PubMed: 19213877]
10. Steitz JA. Polypeptide chain initiation: nucleotide sequences of the three ribosomal binding sites in bacteriophage R17 RNA. *Nature*. 1969; 224:957–64. [PubMed: 5360547]
11. Wolin SL, Walter P. Ribosome pausing and stacking during translation of a eukaryotic mRNA. *EMBO J*. 1988; 7:3559–69. [PubMed: 2850168]
12. Bentley DR, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 2008; 456:53–9. [PubMed: 18987734]
13. Ingolia NT, Lareau LF, Weissman JS. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*. 2011; 147:789–802. [PubMed: 22056041]
14. Hsieh AC, et al. The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature*. 2012
15. Kondo T, et al. Small peptides switch the transcriptional activity of Shavenbaby during *Drosophila* embryogenesis. *Science*. 2010; 329:336–9. [PubMed: 20647469]
16. Ingolia NT. Genome-wide translational profiling by ribosome footprinting. *Methods Enzymol*. 2010; 470:119–42. [PubMed: 20946809]
17. Brar GA, et al. High-Resolution View of the Yeast Meiotic Program Revealed by Ribosome Profiling. *Science*. 2011
18. Trapnell C, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010; 28:511–5. [PubMed: 20436464]
19. Dinger ME, Pang KC, Mercer TR, Mattick JS. Differentiating protein-coding and noncoding RNA: challenges and ambiguities. *PLoS Comput Biol*. 2008; 4:e1000176. [PubMed: 19043537]
20. Robert F, et al. Altering chemosensitivity by modulating translation elongation. *PloS one*. 2009; 4:e5428. [PubMed: 19412536]
21. Fresno M, Jimenez A, Vazquez D. Inhibition of translation in eukaryotic systems by harringtonine. *European journal of biochemistry/FEBS*. 1977; 72:323–30. [PubMed: 319998]
22. Levin JZ, et al. Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nature methods*. 2010; 7:709–15. [PubMed: 20711195]
23. Li GW, Oh E, Weissman JS. The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature*. 2012
24. Qian W, Yang JR, Pearson NM, Maclean C, Zhang J. Balanced Codon Usage Optimizes Eukaryotic Translational Efficiency. *PLoS Genet*. 2012; 8:e1002603. [PubMed: 22479199]
25. Stadler M, Fire A. Wobble base-pairing slows in vivo translation elongation in metazoans. *RNA*. 2011; 17:2063–73. [PubMed: 22045228]
26. Oh E, et al. Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. *Cell*. 2011; 147:1295–308. [PubMed: 22153074]

