

Gene expression

***Ribo-ODDR*: oligo design pipeline for experiment-specific rRNA depletion in Ribo-seq**

Ferhat Alkan [†], Joana Silva[†], Eric Pintó Barberà and William J. Faller ^{*}

Division of Oncogenomics, Netherlands Cancer Institute, 1066 CX Amsterdam, The Netherlands

*To whom correspondence should be addressed.

[†] These authors contributed equally to this work.

Associate Editor: Yann Ponty

Received on August 28, 2020; revised on March 1, 2021; editorial decision on March 9, 2021; accepted on March 10, 2021

Abstract

Motivation: Ribosome Profiling (Ribo-seq) has revolutionized the study of RNA translation by providing information on ribosome positions across all translated RNAs with nucleotide-resolution. Yet several technical limitations restrict the sequencing depth of such experiments, the most common of which is the overabundance of rRNA fragments. Various strategies can be employed to tackle this issue, including the use of commercial rRNA depletion kits. However, as they are designed for more standardized RNAseq experiments, they may perform suboptimally in Ribo-seq. In order to overcome this, it is possible to use custom biotinylated oligos complementary to the most abundant rRNA fragments, however currently no computational framework exists to aid the design of optimal oligos.

Results: Here, we first show that a major confounding issue is that the rRNA fragments generated via Ribo-seq vary significantly with differing experimental conditions, suggesting that a ‘one-size-fits-all’ approach may be inefficient. Therefore we developed *Ribo-ODDR*, an oligo design pipeline integrated with a user-friendly interface that assists in oligo selection for efficient experiment-specific rRNA depletion. *Ribo-ODDR* uses preliminary data to identify the most abundant rRNA fragments, and calculates the rRNA depletion efficiency of potential oligos. We experimentally show that *Ribo-ODDR* designed oligos outperform commercially available kits and lead to a significant increase in rRNA depletion in Ribo-seq.

Availability and implementation: *Ribo-ODDR* is freely accessible at <https://github.com/fallerlab/Ribo-ODDR>.

Contact: f.alkan@nki.nl or w.faller@nki.nl

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

1 Introduction

Since its development, Ribosome Profiling (also known as Ribo-seq) has revolutionized the study of RNA translation (Ingolia *et al.*, 2009). The technique allows the analysis of ribosomally associated mRNA at codon-level resolution, providing a snapshot of the mRNAs bound by ribosomes in the cell. Information on translation efficiencies, open reading frame (ORF) usage, translation start sites, ribosome pause sites, amino acid dependencies and translation elongation rates can be gleaned from the data generated [reviewed in Wang *et al.* (2019)]. Additionally, the level of ribosome binding to an mRNA is a much better predictor of protein levels than the quantity of mRNA that is present, underscoring the importance of this technique (Blevins *et al.*, 2019; Ingolia *et al.*, 2009).

The Ribo-seq protocol takes advantage of the fact that at any instant a ribosome covers a ~28 nucleotide fragment of mRNA. This fragment is protected from nuclease digestion as a result and is hence known as the ribosome protected fragment (RPF). Following

ribosome stalling with translation blockers (e.g. cycloheximide), isolation of a cell lysate and RNase treatment, a cDNA library can be made from the resulting RPFs, and sequenced. By selecting the correct fragment size, the abundance of ribosomes at every location on the transcriptome can be deduced.

Although this process has been somewhat standardized (McGlinchey and Ingolia, 2017), numerous problems remain in generating high quality data. The RNase enzyme used (Gerashchenko and Gladyshev, 2017), or length of digestion (Liu *et al.*, 2019) can significantly bias the resulting data. Additionally, it is a common problem that a high proportion of sequencing reads derive from rRNA sequences, despite the use of rRNA depletion strategies. Indeed, in most experiments rRNA make up the majority of all reads sequenced (McGlinchey and Ingolia, 2017), and more than 90% in some cases (Gerashchenko *et al.*, 2012).

At present, the most common rRNA depletion strategies include the use of commercial kits or custom-designed biotinylated oligos previously reported in the literature. Both approaches make use of

RNA oligos that are complementary to the rRNA, thus binding to their target and allowing its depletion with a simple fishing approach. In addition, some commercial kits, such as Ribo-zero and NEBNext, make use of nuclease digestion as part of the protocol, which have been shown to introduce downstream biases in generated data (Zinshteyn et al., 2020). Moreover, the use of duplex-specific nuclease (DSN) has also been reported. However, DSN is known to also deplete highly expressed genes (Chung et al., 2015).

Both commercial kits and custom oligos assume that the rRNA fragments present in a sample are consistent across experiments. Here, we show that this is not the case, and that the experimental conditions and the tissue being used both introduce variations in the abundance of rRNA fragments produced. This raises the possibility that a standard oligo pool could cause differential rRNA depletion across samples, with significant differences in the rRNA fragments produced. When not addressed, this may lead to failed experiments or cause low sequencing depths for some samples, introducing downstream biases for comparative data analyses due to unbalanced sequencing depth (Zhang et al., 2014).

There are a number of possible approaches that could be taken to circumvent this problem. For example, there may be previously published data that provides a list of oligos that is confirmed to be efficient for Ribo-seq performed in a specific tissue and organism, following a specific protocol.

The most reliable way to confront the problem of differential rRNA fragment production is to perform pilot experiments on identical or similar samples and design novel biotinylated oligos that targets the most abundant rRNA fragments within generated pilot data (Kraus et al., 2019; Zinshteyn et al., 2020). Unfortunately, this approach requires experimental effort and computational work, potentially with a few rounds of optimization. However, this could be avoided due to the increasing number of Ribo-seq datasets from diverse sources that are being published, which could also serve as pilot data for researchers. Using this data, the most abundant rRNA fragments can be identified, and oligos designed to deplete them.

With this study, we first provide evidence that commercial rRNA depletion kits perform suboptimally and rRNA fragments generated by nuclease treatment differ substantially under various experimental conditions. Furthermore, we show that the same variability exists in fragments generated from different organs, even when using identical protocols. To tackle this problem, we present *Ribo-ODDR*, a Ribo-seq focused Oligo Design pipeline for experiment-specific depletion of Ribosomal RNAs. This pipeline addresses and automates the above mentioned problems and allows the design or optimization of oligos with high rRNA depleting potential, based on preliminary or previously published data. At default settings, *Ribo-ODDR* oligos are designed as perfectly complementary to given rRNA sequences. However, when calculating depleting potentials, the contribution of near-complementary bindings are also considered. This approach also enables users to use the pipeline with a collection of alternative rRNA sequences. It is freely accessible via GitHub in order to help researchers improve the power of their Ribo-seq experiments through more efficient rRNA depletion, thus maximizing the information gained from Ribo-seq experiments.

2 Materials and methods

We first introduce the public datasets that are analyzed here. Then, we present *Ribo-ODDR*. Finally, we explain the protocol for the Ribo-seq experiments performed within this study.

