The use of duplex-specific nuclease in ribosome profiling and a user-friendly software package for Ribo-seq data analysis

BETTY Y. CHUNG,[1,3] THOMAS J. HARDCASTLE,[1,3] JOSHUA D. JONES,[2,3] NEREA IRIGOYEN,[2,3]
ANDREW E. FIRTH,[2] DAVID C. BAULCOMBE,[1] and IAN BRIERLEY[2]

[1]Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, United Kingdom
[2]Division of Virology, Department of Pathology, University of Cambridge, Cambridge CB2 1QP, United Kingdom

ABSTRACT

Ribosome profiling is a technique that permits genome-wide, quantitative analysis of translation and has found broad application in recent years. Here we describe a modified profiling protocol and software package designed to benefit more broadly the translation community in terms of simplicity and utility. The protocol, applicable to diverse organisms, including organelles, is based largely on previously published profiling methodologies, but uses duplex-specific nuclease (DSN) as a convenient, species-independent way to reduce rRNA contamination. We show that DSN-based depletion compares favorably with other commonly used rRNA depletion strategies and introduces little bias. The profiling protocol typically produces high levels of triplet periodicity, facilitating the detection of coding sequences, including upstream, downstream, and overlapping open reading frames (ORFs) and an alternative ribosome conformation evident during termination of protein synthesis. In addition, we provide a software package that presents a set of methods for parsing ribosomal profiling data from multiple samples, aligning reads to coding sequences, inferring alternative ORFs, and plotting average and transcript-specific aspects of the data. Methods are also provided for extracting the data in a form suitable for differential analysis of translation and translational efficiency.

Keywords: ribosome profiling; duplex-specific nuclease; *Chlamydomonas*; mouse; translation

INTRODUCTION

Ribosome profiling measures at the codon level the extent to which individual mRNAs species of the transcriptome are engaged in protein synthesis. Initially developed by Ingolia et al. (2009), the method takes advantage of the knowledge that the position of a translating ribosome can be precisely determined by mapping the discrete, ~30 nucleotide (nt) fragments protected by the ribosome from nuclease digestion (Wolin and Walter 1988). Ingolia et al. (2009) exploited advances in deep-sequencing technology to globally analyze ribosome-protected fragments (RPFs), generating high-resolution views of the location of translating ribosomes on the transcriptome at any one time (Ingolia 2010, 2014; Ingolia et al. 2009, 2011, 2012, 2013). Profiling has proven to be increasingly valuable in studies of the translation process, for example, in the discovery of novel open reading frames (ORFs), the determination of elongation rates, the identification of sites of ribosome pausing and in the study of protein folding (for review, see Morris 2009; Weiss and Atkins 2011; Michel and Baranov 2013; Ingolia 2014; Jackson and Standart 2015). It also has broad application in the analysis of global gene expression and has been exploited in studies of infectious diseases (Stern-Ginossar et al. 2012, 2015; Liu et al. 2013; Arias et al. 2014; Caro et al. 2014; Jensen et al. 2014; Muzzey et al. 2014; Vasquez et al. 2014; Yang et al. 2015), cell growth, differentiation and development (Brar et al. 2012; Huang et al. 2013; Lee et al. 2013; Stadler and Fire, 2013; Stumpf et al. 2013; Subramaniam et al. 2013; Baudin-Baillieu et al. 2014; Brubaker et al. 2014; Duncan and Mata 2014; Gonzalez et al. 2014; Hendriks et al. 2014; Katz et al. 2014; Kronja et al. 2014; Schrader et al. 2014; Vaidyanathan et al. 2014; de Klerk et al. 2015), apoptosis (Wiita et al. 2013), mitochondrial gene expression and disease (Rooijers et al. 2013; Williams et al. 2014), cell stress (Gerashenko et al. 2012; Labunskyy et al. 2014; Zid and O'Shea 2014; Sidrauski et al. 2015), cell toxicity (Haft et al. 2014), and cell evolution (Artieri and Fraser 2014; McManus et al. 2014).

In the ribosome profiling methodology (Fig. 1), often referred to as Ribo-seq, cells are lysed under conditions optimized to minimize further ribosome movement (addition

[3]**Joint first authors.**

Corresponding authors: bcy23@cam.ac.uk, ib103@cam.ac.uk

Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.052548.115. Freely available online through the *RNA* Open Access option.