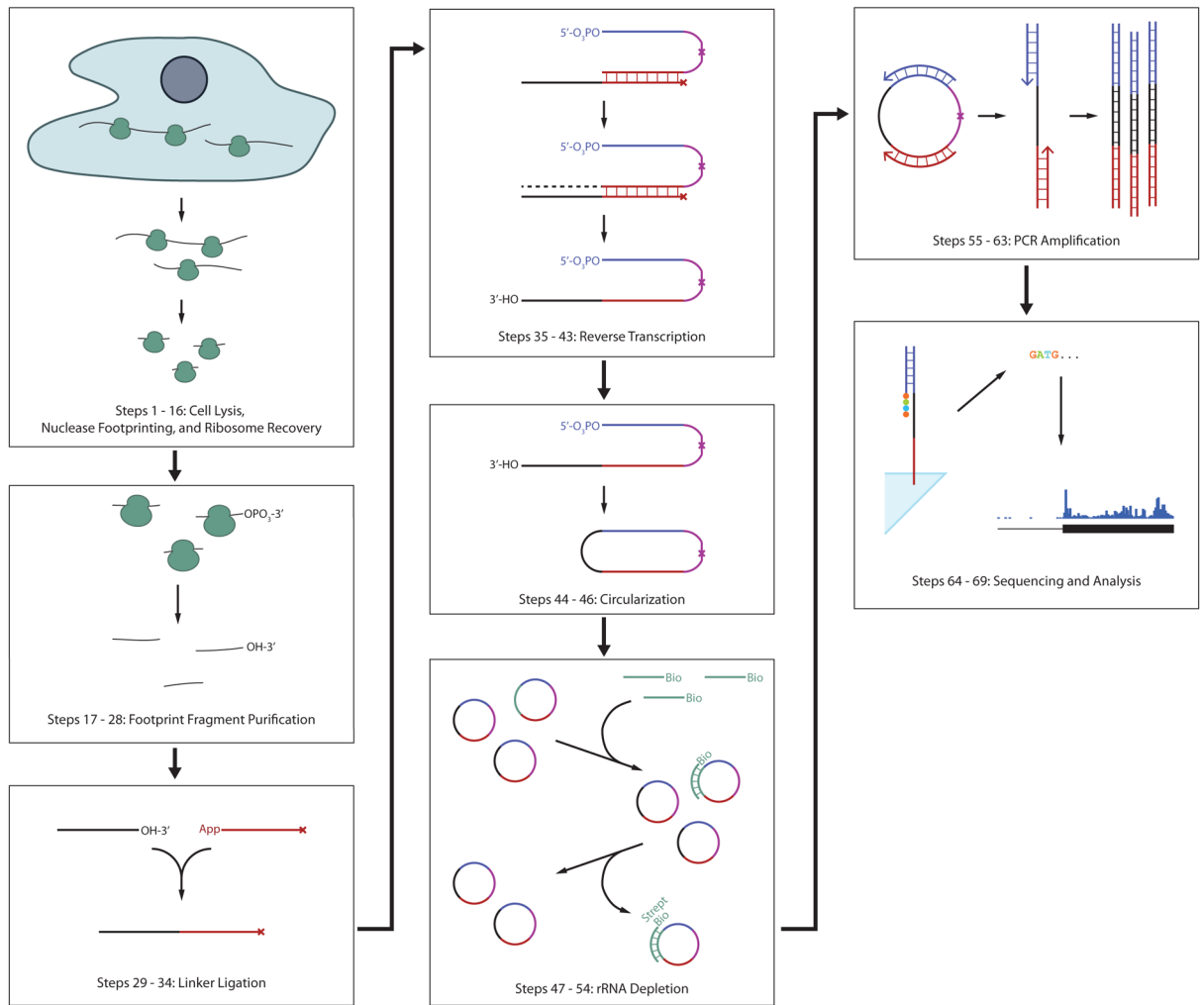
27. Reid DW, Nicchitta CV. Primary role for endoplasmic reticulum-bound ribosomes in cellular translation identified by ribosome profiling. *The Journal of biological chemistry*. 2012; 287:5518–27. [PubMed: 22199352]
28. Arava Y, et al. Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A*. 2003; 100:3889–94. [PubMed: 12660367]
29. Sampath P, et al. A hierarchical network controls protein translation during murine embryonic stem cell self-renewal and differentiation. *Cell stem cell*. 2008; 2:448–60. [PubMed: 18462695]
30. del Prete MJ, Vernal R, Dolznig H, Mullner EW, Garcia-Sanz JA. Isolation of polysome-bound mRNA from solid tissues amenable for RT-PCR and profiling experiments. *RNA*. 2007; 13:414–21. [PubMed: 17237355]
31. Blanchard SC, Kim HD, Gonzalez RL Jr, Puglisi JD, Chu S. tRNA dynamics on the ribosome during translation. *Proceedings of the National Academy of Sciences of the United States of America*. 2004; 101:12893–8. [PubMed: 15317937]
32. Jelenc PC. Rapid purification of highly active ribosomes from *Escherichia coli*. *Anal Biochem*. 1980; 105:369–74. [PubMed: 7006445]
33. Heiman M, et al. A translational profiling approach for the molecular characterization of CNS cell types. *Cell*. 2008; 135:738–48. [PubMed: 19013281]
34. Sanz E, et al. Cell-type-specific isolation of ribosome-associated mRNA from complex tissues. *Proc Natl Acad Sci U S A*. 2009; 106:13939–44. [PubMed: 19666516]
35. Lau NC, Lim LP, Weinstein EG, Bartel DP. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science*. 2001; 294:858–62. [PubMed: 11679671]
36. Hansen KD, Brenner SE, Dudoit S. Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic acids research*. 2010; 38:e131. [PubMed: 20395217]
37. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome biology*. 2010; 11:R106. [PubMed: 20979621]
38. Bullard JH, Purdom E, Hansen KD, Dudoit S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC bioinformatics*. 2010; 11:94. [PubMed: 20167110]
39. Li J, Jiang H, Wong WH. Modeling non-uniformity in short-read rates in RNA-Seq data. *Genome biology*. 2010; 11:R50. [PubMed: 20459815]
40. Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L. Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome biology*. 2011; 12:R22. [PubMed: 21410973]
41. Trapnell C, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology*. 2010; 28:511–5.
42. Hafner M, et al. RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. *RNA*. 2011; 17:1697–712. [PubMed: 21775473]
43. Linsen SE, et al. Limitations and possibilities of small RNA digital gene expression profiling. *Nature methods*. 2009; 6:474–6. [PubMed: 19564845]
44. Zhuang F, Fuchs RT, Sun Z, Zheng Y, Robb GB. Structural bias in T4 RNA ligase-mediated 3'-adapter ligation. *Nucleic acids research*. 2012
45. Garber M, Grabherr MG, Guttman M, Trapnell C. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nature methods*. 2011; 8:469–77. [PubMed: 21623353]
46. Guo H, Ingolia NT, Weissman JS, Bartel DP. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*. 2010; 466:835–40. [PubMed: 20703300]
47. Trapnell C, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols*. 2012; 7:562–78.
48. Goecks J, Nekrutenko A, Taylor J. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome biology*. 2010; 11:R86. [PubMed: 20738864]
49. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*. 2009; 10:R25. [PubMed: 19261174]