2.1 Ribo-seq with commercial rRNA depletion kits

Suboptimal performance of commercial depletion kits in Ribo-seq is a known issue (Chung et al., 2015) and we provide evidence on this by analyzing two public datasets (Simsek et al., 2017; Zinshteyn et al., 2020), in addition to one experiment performed here using the RiboCop kit (Lexogen, catalog no. 037). We accessed the datasets through NCBI using the GSE147324 (Zinshteyn et al., 2020) and GSE96998 (Simsek et al., 2017) ids. For the former, adapter trimming was performed with given instructions (Zinshteyn et al.,

2020), and, for the latter, cutadapt tool (Martin, 2011) was used with the following parameters, `-action=trim -discard-untrimmed -m 18 -adapter CTGTAGGCACCATCAAT -O 15 -error-rate = 0.1`. The rRNA fragments were then identified through mapping the trimmed reads to human and mouse 28S, 18S, 5.8S and 5S rRNA sequences, using TopHat (Kim et al., 2013) with the following parameters, `-no-novel-juncs -no-novel-indels -no-coverage-search`. We also mapped them to protein-coding transcripts (gencode v34 and vM21) after cleaning rRNA fragments using the *SortMeRNA* tool (Kopylova et al., 2012), and calculated the rRNA percentages by dividing the number of rRNA-mapping reads by the total number of reads that maps to rRNAs or protein coding transcripts.

2.2 Organ-specific *in vivo* Ribo-seq dataset

To provide evidence on the necessity of experiment-specific rRNA depletion, we made use of a comprehensive public dataset that generated *in vivo* Ribo-seq data for multiple tissues in mouse (Gerashchenko et al., 2021). One should note that no rRNA depletion protocol was applied within included experiments, however, RNA digestion was performed differently for two groups of samples. This difference enables us to analyze not only the tissue specificity of rRNA fragments but also their technical dependency on the used experimental protocol. Dataset was accessed through NCBI with the GSE112223 id and raw reads were trimmed using the cutadapt tool (Martin, 2011) (same parameters as above but for adapter `AGATCGGAAGAGCACACGTCT`) before running the *Ribo-ODDR* pipeline (*design-mode*).

2.3 Other public Ribo-seq datasets

To highlight the effect of ribonuclease selection on produced rRNA fragments, we made use of an *in vivo* Ribo-seq dataset (GSE82220), where experiments were separately performed in mouse liver using three different RNase pools (T1, S7 and T1&S7 together) (Gerashchenko and Gladyshev, 2017). Additionally, to address the rRNA fragment differences across experiments with identical protocols performed in the same tissue, we made use of a *in vivo* Ribo-seq dataset (GSE105147) where Ribo-seq was performed in mouse liver tumors that have different oncogenic driver mutations, and in healthy liver samples from WT mouse (Xu et al., 2019). We accessed the raw sequencing data of both datasets using the given accession ids and trimmed the adapter sequences using the cutadapt tool (Martin, 2011) (same parameters as in previous subsection). Later, we identified the rRNA fragments with the same approach used for the first two datasets.

2.4 The *Ribo-ODDR* pipeline

The primary aim of the *Ribo-ODDR* pipeline is to aid the biotinylated oligo design process for the depletion of rRNA fragments in Ribo-seq experiments. This includes designing novel oligos based on pilot experimental data and cross-species optimization of pre-compiled oligo sets. The tool does not contain any information on rRNA sequences, enabling *Ribo-ODDR* to be applied to any organism as long as rRNA (or other depletion-intended RNA) sequences are provided by the user.

In the following sections, we first describe how *Ribo-ODDR* performs the cross-species optimization. Then, we explain the methodology for designing novel oligos based on the pilot experimental data. All oligo designs are reported in FASTA, CSV, BED and GFF3 file formats, which contain various relevant information on oligo designs. One should note that *Ribo-ODDR* does not provide a final optimal set of oligos to deplete rRNA fragments. Instead, it reports the depleting potential of all high potential oligos to the user. However, the '*Ribo-ODDR oligo-selector*' user interface, introduced below, can aid the oligo selection process.

2.4.1 Cross-species optimization mode

In this mode, *Ribo-ODDR* requires two inputs from the user. The first input is the rRNA sequences of the organism in which the Ribo-seq experiments are going to be performed. Second input is the

sequences of pre-compiled oligos which are designed to deplete the rRNAs of another organism. We refer to these oligos as source oligos. *Ribo-ODDR* uses the following steps to optimize the source oligos for given rRNA sequences. Using *RISearch2* RNA-RNA interaction prediction tool (Alkan *et al.*, 2017), *Ribo-ODDR* first identifies the potential target regions of the source oligos in given rRNA sequences. Later, for each source oligo, the interaction prediction with lowest hybridization energy is accepted as the intended depletion target of that oligo. New oligos, that will replace the source oligos, are then designed as perfectly complementary to their identified depletion target. Later, to reach equivalent coverage as the source oligo, each new oligo design is extended on both ends according to its dangling ends within the source oligo-target rRNA interaction. This is done by extending the new oligo design by one complementary base for each unpaired nucleotide on 5' and 3' ends of the oligo. Note that if depletion target regions of source oligos are identical in both organisms, *Ribo-ODDR* will report the source oligos as optimized oligos unchanged.

2.4.2 Novel oligo design mode

The aim of this mode is to compute the depleting potential of all novel oligos on the given pilot Ribo-seq data. In a simple use-case, it requires the user to provide the rRNA sequences, the length range of desired oligos and the pilot data in which rRNA fragments are abundant. The workflow of the *Ribo-ODDR* (*design mode*) is shown in Figure 1.

Identifying rRNA fragments. Several variations of the Ribo-seq protocol exist, and for most the generated sequencing data requires trimming of adapter sequences and/or cleaning of used size selection markers before alignment. *Ribo-ODDR* does not perform these pre-processings itself, therefore, requires the user to preprocess the sequencing data prior to *Ribo-ODDR* use. Under default settings, trimmed & cleaned reads, provided as input, are first aligned to rRNA sequences using TopHat (Kim *et al.*, 2013) with the following parameter settings, `-n 2 -no-novel-juncs -no-novel-indels -no-coverage-search -segment-length 25`. However, users can also perform this alignment and provide their rRNA alignment files as input.