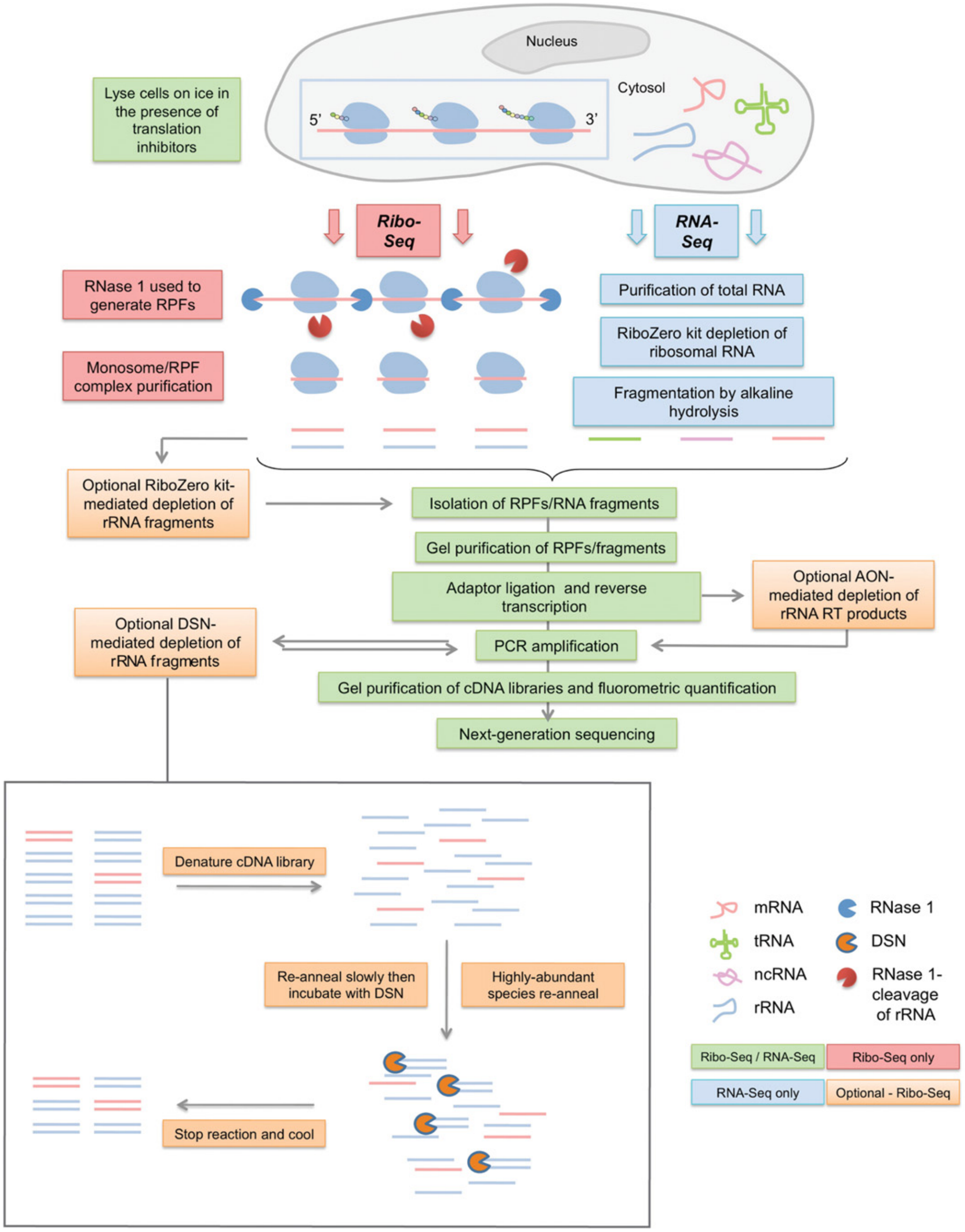
Nucleus
Cytosol
Lyse cells on ice in the presence of translation inhibitors
5'
3'
Ribo-Seq
RNA-Seq
RNase 1 used to generate RPFs
Monosome/RPF complex purification
Purification of total RNA
RiboZero kit depletion of ribosomal RNA
Fragmentation by alkaline hydrolysis
Optional RiboZero kit-mediated depletion of rRNA fragments
Isolation of RPFs/RNA fragments
Gel purification of RPFs/fragments
Adaptor ligation and reverse transcription
Optional AON-mediated depletion of rRNA RT products
Optional DSN-mediated depletion of rRNA fragments
PCR amplification
Gel purification of cDNA libraries and fluorometric quantification
Next-generation sequencing
Denature cDNA library
Re-anneal slowly then incubate with DSN
Highly-abundant species re-anneal
Stop reaction and cool
mRNA
tRNA
ncRNA
rRNA
RNase 1
DSN
RNase 1-cleavage of rRNA
Ribo-Seq / RNA-Seq
Ribo-Seq only
RNA-Seq only
Optional - Ribo-Seq