50. Li H, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25:2078–9. [PubMed: 19505943]

\$watermark-text

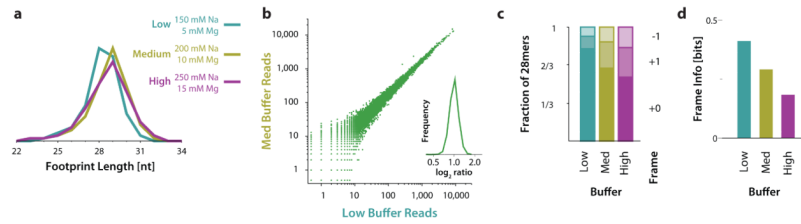
\$watermark-text

\$watermark-text



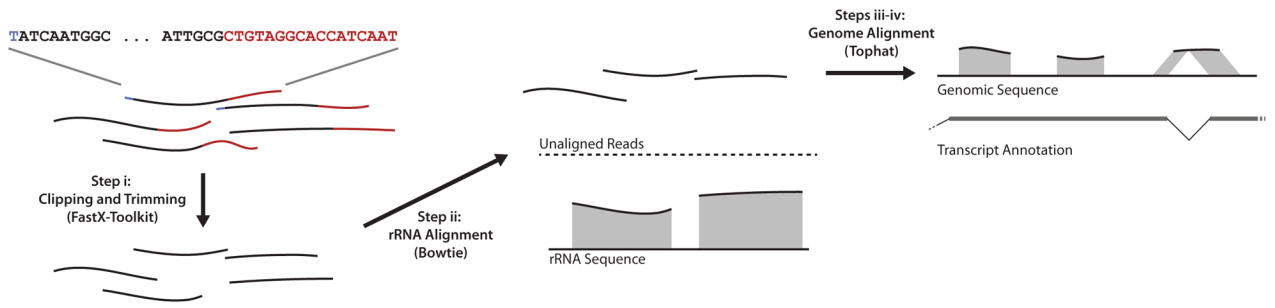
**Figure 1.**  
Overview of the ribosome profiling protocol.





**Figure 2.**

Buffer conditions affect footprint precision but not expression measurements. (a) Length distribution of ribosome footprints obtained from nuclease digestion in different buffer conditions, as indicated. (b) Expression measurements (average footprint density across each message) by ribosome profiling in different buffer conditions. Each point on the scatter plot represents a human gene, and the expression shown on the two axes is the total number of ribosome footprints in each sample that aligned to the canonical isoform of the gene, excluding those mapping to the first 15 or last 5 codons, where drug treatment distorts ribosome occupancy. The histogram shows the distribution of  $\log_2$  ratios between medium and low salt/magnesium buffer conditions for genes with at least 200 total footprints in the two measurements. This criterion avoids low-expression genes where statistical sampling error dominates inter-replicate differences. (c) Sub-codon position of ribosome footprint 5' termini, obtained by nuclease digestion in different buffer conditions, as indicated. Reads that aligned with the annotated CDS of canonical transcripts in the UCSC Known Genes dataset were used, and the position of their 5' terminus was determined relative to codon boundaries in this CDS. (d) Conditional entropy of the ribosome footprint length and reading-frame position distribution in different samples.