Oligo-set generation and depleting potential computation. Next, based on the user-given oligo length range constraint, depletion oligos are generated in a position specific manner. Oligo designs correspond to fixed length regions within user-given rRNAs, an oligo sequence being perfect complementary to its region. Note that the final oligo-set spans all possible regions across all given rRNAs. Therefore, oligo designs overlap with each other but the depleting potential of each oligo is computed separately following a novel heuristic approach. For each pilot sample and oligo, *Ribo-ODDR* determines the sample-specific oligo depleting potential by computing the oligo-specific percentage of depletable rRNA fragments within all rRNA fragments of that sample. Depletable fragments correspond to the reads that are primary-aligned to the oligo-specific depletion regions within given rRNA sequences. *Ribo-ODDR* identifies these regions using *RISearch2* with the following parameter settings, `-s max([l × 2/3], 10) -e 0 -w 0`, l denoting the oligo length. This allows *Ribo-ODDR* to identify all rRNA regions that are

complementary to the oligo sequence for a stretch of at least 10 nucleotides. However, not all reads mapping to these regions are considered depletable. A rRNA fragment (read) is considered depletable only if it satisfies the following constraints. The rRNA fragment has to cover a minimum of 10 nucleotides or two thirds of the oligo length, whichever is higher, within the depletion region under consideration. Additionally, the rRNA fragment can have a maximum of 10 nucleotides or one third of the oligo length, whichever is lower, outside the depletion region to be considered as depletable. This novel approach, as a whole, allows us not only to include the contribution of suboptimal bindings to oligo depleting potential but also to handle the multiple copies of rRNA sequences, when provided by the user.

Filtering oligos based on depleting potential. For fast computation of oligo features, *Ribo-ODDR* filter outs some of the low potential oligos based on customizable thresholds. Under default settings, oligos with a depleting potential less than 0.05 in more than 75% of the provided pilot samples are discarded. However, these thresholds can be altered by the user.

Computation of other oligo features. In addition to sample-specific depleting potential of oligos in pilot samples, *Ribo-ODDR* reports a few other informative statistics on designed oligos, including *GC_content* and *target_rRNA_position*. For each oligo, an *overall_depletion_score* is also computed, that is the ratio of samples oligo has a depleting potential more than the user-given threshold, (0.05 by default). Additionally, for each oligo, *Ribo-ODDR* reports a *minimum_hybridization_energy* that is the free energy of the full perfect complementary binding to an rRNA fragment at 37°C computed by *RISearch2* (Alkan *et al.*, 2017). Using the *RNAfold* program (Lorenz *et al.*, 2011), self-folding of the oligo is also predicted. This is reported in three different features, predicted *structure*, the *MFE* as the free energy of the predicted structure, and the *base_pairing_percentage* within the given structure.

Off-target prediction for designed oligos. If protein-coding transcript sequences of the organism are provided by the user, *Ribo-ODDR* computes the off-targeting potential of oligos as well. Denoting the minimum binding free energy across all oligos as E_{min} and the minimum oligo length as l_{min} , oligo off-targets are predicted using *RISearch2* (Alkan *et al.*, 2017) with the following parameter settings, `-s l'_{min} -e E'_{min}` , where $l'_{min} = [0.75 \times l_{min}]$ and $E'_{min} = \min(-25, 0.5 \times E_{min})$. This allows the detection of potential off-target regions, where generated RPFs are susceptible to undesired depletion. The number of predicted off-targets are reported to the user as an oligo feature, however, the additional information on individual off-target predictions are outputted separately.

2.4.3 Selecting final oligos with *Ribo-ODDR* oligo-selector

To aid the final selection from all oligos outputted by the *Ribo-ODDR* *design mode*, we developed the *Ribo-ODDR* oligo-selector user interface using the R-shiny environment. With it, users can explore the features of the oligo designs, filter them according to different filters on reported features and add the desired ones to the selection list, which results in removing the highly similar oligos from the available oligo list. A snapshot from this interface is shown in Supplementary Figure S1.

2.5 Experimental details on Ribo-seq experiments

C57BL/6 female and male mice between 8 and 12 weeks of age were used for experiments. For Figure 6 experiments, *Lgr5Cre^{ERT2}RPL22.HA* and *VillinCre^{ERT2}RPL22.HA* mice were generated by crossing *Lgr5Cre^{ERT2}* (Barker *et al.*, 2007) and *VillinCre^{ERT2}* (el Marjou *et al.*, 2004) mice with the RiboTag mouse (Sanz *et al.*, 2009). Due to differences in recombination efficiency and total number of cells, for the former, recombination was induced by a single intraperitoneal injection of 120 mg/kg tamoxifen and samples were taken after 24 h and 48 h. Whereas, for the latter, two consecutive injections of 80 mg/kg tamoxifen were performed and samples were taken after 120 h. Due to availability of strains, for Figure 7, *Lgr5Cre^{ERT2}Rptor^{fl/fl}RPL22.HA* animals were used for experiments. Recombination was induced by a single injection of

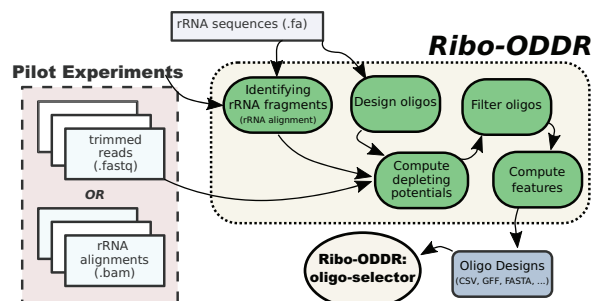


Fig. 1. The workflow diagram for the *Ribo-ODDR* pipeline (*novel oligo design mode*)

120 mg/kg tamoxifen and samples were taken after 24 h. Mice were bred in-house at the Netherlands Cancer Institute and all experimental protocols were approved by the NKI Animal Welfare Body.

2.5.1 Sample preparation from *in vivo* small intestines

Mice were euthanized by CO₂ and small intestines were immediately dissected, flushed with cold PBS supplemented with 100 µg/mL of cycloheximide and snap frozen using liquid nitrogen. Frozen tissues were ground by pestle and mortar while submerged in liquid nitrogen. The resulting powder was rapidly dissolved in cold lysis buffer (20 mM Tris HCl pH 7.4, 10 mM MgCl₂, 150 mM KCl, 1% NP-40, 100 µg/mL cycloheximide and 1x EDTA-free proteinase inhibitor cocktail (Roche, 04693132001)) and incubated on ice for 30 min. Samples were then homogenized using a Tissue Lyser (3 rounds of 45 sec at 50 oscillations per second) and centrifuged at max speed for 20 min at 4°C.

2.5.2 Sample preparation from *in vitro* crypt cultures

Crypt cultures were generated from the *VillinCre^{ERT2}RPL22.HA* mice as described previously (Faller et al., 2015). Around 120 plugs of 30 µL BME (Amsbio #3533-010-02) were used for each sample. Cells were incubated with 100 µg/mL cycloheximide for 3-5 min at 37°C, after which all steps were done on ice. Cells were collected and washed twice in cold PBS supplemented with 100 µg/mL cycloheximide, and homogenized with a 25 G needle in cold lysis buffer. After incubating the lysates on ice for 20 min, samples were centrifuged at max speed for 20 min at 4°C.