FIGURE 1. Ribosomal profiling strategy and points of rRNA depletion. The profiling methodology was based largely on that described by Ingolia and colleagues (Ingolia et al. 2009, 2012) except, following purification of ribosome-protected fragments (RPFs), library amplicons were constructed using a small RNA cloning strategy (Guo et al. 2010). Steps in the protocol specific to Ribo-seq, RNA-seq, or present in both are color-coded as indicated. Tested rRNA removal strategies are shown in orange boxes, with DSN treatment detailed separately at the *bottom* of the figure.

of translation inhibitors, rapid freezing), the lysate is treated with ribonuclease (often RNase 1) to degrade regions of mRNAs that are not physically protected, and the ribosomes harvested on sucrose gradients or through a sucrose cushion. The ribosome pellet is de-proteinized, the RPFs harvested by elution from a polyacrylamide gel, ligated to adapters, subjected to RT-PCR, deep sequenced and mapped back to the genome to reveal the location and abundance of ribosomes on mRNAs. The transcriptome itself is determined from the same lysate; total RNA is harvested, fragmented, cloned, and sequenced to generate an RNA-seq library.

Despite its increasing use, Ribo-seq is still in development (e.g., see Gerashcenko and Gladyshevet al. 2014; Gao et al. 2015) and there remain issues which limit its application. Not least among these is the problem of rRNA contamination, which can account for >90% of total reads (Gerashchenko et al. 2012). Hybridization-subtraction methods have been developed to reduce levels of major rRNA contaminants, but these are only partially effective, are time consuming, and can potentially introduce bias (Ingolia et al. 2011; Gerashchenko et al. 2012; Wickersheim and Blumenstiel 2013; Bazzini et al. 2014; Subtelny et al. 2014). Treatment of monosomes with EDTA, which dissociates the subunits, as well as minimizing the size range of RPFs sliced from polyacrylamide gels can also reduce rRNA contamination (Guo et al. 2010), but at the risk of losing important information that can be derived from analysis of a broader RPF size range (Rooijers et al. 2013; Lareau et al. 2014). Here we describe the use of duplex-specific nuclease (DSN) as a simple, species-independent way to achieve significant reductions in rRNA contamination. DSN, isolated from the hepatopancreas of the Kamchatka crab (*Paralithodes camtschaticus*), cleaves double-stranded DNA and RNA–DNA hybrid duplexes, with increased activity on perfectly matched duplexes (Shagin et al. 2002). DSN has been used in the normalization of cDNA libraries prior to next generation sequencing (Shagina et al. 2010) and the depletion of rRNA from RNA-seq libraries (Christodoulou et al. 2011; Yi et al. 2011; Matvienko et al. 2013; Miller et al. 2013). These experiments exploited the knowledge that the rate of DNA hybridization is proportional to the product of the concentration of the two separate DNA strands. Following denaturation of an RNA-seq cDNA (amplicon) library, the most abundant sequences re-anneal first and can be selectively degraded by addition of DSN (at 68°C), while less abundant sequences remain as ssDNA (Yi et al. 2011). To test the effectiveness of DSN in ribosome profiling, we generated Ribo-seq libraries from mouse tissue culture cells and from the green alga *Chlamydomonas reinhardtii*. We found that DSN reduced rRNA contamination substantially with only slight depletion of the most abundant mRNA RPF species, even within libraries of *C. reinhardtii*, whose transcriptome is highly GC-rich (Merchant et al. 2007).