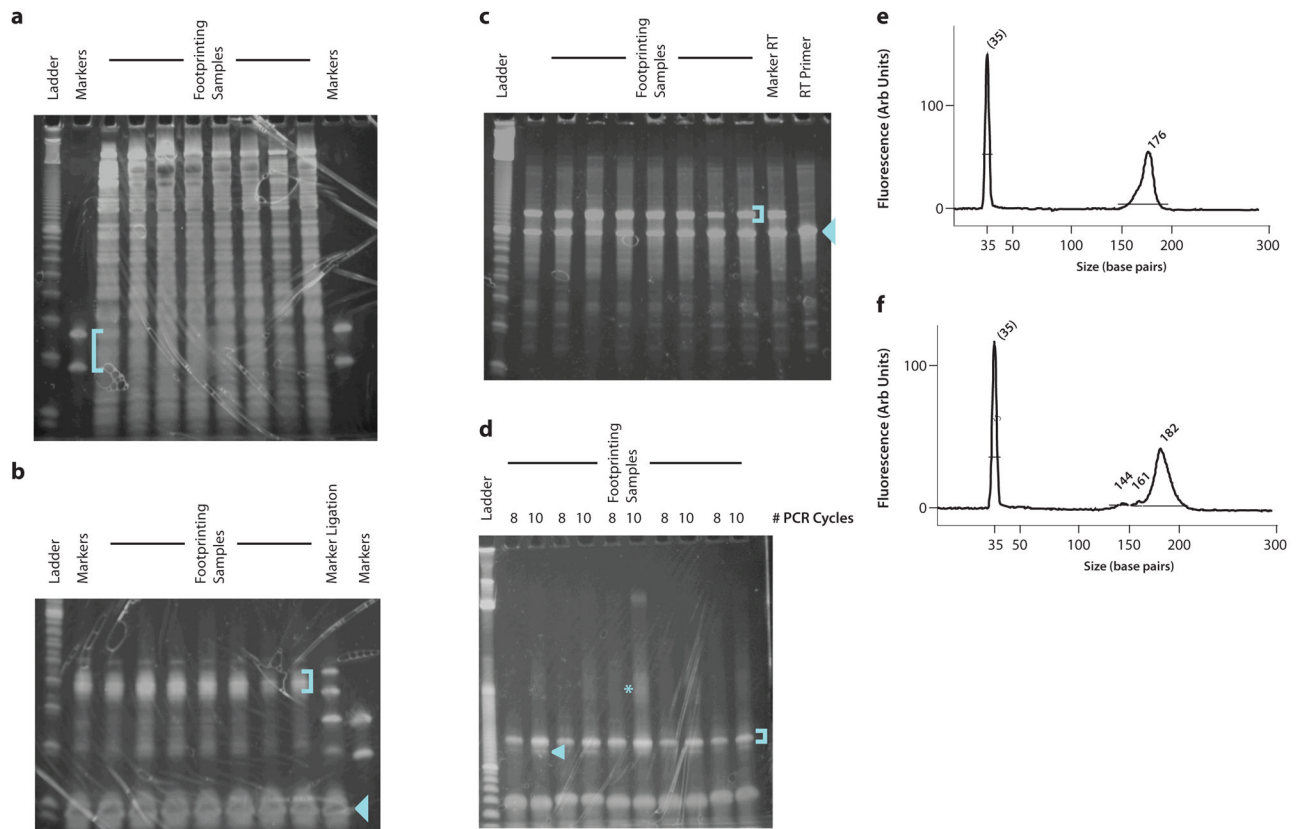


**Figure 3.** Overview of ribosome footprinting sequence pre-processing and alignment.

\$watermark-text

\$watermark-text

\$watermark-text



**Figure 4.**

Representative gels from intermediate product purification. (a) Size selection of ribosome footprint fragments. The footprinting samples are derived from HeLa lysates with 5–15  $\mu$ g input RNA. The blue bracket indicates the gel region that should be excised. (b) Purification of ligation products. Two marker samples are shown, one of which contains only the lower and upper marker oligonucleotides, the other of which was produced by carrying forward the markers from the size selection gel through dephosphorylation and ligation. The blue bracket indicates the gel region that should be excised. The blue arrowhead indicates the unreacted linker. (c) Purification of reverse transcription products. The blue bracket indicates the gel band that should be excised. The blue arrowhead indicates the unextended RT primer, which should be avoided. (d) Purification of PCR products. The blue bracket indicates the ~175 nt product band that should be purified. The blue arrowhead indicates the ~145 nt background band derived from unextended RT primer that should be avoided. The blue asterisk indicates the partial duplexes resulting from re-annealing as the PCR amplification approaches saturation. (e) BioAnalyzer profile of a high-quality sequencing library. A single 176 nt peak is present. (f) BioAnalyzer profile of a sequencing library with significant background from unextended RT primer. The background manifests as smaller DNA fragments that comprise 5–10% of the total DNA present in the sample; completely unextended RT primer yields a 144 bp PCR product. The DNA in this peak will produce sequencing data, but the sequence will consist of the linker sequence with no footprint.