2.5.3 Ribosome profiling

Pull down of HA-tagged ribosomes. All supernatants were cleared for 20 min at 4°C, using Pierce™ Control Agarose Matrix (ThermoFisher #26150), after which they were incubated with pre-washed Anti-HA.11 Epitope Tag Affinity Matrix (BioLegend #900801) overnight at 4°C. Ribosomes were eluted in lysis buffer containing 200 µg/mL HA peptide (ThermoFisher #26184) and supplemented with 100 µg/mL cycloheximide for 15 min at 30°C. Exposed RNA was digested with RNase I (ThermoFisher #AM2294) for 40 min at 25°C and was stopped by adding SUPERASE (ThermoFisher #AM2694). RPFs were purified with miRNeasy minikit (Qiagen #217004) following the manufacturer's protocol and used for the library preparation.

Library preparation. The library preparation was conducted as previously described (Loayza-Puch et al., 2013) with some modifications. Briefly, RPFs were run in a 10% TBE-Urea polyacrylamide gel and size selected between 19 nt and 32 nt using marker RNAs. Gel slices were crushed, eluted and ethanol precipitated. Samples were then dephosphorylated in the 3' using T4 polynucleotide kinase (PNK) (NEB #M0201) and 1.5xMES buffer (150 mM MES-NaOH, 15 mM MgCl₂, 15 mM β-mercaptoethanol and 450 mM NaCl, pH 5.5) and incubated at 37°C for 4 h. RNAs were purified using Trizol and the 3' adapter was added using T4 RNA ligase I (NEB #M0204) at 24°C overnight. The ligated products were size selected and 5' phosphorylated with T4 PNK for 30 min at 37°C. After purifying the RNAs, the 5' adapter was added with T4 RNA ligase I for 2,5 h at 37°C and the final products were size selected on a 10% TBE-Urea polyacrylamide gel (see Supplementary Table S1 for adapter and primer sequences). This was followed by rRNA depletion, that is performed according to manufacturer's instructions when using the RiboZero (Illumina, catalog no. 20020598) and RiboCop (Lexogen, catalog no. 037) kits. For depletion with custom oligos, samples were incubated with 2 µL of the biotinylated oligo pool (10 µM each oligo, Supplementary Tables S2–S5) in 20 µL with 2xSSC (ThermoFisher #15557044). Samples were then denatured at 100°C for 1 min, followed by an incubation at 37°C for 15 min. In the meantime, 40 µL of MyOne Streptavidin C1 DynaBeads (ThermoFisher #65001) were washed and re-suspended in 20 µL of 2x wash/bind buffer (2 M NaCl, 1 mM EDTA, 5 mM Tris and 0.2% Triton X-100) and mixed with the sample at 1000 rpm for 30 min and at 37°C. Supernatants were collected and RNAs were precipitated with isopropanol and re-suspended in 8 µL of RNase-free water. Reverse transcription was performed with SuperScript III

(ThermoFisher #18080051) following the manufacturer's instructions and using the RTP primer. cDNA was then purified using G-50 columns (Merck GE27-5330-01) and used as a template for the PCR reaction with Phusion High-Fidelity DNA Polymerase (ThermoFisher #F530L) for 18 cycles, with primers listed in Supplementary Table S1. PCR products were purified using the QIAquick PCR purification kit (Qiagen #28104) followed by a E-Gel SizeSelect II 2%, (ThermoFisher #G661012). The quality and molarity of the samples were evaluated with the Agilent 2100 Bioanalyzer and the libraries were sequenced on the Illumina HiSeq2500.

Data processing. Raw reads are trimmed and cleaned from the size selection markers using the cutadapt tool (Martin, 2011), where trimming parameters are set to `-action=trim -discard-untrimmed -m 18 -adapter TGG AATTCTCGGGTGCCAAGG -error-rate=0.1`, whereas for cleaning `-g AGTGTACTCCGAAGAGGAC; anywhere -g GGCATTAACGCGAACTCGGCCTACAATAGTGA; anywhere -discard-trimmed -error-rate 0.1 -overlap 15` parameters are used. Then, *Ribo-ODDR* (design-mode) is used to align reads to mouse rRNA sequences (28S: NR_003279.1, 18S: NR_003278.3, 5.8S: NR_003280.2, 5S: NR_030686.1) and to design depletion oligos. Preprocessed reads are cleaned from rRNA fragments using the *SortMeRNA* tool (Kopylova et al., 2012) and remaining reads are mapped to genome release M21 protein-coding transcripts and GRCm38.p6 (mm10) mouse genome using TopHat (Kim et al., 2013) (`-no-novel-juncs -no-novel-indels -no-coverage-search`).

3 Results

3.1 Suboptimal rRNA depletion of commercial kits

Inefficient rRNA depletion is a known issue in Ribo-seq, and recently a comparative analysis of different rRNA depletion approaches has been published (Zinshteyn et al., 2020). This analysis included several commercially available kits (Ribo-Zero, Ribo-Zero Plus, RiboCop, NEBNext and QIAseq FastSelect), as well as a pool of biotinylated custom oligos (riboPool). Surprisingly, analysis of this data showed that despite rRNA depletion, there was still a high abundance of rRNA fragments in all samples. Of all reads that could be mapped to rRNA and protein coding transcripts, an average of 85% of them were rRNA fragments (see Fig. 2). These unexpectedly high percentages significantly reduce the resolution of the performed experiments as they decrease the sequencing depth in open reading frames (ORFs) and thus limit downstream analyses.

In order to further understand the inefficiencies of these kits, we visualized this data using the *svist4get* tool (Egorov et al., 2019). In line with the published analysis, we observed that rRNA depletion using commercial kits resulted in the incomplete depletion of 28S

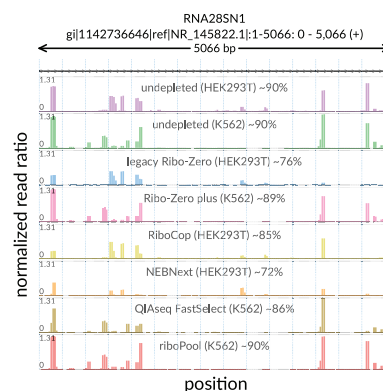


Fig. 2. Suboptimal performance of commercial rRNA depletion kits. Visualization is based on a previously published dataset (Zinshteyn et al., 2020) where several different kits were tested for rRNA depletion in human cell lines. Each track shows the positional abundance profile of 28S rRNA fragments coming from individual samples. For every position in the x-axis, y-axis represents the normalized read ratio, number of rRNA reads mapped to that position divided by the total number of reads mapped to all protein coding transcripts. Sample-specific total rRNA percentages are given in track labels together with sample identifiers

fragments, particularly those originating from several experiment-specific hotspots. Interestingly however, there was significant variability in the rRNA fragments present, suggesting that there was a depletion protocol driven heterogeneity in the rRNA fragments sequenced (Fig. 2). Analysis of an additional dataset generated using Ribo-Zero (Simsek *et al.*, 2017) also showed this heterogeneity, suggesting experiment-specific rRNA inconsistencies introduce variables that result in decreased protein coding sequencing depth (see Supplementary Fig. S2).