Another limitation of Ribo-seq is in data analysis, which requires considerable expertise in bioinformatics. Programs that allow non-specialists to easily interpret Ribo-seq data sets have only recently become available (Crappé et al. 2015; Legendre et al. 2015) and many analyses are not yet supported in published packages. Here we describe an R package riboSeqR (released under Bioconductor, 2014) that provides a set of methods for parsing ribosomal profiling data from multiple samples, aligning to coding sequences, inferring alternative reading frames, and plotting average and transcript-specific behavior of these data. A unique feature of RPFs when aligned to the transcriptome is that they reflect the triplet periodicity of the translation process, where, during elongation, the ribosome moves in steps of three nucleotides (i.e., one codon) at a time along the mRNA. By analyzing the phase of the triplet periodicity of aligned RPFs, it is possible to determine the reading frame of translation on an mRNA. This is particularly relevant for identifying short and/or non-AUG initiated ORFs and for characterization of translated ORFs which may overlap the "main" coding ORF, or be present downstream (Michel et al. 2012; Dunn et al. 2013; O'Connor et al. 2013). Thus, riboSeqR uses this feature to identify unannotated coding ORFs.

We tested and validated the package using data derived from the libraries described above. In addition to processing and displaying profiling data, the riboSeqR package also allowed us to visualize a variety of translational control events. The use of DSN and the riboSeqR package facilitates the application of ribosome profiling and will be of value to both old and new users of the technique.

RESULTS

Duplex-specific nuclease: a sequence-independent rRNA depletion strategy

Ribo-seq and RNA-seq libraries were prepared from mouse tissue culture cells and the green alga *Chlamydomonas reinhardtii* and sequenced on MiSeq or HiSeq 2000 platforms (Table 1). The protocol, detailed in Figure 1 and Materials and Methods, includes a smallRNA cloning step to allow inexpensive in-house adapter activation and is adapted to Illumina smallRNA v2 to facilitate multiplexing. DSN treatment was performed at the library amplicon stage (post-RT-PCR; see Fig. 1). For each library, either one or two cycles of denaturing, annealing, and DSN treatment were performed; for *Chlamydomonas*, each treatment was carried out for 25, 50, or 90 min; for the mouse library, the reaction was for 25 min only. As shown in Table 1 and Figure 2A, for *Chlamydomonas*, each treatment decreased the proportion of

TABLE 1. Ribo-seq reads

Sample	Trimmed reads	rRNA reads	% rRNA	mRNA reads	% mRNA
C. reinhardtii cells untreated	2,708,355	2,527,482	93.3	52,135	1.9
1× DSN-treated *C. reinhardtii* cells	1,885,868	1,704,536	90.4	66,101	3.5
2× DSN-treated *C. reinhardtii* cells	1,722,290	1,448,560	84.1	123,812	7.2
Murine 17 clone 1 cells untreated	5,348,486	4,756,953	88.9	351,343	6.6
1× DSN-treated murine 17 clone 1 cells	2,834,455	1,984,716	70.0	531,908	18.8
2× DSN-treated murine 17 clone 1 cells[a]	27,124,163	13,985,423	51.6	6,840,263	25.2

DSN treatment was for 25 min.
[a]Samples were sequenced by MiSeq or HiSeq.