**Table 1**

## Biotinylated rRNA depletion oligos

Reference	Start	End	Sequence
NR_003278.1	204	230	GGGGGGATGCGTGCATTATCAGATCA
NR_003278.1	286	320	TTGGTGACTCTAGATAACCTCGGGCCGATCGCACG
NR_003278.1	836	871	GAGCCGCCTGGATACCGCAGCTAGGAATAATGGAAT
NR_003279.1	183	216	TCGTGGGGGGCCCAAGTCTTCTGATCGAGGCC
NR_003287.1	919	950	GCACTCGCCGAATCCCGGGCCGAGGGAGCGA
NR_003279.1	921	948	GGGGCCGGGCCGCCCTCCACGGCGCG
NR_003287.1	1053	1080	GGGGCCGGGCCACCCCTCCACGGCGCG
NR_003279.1	1012	1052	CCCAGTGCGCCCCGGGCGTCGTCGCGCCGTCGGGTCCGGG
NR_003279.1	1257	1289	TCCGCCGAGGGCGCACCAACGGCCCGTCTCGCC
NR_003279.1	3754	3781	AGGGGCTCTCGCTTCTGGCGCCAAGCGT
NR_003279.1	4395	4429	GAGCCTCGGTTGGCCCCGATAGCCGGGTCCCCGT
NR_003287.1	4711	4745	GAGCCTCGGTTGGCCTCGGATAGCCGGTCCCCGC
NR_003287.1	4990	5024	TCGCTGCGATCTATTGAAAGTCAGCCCTCGACACA
NR_003280.1	125	157	TCCTCCGGGGCTACGCCTGTCTGAGCGTCGCT

**Table 2**

## Indexed library PCR primers

Forward Index	Indexed reverse library PCR primer
ACGACT	CAAGCAGAAGACGGCATAACGAGAT <u>AGTCG</u> TGTGACTGGAGTTCAGACGTGTGCTCTTCCG
ATCAGT	CAAGCAGAAGACGGCATAACGAGAT <u>ACTGAT</u> TGTGACTGGAGTTCAGACGTGTGCTCTTCCG
CAGCAT	CAAGCAGAAGACGGCATAACGAGAT <u>ATGCTGG</u> TGTGACTGGAGTTCAGACGTGTGCTCTTCCG
CGACGT	CAAGCAGAAGACGGCATAACGAGAT <u>ACGTCGG</u> TGTGACTGGAGTTCAGACGTGTGCTCTTCCG
GCAGCT	CAAGCAGAAGACGGCATAACGAGAT <u>AGCTGCG</u> TGTGACTGGAGTTCAGACGTGTGCTCTTCCG
TACGAT	CAAGCAGAAGACGGCATAACGAGAT <u>ATCGTAG</u> TGTGACTGGAGTTCAGACGTGTGCTCTTCCG
CTGACG	CAAGCAGAAGACGGCATAACGAGAT <u>CGTCAGG</u> TGTGACTGGAGTTCAGACGTGTGCTCTTCCG
GCTACG	CAAGCAGAAGACGGCATAACGAGAT <u>CGTAGCG</u> TGTGACTGGAGTTCAGACGTGTGCTCTTCCG