In samples that are difficult to work with, this may be a terminal issue for profiling experiments. Ribo-seq in intestinal epithelial cells, for example, is known to result in a very high level of rRNA sequencing reads (unpublished data from this lab and personal communication from others). We therefore carried out a Ribo-seq experiment using *in vitro* mouse intestinal organoids, with rRNA fragments depleted with the RiboCop kit. This experiment resulted in 89% rRNA reads (only 11% protein coding reads). In our analysis, we identified three hotspots (one for each 28S, 18S and 5-8S rRNA), where each hotspot individually accounted for more sequencing reads than all protein-coding reads (see Supplementary Fig. S3).

These observations demonstrate that commercial rRNA depletion kits perform suboptimally in Ribo-seq and suggest that custom-designed oligos would be a powerful way to increase sequencing depth in such experiments.

3.2 Tissue and RNase specificity of rRNA fragments

The use of custom-designed biotinylated oligos serves as a good alternative to overcome the inefficiency of commercial kits. However, there is no consensus on which oligos to use for maximal rRNA depletion, or even whether the same oligos are suitable for different experiments. Our results above would suggest that this is not the case.

In order to assess this, we measured the variability in rRNA fragment position and abundance in samples generated using different protocols and tissues of origin. We made use of a previously published dataset in which the authors performed *in vivo* Ribo-seq in nine different mouse organs without any rRNA depletion (Gerashchenko *et al.*, 2021). In this dataset, six sets of samples (brain, heart, kidney, liver, skeletal muscle and testis) were digested using a mix of RNaseT1 and RNaseS7, with the remaining 3 (lung, pancreas and spleen) being digested with only RNaseT1 as part of the Ribo-seq protocol. After identifying 28S, 18S, 5-8S and 5S rRNA fragments separately for each sample, we compared their rRNA fragment profiles (based on number of fragments mapped to each position in rRNAs) with a principal component analysis (PCA) (Fig. 3). This analysis revealed a striking heterogeneity in rRNA fragments in samples generated using different protocols, suggesting that rRNA depletion oligos that are efficient in one experiment may not be suitable for another. This protocol-derived heterogeneity of rRNA fragments can be clearly observed in Figure 3A, where the positional abundance profiles 28S rRNA fragments are shown for individual organs (one representative sample for each).

Moreover, the PCA and abundance profiles also reveal significant rRNA fragment differences in samples generated from different organs even when using the same protocol. While our analyses showed that there is a strong agreement between replicate measures of each organ in terms of rRNA fragments produced (Supplementary Figs S4–S12), we observed clear profile separation between organs (Fig. 3 and Supplementary Fig. S13).

A complementary analysis with another dataset (Gerashchenko and Gladyshev, 2017) showed that rRNA fragments still differ between experiments in the same tissue (liver), when experimental protocols are altered by the choice of RNase (see Supplementary Fig. S14). This is in agreement with our observations of rRNA fragment variability produced by different RNases. Furthermore, analysis of an additional dataset (Xu *et al.*, 2019) showed that, even when experimental protocols are identical, rRNA fragments can still vary between Ribo-seq in the same tissue, in this case in liver samples with different oncogenic driver mutations (see Supplementary Fig. S15).

All in all, this suggests that rRNA fragment heterogeneity is a common occurrence, and clearly shows that a ‘one size fits all’ approach is not appropriate for rRNA depletion.

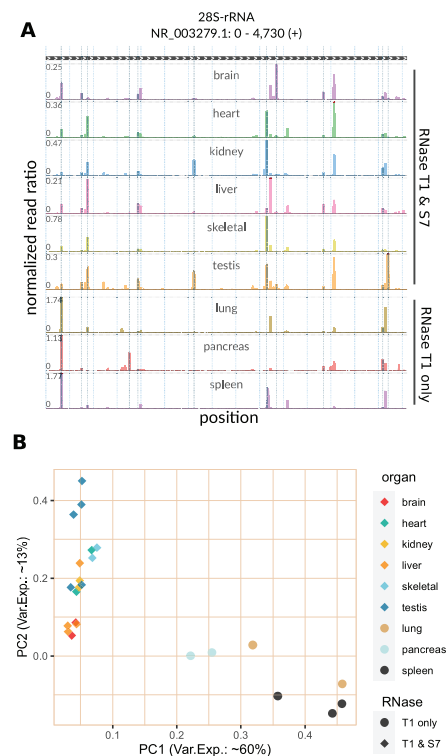


Fig. 3. Tissue and RNase specificity of rRNA fragments in the Ribo-seq data from (Gerashchenko *et al.*, 2021). (A) Each track shows the positional abundance profile of 28S rRNA fragments within the representative sample of the labeled organ. For every position in the x-axis, y-axis represents the normalized read ratio, number of rRNA reads mapped to that position divided by the total number of reads mapped to all protein coding transcripts. (B) Using the sample-specific abundance profiles on all rRNAs (28S, 18S, 5-8S and 5S), principal component analysis (PCA) was performed for all samples of the analyzed dataset (number of replicates varies for each organ). The first and second principal components are plotted against each other, PC1 in x-axis and PC2 in y-axis. The percentage of variance explained by each component is given in corresponding axis labels

3.3 Comparing the depleting potential of oligos across different experiments

In order to understand the effect that this rRNA fragment heterogeneity has on the efficiency of rRNA depletion oligos, we developed *Ribo-ODDR*. Based on given pilot Ribo-seq data, this pipeline measures the depleting potential of all possible oligos. For each oligo, this potential is simply equal to the percentage of rRNA fragments produced from the oligo target region on the rRNA, where the oligo binds with near-perfect complementarity.

We ran *Ribo-ODDR* on the organ-specific data used above and obtained the sample-specific depleting potentials of all 25 nt long oligos ($n = 6782$) that can deplete mouse rRNA fragments. For each individual oligo, the organ-specific depleting potential was calculated by simply averaging the values computed for each replicate of that organ.

In Figure 4, we compare the depleting potentials of oligos across all organ pairs with a cross-organ correlation analysis. These data makes it clear that the correlation in oligo depleting potential between samples treated using the same RNase digestion strategy is significantly higher than those using another strategy. In the RNaseT1-only digestion group, intra-group Pearson’s correlation coefficients are between 0.65 and 0.75 (mean value of 0.69), and for the RNaseT1/S7 group this is between 0.34 and 0.88 (mean value of 0.64), whereas inter-group coefficients are between 0.14 and 0.64 (mean value of 0.28). This confirms our observations detailed in Figure 3 that differing experimental conditions results in substantial differences in rRNA fragments created.