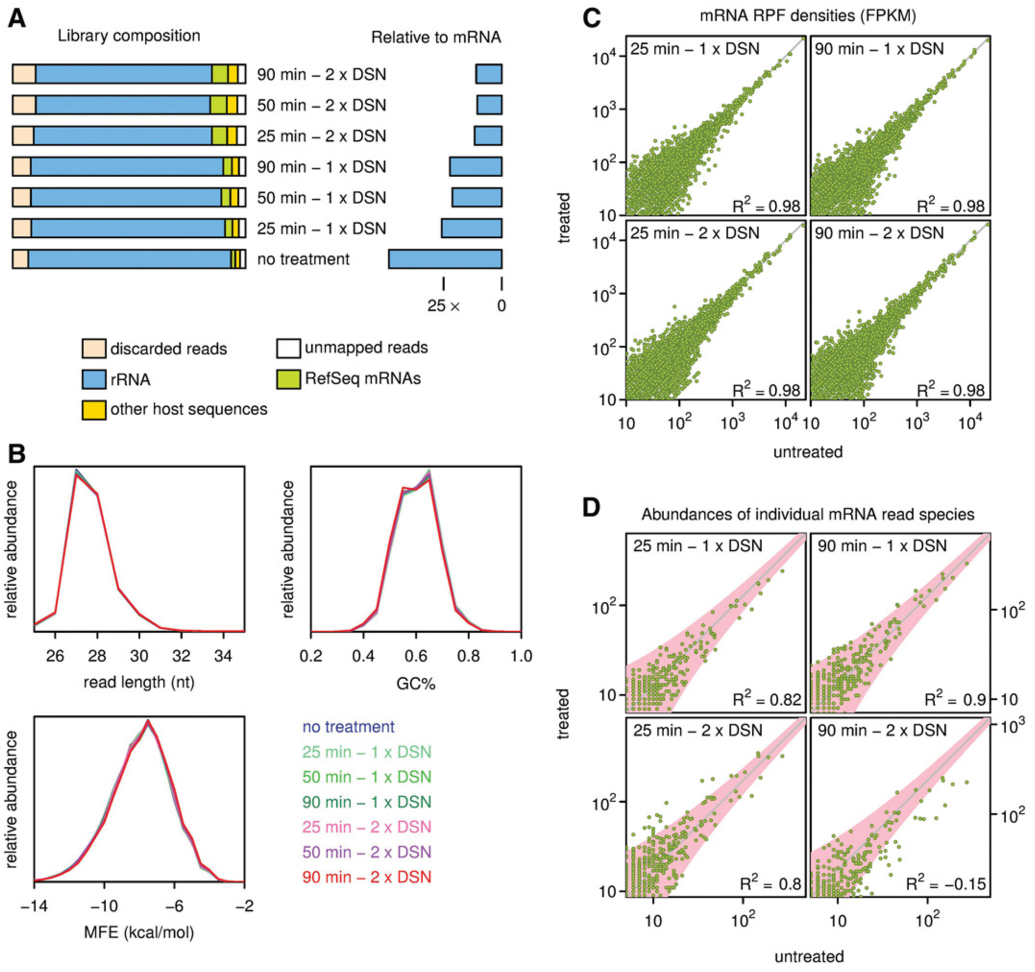

FIGURE 2. Analysis of the DSN-based rRNA depletion strategy using *Chlamydomonas* samples. (*A*) Relative rRNA depletion for different DSN treatments. Library composition is shown on the *left* and the amount of rRNA contamination relative to mRNA is on the *right* (see also Table 1). (*B*) Read abundance in DSN-treated and untreated libraries expressed as a function of read length, GC composition, and minimum free folding energy (MFE) for reads mapping to mRNAs. (*C*) RPF densities based on all RPFs mapping to NCBI RefSeq mRNAs for DSN-treated and untreated samples, expressed as fragments per kilobase per million mapped reads (FPKM). (*D*) Abundances of distinct mRNA-derived read species in DSN-treated and untreated samples. The gray guideline indicates the expected relationship if there is no depletion of mRNA—the slope is the ratio of the total number of mapped mRNA-derived reads in each sample. A theoretical 95% envelope based on χ^2 statistics is shown in pink. R^2 is calculated for distinct RPF species that have $\geq$5 occurrences in the untreated sample and is relative to the expected relationship indicated by the gray line (not a linear regression line), hence the potential for negative R^2 values.

rRNA substantially, increasing the proportion of mRNA in the sample by about fourfold after the two treatments. We did not see a noticeable effect of the time of incubation on the amount of rRNA depletion (Fig. 2A). In considering the use of DSN for depletion of rRNA from these libraries, we were mindful that post-hybridization nuclease treatment could potentially introduce biases arising from digestion of abundant mRNA-derived cDNAs. The possibility that DSN could also deplete annealed, GC-rich cDNAs, or single-stranded cDNA with a high propensity to form intramolecular structures was also considered. The highly GC-rich *Chlamydomonas* transcriptome (Merchant et al. 2007; Fig. 2B) was especially relevant in this regard, although the preferential activity of DSN on perfectly matched duplexes (Shagin et al. 2002) limited our concerns somewhat. An analysis of the physical profile of the reads is presented in Figure 2B, showing the length, GC content, and minimum free folding energy distributions of mRNA-derived RPFs for DSN-treated and untreated samples. The profiles of the different samples were found to be almost identical, indicating that DSN treatment introduces negligible bias with respect to these parameters for mRNA-derived RPFs. RPF densities on mRNA transcripts were found to closely follow a zero-intercept linear relationship when DSN-treated samples were compared with untreated samples, indicating that DSN did not noticeably deplete the most abundant mRNA transcripts ($R^2 = 0.98$; Fig. 2C). We also compared the abundances of individual mRNA read species between untreated and DSN-treated samples (Fig. 2D). Due to the lower counts involved, R^2 values were lower. However, there was relatively little depletion of abundant read species except for the 2 × 90 min treated sample.