Furthermore, if the same RNase digestion protocol is used, oligos designed for one tissue (assuming only high potential oligos are selected) do not necessarily provide efficient depletion in another. In

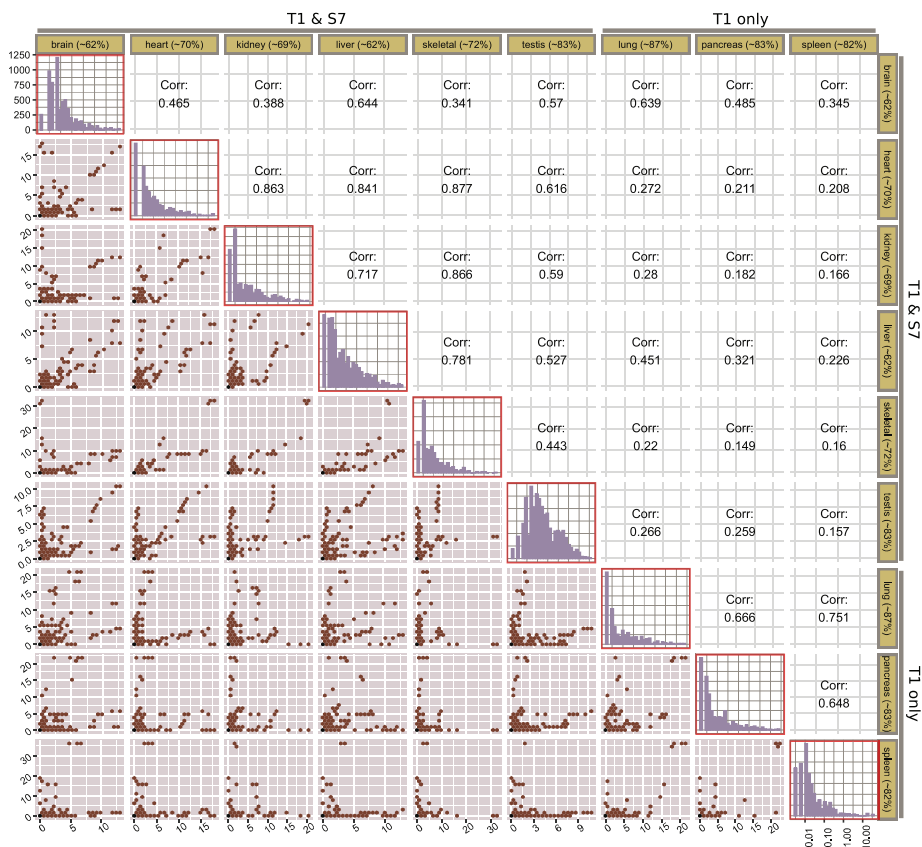


Fig. 4. Cross-organ correlation analysis of *Ribo-ODDR* computed oligo depleting potentials for all 25 nt oligos ($n = 6782$) targeting mouse 28S, 18S, 5-8S and 5S rRNA fragments. Each row and column corresponds to an organ. The diagonal plots (red boxes) show the histogram of oligo depleting potentials, computed for that organ based on the analyzed dataset. Axes of diagonal plots are shared, x-axis (shown in bottom-right corner) representing the log-transformed depleting potential and y-axis (shown in top-left corner) representing the number of oligos with that potential. Lower hex-binned scatter plots compares the depleting potential of all oligos between organ pairs (column versus row) with the Pearson's correlation coefficient given in their diagonal mirrors. In these plots, each bin contains one or more oligos with organ-specific rRNA depleting potential given in x- and y-axes for column and row organs, respectively. Percentages in row and column labels show the average rRNA percentage for that organ



Fig. 5. Oligo sets used in this paper. SET-1 consists of mouse-optimized version of eight human oligos and four additional ones, manually selected based on pilot *in vitro* experiments. SET-2 includes all SET-1 oligos and 5 new oligos, designed by *Ribo-ODDR* with data from pilot *in vivo* experiments using SET-1 oligos for rRNA depletion

some cases, the oligos with high depleting potential in one tissue show high depleting potential in others (kidney-versus-skeletal muscle, for example), however, this is only the case for a minority of tissues. Most tissue pairs show a low correlation in rRNA depleting potential of oligos. For example, pancreas-versus-heart and skeletal muscle-versus-lung both have Pearson's correlation coefficients of below 0.23, demonstrating that rRNA depletion oligos used successfully in one organ are unlikely to work in another. These observations are in agreement with our previous analysis of other publically available datasets (Fig. 2), and suggest that maximizing the information gained in Ribo-seq experiments may require experiment-to-experiment optimization.

3.4 Improving overall rRNA depletion efficiency using *Ribo-ODDR*, *in vivo* oligo design example

To test the power of the *Ribo-ODDR* design platform, we performed *in vivo* Ribo-seq experiments in the mouse intestine (see Fig. 5). We began by optimizing previously published human oligos (Ingolia et al., 2009; Loayza-Puch et al., 2013) to mouse ribosomal sequences which we named SET-1 oligos (see Supplementary Fig. S16 and Supplementary Tables S2 and S3). This experiment resulted in an average of only ~6% of sequencing reads mapped to protein coding regions, confirming the high levels of rRNA contamination found in intestinal epithelial samples. With this pilot data, we ran the *Ribo-ODDR* pipeline to design 5 additional oligos with high rRNA depleting potentials and added them to the oligo pool, creating SET-2 (see Supplementary Table S4). In Figure 6 and Supplementary Figure S17, we show that positional abundance profile of rRNA fragments are highly conserved between replicates in each experiment group, and newly designed oligos in SET-2 were successful at depleting the fragments in their corresponding regions. Crucially, rRNA depletion was far more efficient after the addition of five *Ribo-ODDR* designed oligos, resulting in a ~5-fold increase in protein-coding transcript reads (~28% versus ~6%), with SET-2 oligos giving ~72% rRNA fragments on average, compared to ~94% rRNA fragments on average for experiments using SET-1 oligos. This substantial increase in rRNA depletion efficiency demonstrates the power of experiment-specific rRNA depletion in Ribo-seq experiments and how using *Ribo-ODDR* can help this process.

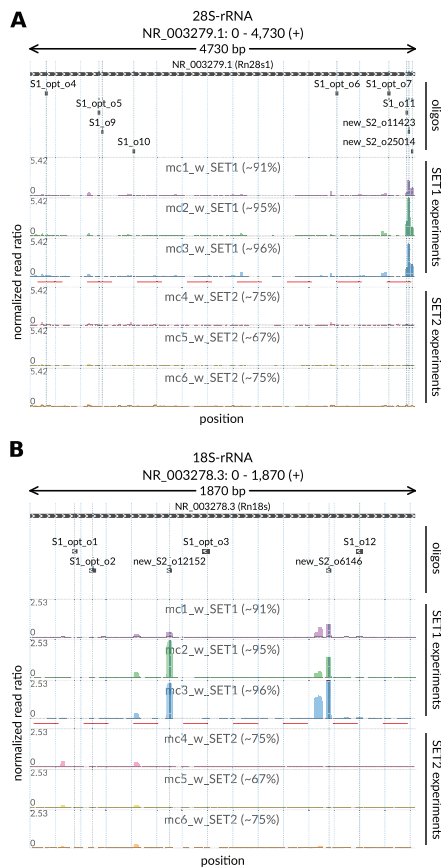


Fig. 6. Positional abundance profiles of 28S (A) and 18S (B) rRNA fragments coming from *in vivo* (mouse intestine) Ribo-seq experiments performed with two different sets of rRNA depletion oligos, SET-1 and SET-2. The latter set includes all oligos from the former and 5 additional oligos designed with *Ribo-ODDR*, based on pilot data generated using only SET-1 oligos. In each figure, top track indicates the target regions of used oligos within that rRNA, where additional oligos of the SET-2 are labeled as ‘new’. In all tracks, x-axis corresponds to position within rRNAs. In profile tracks, y-axis is fixed for all samples and shows the normalized read ratio, number of rRNA reads mapped to the position divided by the total number of reads mapped to all protein coding transcripts. The percentages given within sample labels indicates the sample-specific percentage of rRNA fragments, within all reads that is mapped to rRNAs and protein-coding transcripts