To directly compare the efficacy of DSN with other rRNA depletion strategies, we prepared four new mouse libraries, two that had been subjected either to one or two cycles of DSN (25 min), and a further two in which rRNA removal was achieved using either a pool of antisense oligonucleotides (AON) (Ingolia et al. 2012) or a RiboZero kit. The step in library preparation where these depletion methods were used is indicated in Figure 1. Libraries were sequenced on the NextSeq platform and RiboSeq read counts are displayed in Table 2. Treatment with RiboZero produced the greatest increase in library mRNA fraction (18-fold), followed by 2× and 1× DSN (11- and ninefold, respectively), while AON depletion gave only a threefold increase in mRNA fraction (Fig. 3A; Table 2). Again, DSN did not introduce any bias relative to mRNA read length or minimum free folding energy (Fig. 3B), although for unknown reasons in this experiment, all treatments led to a very slight bias in GC% compared with the untreated sample (Fig. 3B). As before, DSN did not noticeably deplete mean RPF densities for the most abundant mRNAs ($R^2 = 0.97$–0.98; Fig. 3C). When we compared the abundances of individual mRNA read species, DSN was found to introduce more variability than RiboZero ($R^2 = 0.67$ for RiboZero; 0.59 and 0.45 for 1× and 2× DSN, respectively; Fig. 3D). Individual RPF species showing evidence of depletion (below the diagonal line in Fig. 3D; 1 × DSN panel) tended to have slightly higher GC content (56.2% GC) than those without depletion (above the diagonal in Fig. 3D; 1 × DSN panel) (51.7% GC). Thus, DSN depletion may lead to slight underestimates of ribosome density at a few specific sites (e.g., strong ribosome pause sites in highly expressed transcripts); however, for the vast majority of applications, this is unlikely to be problematic. RiboZero on the other hand was found to introduce more bias than DSN for mRNA-derived reads that have stronger binding potential to the RiboZero probe (Fig. 3B, lower right panel). Given the robustness and specificity of rRNA depletion by DSN in the context of Riboseq, we anticipate this sequence-independent approach will allow application of ribosome profiling to a wide array of organisms.

A user-friendly bioinformatic package for Ribo-seq processing; riboSeqR

We have developed the riboSeqR R package (available at the Bioconductor website: http://www.bioconductor.org/packages/riboSeqR and also implemented at http://ribogalaxy.ucc.ie/) to provide a set of methods for user-friendly analysis of ribosome profiling data. The package parses data aligned to a (potentially de novo) transcriptome, providing frame-calling and plotting functions. The package optionally identifies potential coding sequences based on the identification of start/stop codons within the sequence of FASTA files, with RPFs mapping in-frame to corresponding ORFs. Alternatively, known coding sequences can be used. The versatility of this package is illustrated here through analyses of the *Chlamydomonas* and mouse Ribo-seq data sets, with a combined size of RNA-seq and Ribo-seq alignment files, respectively, of 1.2 GB and 2.3 GB. Scripts used to perform these analyses are provided in Supplemental Figures S1 and S2, and the run-time for these analyses was ∼14 min for *Chlamydomonas* and 31 min for mouse data, on a single 2.50 GHz processor with 16 GB of RAM.

We began the riboSeqR analysis by examining read-length distributions. For *C. reinhardtii*, which possesses chloroplastic, mitochondrial, and cytoplasmic ribosomes similar to plant ribosomes (Manuell et al. 2007), RNase 1 treatment typically produced cytoplasmic RPFs with a length size distribution sharply peaked at 27–28 nt (Fig. 4A). For the 27-nt size class, the 5′ ends of *C. reinhardtii* RPFs mapped overwhelmingly to the second nucleotide position of codons (Fig. 4B). For the second most abundant RPF size class, i.e., 28 nt, a large majority of 5′ ends mapped to the first nucleotide of codons, indicating, in this case, the addition of one nucleotide at the 5′ end of such RPFs relative to 27-nt RPFs (Fig. 4B). riboSeqR uses this initial analysis to filter the data, considering for further analysis those RefSeq annotated coding sequences that contain at least fifty 27-nt reads, mapping to at least 10 distinct locations within the coding region. Note that for highly translated coding regions, the small proportion of out-of-phase reads may cause overlapping but out-of-phase putative coding regions to pass this filtering step. riboSeqR thus further filters those putative coding regions by identifying those cases where the phase with the