3.4.1 Evaluating potential off-target effects of custom oligos

A potential drawback of oligo-based depletion of rRNA is the possibility of complementarity to protein-coding fragments that can result in off-target depletion of mRNA. To ensure that this is not the case with *Ribo-ODDR* designed oligos, the tool reports the off-target potential of all oligos, allowing the selection of those with minimal complementarity to mRNAs. Indeed, in-depth read count analysis of potential off-target sites, shows that such depletion can be avoided. As shown in [Supplementary Figure S18](#), the average read count of potential off-target regions of the *Ribo-ODDR* designed SET-2 oligos does not change between experiments that use SET-1 or SET-2 oligos. This observation suggests that these 5 oligos, designed and selected with *Ribo-ODDR*, do not cause undesired depletion of informative off-target fragments.

3.5 *Ribo-ODDR* oligo-based depletion versus commercial kits

For comprehensive evaluation of rRNA depletion performances of commercial kits compared to *Ribo-ODDR* designed oligos, we performed six more *in vivo* Ribo-seq experiments in the mouse intestine, comparing RiboZero, RiboCop (two of the most commonly used commercially available kits) and *Ribo-ODDR* designed oligos. We added 5 additional *Ribo-ODDR*-designed oligos, creating the

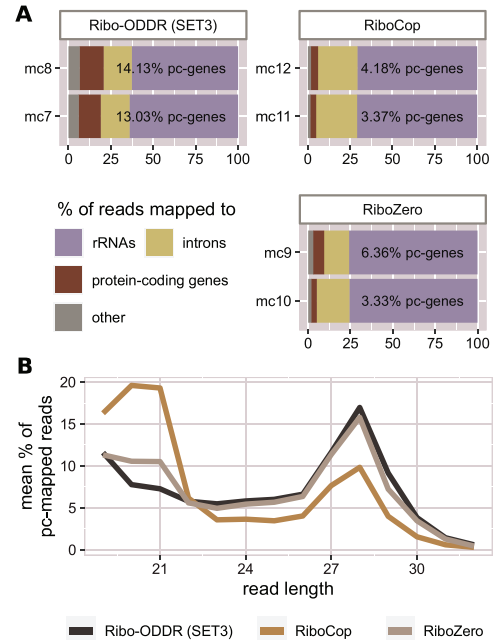


Fig. 7. Performance comparison of *Ribo-ODDR* oligos (SET-3) with RiboZero and RiboCop rRNA depletion kits. (A) Fractions of sequencing reads that are mapped to protein-coding genes, rRNAs, other non-coding transcripts and introns in 6 Ribo-seq experiments, performed *in vivo* in mouse intestine (2 replicates per group). (B) Read length histogram for protein-coding gene mapped reads in three experiment groups. Percentages are given as average of two replicates for each group

SET-3. In the resulting data we measured the percentage of reads that map to the protein coding, ribosomal, intronic and other RNAs, as well as the size of the RNA fragments produced.

Analysis of these Ribo-seq experiments showed that, as expected from intestinal tissue, RiboCop and RiboZero kits produced less than 5% protein coding mapped reads on average (see [Fig. 7](#)), severely limiting the sequencing resolution of this experiment. However, *Ribo-ODDR* oligos yielded ~14% protein-coding mapped reads, showing ~3 to ~4 times better performance than commercial kits. Further analysis of these data showed that the commercial kits differed in the quality of sequencing reads, with the RiboCop kit resulting in a higher number of short sequencing reads compared to RiboZero or *Ribo-ODDR* designed oligos. These experiments clearly show that experiment-specific custom oligos are superior to commercially available kits for rRNA depletion in Ribo-seq experiments, and that *Ribo-ODDR* provides a suitable tool for the design of such oligos. Furthermore, we have shown that in samples that have a low sequencing depth of protein coding RNAs, this increased rRNA depletion can turn a failed experiment into a successful one.

4 Discussion

Ribosome profiling has become a mainstay experiment in the analysis of RNA translation. It is one of the most informative techniques available for studying the transcriptome and has become very widely used in the decade since its development ([Ingolia et al., 2009](#)). However, as the technique focuses on ribosomally bound RNAs, the enrichment of rRNAs is an unfortunate necessity of the protocol. The nuclease cleavage of rRNAs produces fragments of a similar size to those being analyzed, creating an obvious technical challenge. Indeed, rRNA fragments commonly far outnumber reads from protein coding genes. As a result, rRNA depletion is a vital step in generating high quality Ribo-seq data.

The most common approach to overcome this issue is the use of commercially available rRNA depletion kits. However, our data shows that the efficiency of depletion using this method is variable,

and suggests that combining this method with a small number of custom designed oligos could significantly increase rRNA depletion. Additionally, previously published studies have suggested that the use of commercial kits can result in bias in individual mRNA fragments (Chung et al., 2015; Zinshteyn et al., 2020), emphasizing that the rRNA depletion strategy must be considered when planning experiments.

Using publically available data, we have also shown that this issue is compounded by variability in the specific rRNA fragments that are introduced by differing experimental conditions. Both the tissue origin, and the nuclease used for digestion significantly change the rRNA fragment population, showing that a depletion strategy that works for one experiment will not necessarily work for another. The source of the tissue-specificity of rRNA fragments is unknown, however it may be due to differences in the accessibility of the rRNA to the nuclease and/or to the presence of different intrinsic RNases. This in turn may result in differing abundances of rRNA fragments between tissues, saturating the binding capacity of the probes included in commercial kits in one tissue, and not in another. Ultimately, regardless of the cause of this variability, significant sequencing depth can be gained by improving the rRNA depletion. This may be particularly important in samples and tissues that have previously proven difficult to assay using Ribo-seq, such as the intestinal epithelium and other *in vivo* tissues. Furthermore, with the recent development of disome and trisome sequencing (Han et al., 2020; Meydan and Guydosh, 2020), the ability to improve the sequencing depth has the potential to significantly increase the information that can be gained by such experiments.

We developed *Ribo-ODDR* to aid with the design of custom oligos in an experiment by experiment manner. The tool enables users to run the *design mode* using preliminary or previously published data, allowing them to calculate the depleting potential of different oligo designs and select a small number with high depleting potential. We have shown that using such an approach can result in a 4-fold increase in the percentage of protein coding transcripts detected.

Thus far, manual inspection of Ribo-seq reads and their alignments to rRNA sequences has already been in practice for custom depletion oligo design. However, manual inspection cannot assess the suboptimal bindings of the oligos to determine their true potential. *Ribo-ODDR*'s novel depleting potential calculation approach addresses this by default. Additionally, this approach also enables users to add alternative rRNA sequences to the pipeline. Moreover, some oligo features, such as off-targeting potential, are often ignored when using the manual inspection approach. However, *Ribo-ODDR* integrates all the necessary design steps and enables users to assess the different features in order to select the most optimal oligos for highest rRNA depletion.

An obvious drawback of custom oligo design approach is the need for preliminary data to optimize the depletion strategy. Optimally, it is advisable to generate such preliminary data using the exact protocol as planned under experimental conditions, particularly when using tissues that have previously proven difficult to work with. However, as a result of the increasing number of Ribo-seq studies being published, in many cases it may be sufficient to use data from a similar source tissue that has been previously published. This could then be analyzed using *Ribo-ODDR* to create an oligo set that is likely to efficiently deplete rRNAs.

It is also important to point out that *Ribo-ODDR* is not necessarily a stand-alone method. We envision that *Ribo-ODDR* will be used alone in some cases, and in conjunction with other depletion strategies in others. For instance, our data suggests that commercial kits can benefit from the addition of a small number of custom designed oligos.

Ribo-ODDR provides a platform to assess the most optimal custom oligos, allowing for increased depth of mRNA fragment sequencing, and maximizing the information gained in Ribo-seq experiments.

5 Conclusion

In this study, we show that the use of commercial rRNA depletion kits may result in suboptimal depletion in Ribo-seq experiments,

and that different tissues and experimental conditions result in heterogeneity of produced rRNA fragments. Both of these findings demonstrate the necessity of experiment-specific custom oligo design for efficient rRNA depletion. To aid the computational part of the oligo design process, we have developed *Ribo-ODDR*. Oligos designed using this platform resulted in a substantial increase in rRNA depletion *in vivo* Ribo-seq experiments in mouse intestine, with much higher depletion performance when compared to commercial kits. Ultimately, this allows higher sequencing depth on the translome and more powerful downstream data analyses. The tool is easy to use, and will allow the optimization of this crucial step in the Ribo-seq protocol, particularly for samples that have proven difficult to assay. *Ribo-ODDR* is an open source software and freely accessible at <https://github.com/fallerlab/Ribo-ODDR>.

Acknowledgements

The authors thank past and present members of the Faller Lab and Abhijeet Pataskar for critical reading and fruitful discussions.

Funding

This work was funded by the Dutch Cancer Society (KWF Kankerbestrijding) Project [NKI-2016-10535]. JS was funded by an EMBO Long Term Fellowship [210-2018].

Conflict of Interest: none declared.

Data availability

All sequencing data generated for this paper has been deposited to NCBI and can be accessed using the PRJNA598159 or SRP239066 accession ids. Ribo-seq related QC plots of these data are provided in the Supplementary Document.

References

- Alkan, F. et al. (2017) Rlsearch2: suffix array-based large-scale prediction of RNA-RNA interactions and siRNA off-targets. *Nucleic Acids Res.*, **45**, e60.
- Barker, N. et al. (2007) Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature*, **449**, 1003–1007.
- Blevins, W.R. et al. (2019) Extensive post-transcriptional buffering of gene expression in the response to severe oxidative stress in baker's yeast. *Sci. Rep.*, **9**, 11005.
- Chung, B.Y. et al. (2015) The use of duplex-specific nuclease in ribosome profiling and a user-friendly software package for Ribo-seq data analysis. *RNA*, **21**, 1731–1745.
- Egorov, A.A. et al. (2019) svist4get: a simple visualization tool for genomic tracks from sequencing experiments. *BMC Bioinformatics*, **20**, 113.
- el Marjou, F. et al. (2004) Tissue-specific and inducible Cre-mediated recombination in the gut epithelium. *Genesis*, **39**, 186–193.
- Faller, W.J. et al. (2015) mTORC1-mediated translational elongation limits intestinal tumour initiation and growth. *Nature*, **517**, 497–500.
- Gerashchenko, M.V. et al. (2012) Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress. *Proc. Natl. Acad. Sci. USA*, **109**, 17394–17399.
- Gerashchenko, M.V. et al. (2021) Translation elongation rate varies among organs and decreases with age. *Nucleic Acids Res.*, **49**, e9.
- Gerashchenko, M.V. and Gladyshev, V.N. (2017) Ribonuclease selection for ribosome profiling. *Nucleic Acids Res.*, **45**, e6.
- Han, P. et al. (2020) Genome-wide survey of ribosome collision. *Cell Rep.*, **31**, 107610.
- Ingolia, N.T. et al. (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*, **324**, 218–223.
- Kim, D. et al. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.*, **14**, R36.
- Kopylova, E. et al. (2012) SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*, **28**, 3211–3217.
- Kraus, A.J. et al. (2019) Efficient and specific oligo-based depletion of rRNA. *Sci. Rep.*, **9**, 12281.
- Liu, B. et al. (2019) Optimization of ribosome profiling using low-input brain tissue from fragile X syndrome model mice. *Nucleic Acids Res.*, **47**, e25.

- Loayza-Puch,F. *et al.* (2013) p53 induces transcriptional and translational programs to suppress cell proliferation and growth. *Genome Biol.*, **14**, R32.
- Lorenz,R. *et al.* (2011) ViennaRNA Package 2.0. *Algorithms Mol. Biol.*, **6**, 26.
- Martin,M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet-journal*, **17**, 10–12.
- McGlinchy,N.J. and Ingolia,N.T. (2017) Transcriptome-wide measurement of translation by ribosome profiling. *Methods*, **126**, 112–129.
- Meydan,S. and Guydosh,N.R. (2020) Disome and Trisome Profiling Reveal Genome-wide Targets of Ribosome Quality Control. *Mol Cell*, **79**, 588–602.
- Sanz,E. *et al.* (2009) Cell-type-specific isolation of ribosome-associated mRNA from complex tissues. *Proc. Natl. Acad. Sci. USA*, **106**, 13939–13944.
- Simsek,D. *et al.* (2017) The mammalian ribo-interactome reveals ribosome functional diversity and heterogeneity. *Cell*, **169**, 1051–1065.
- Wang,Y. *et al.* (2019) Recent advances in ribosome profiling for deciphering translational regulation. *Methods*, **176**, 46–54.
- Xu,Y. *et al.* (2019) Translation control of the immune checkpoint in cancer and its therapeutic targeting. *Nat. Med.*, **25**, 301–311.
- Zhang,Z.H. *et al.* (2014) A comparative study of techniques for differential expression analysis on RNA-Seq data. *PLoS One*, **9**, e103207.
- Zinshteyn,B. *et al.* (2020) Nuclease-mediated depletion biases in ribosome footprint profiling libraries. *RNA*, **26**, 1481–1488.