TABLE 2. Ribo-seq reads

Sample	Trimmed reads	rRNA reads	% rRNA	mRNA reads	% mRNA
Murine 17 clone 1 cells untreated	25,961,533	23,436,379	90.3	789,945	3.0
Murine 17 clone 1 cells 1× DSN	18,810,325	10,955,703	58.2	5,117,415	27.2
Murine 17 clone 1 cells 2× DSN	11,866,223	6,040,718	50.9	3,935,374	33.2
Murine 17 clone 1 cells AON depletion	28,574,393	23,531,543	82.4	2,367,714	8.3
Murine 17 clone 1 cells RiboZero	32,125,425	6,792,302	21.1	17,039,756	53.0

Samples were sequenced by NextSeq.

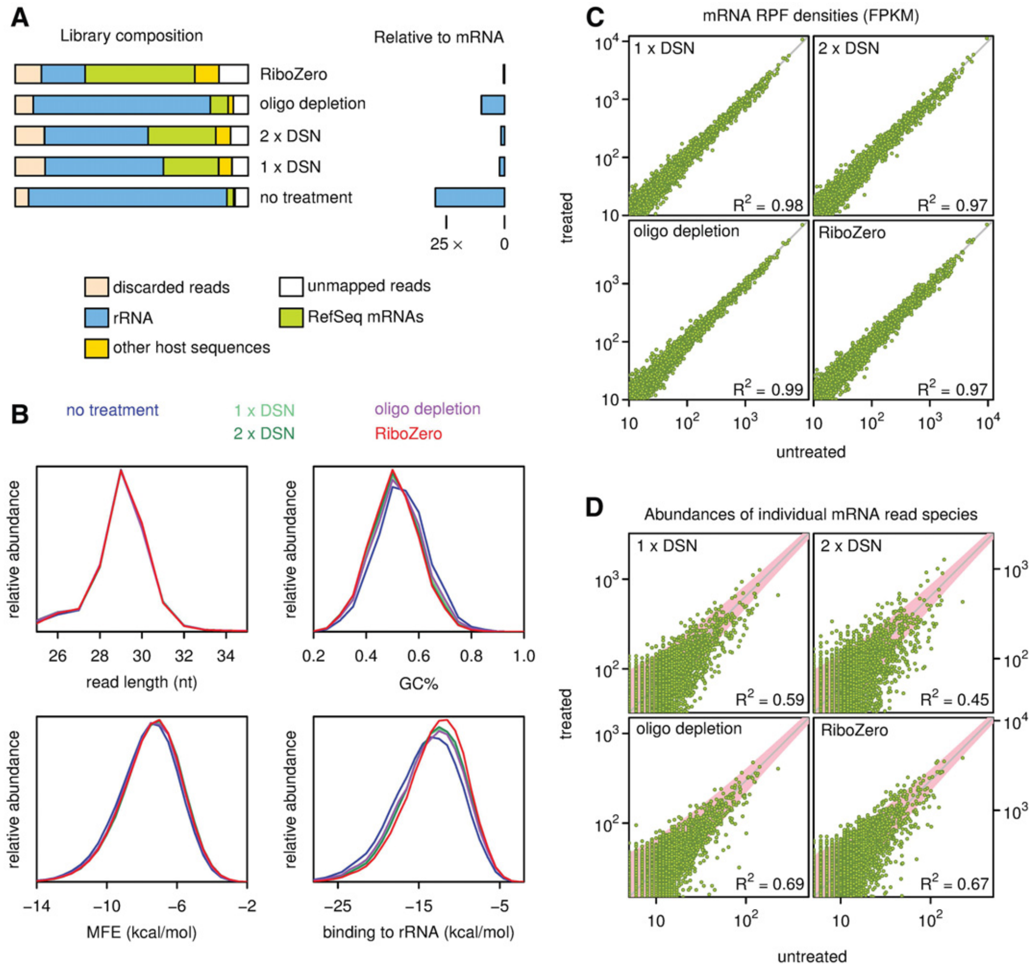

FIGURE 3. Comparison of rRNA depletion strategies using mouse samples. (*A*) Relative rRNA depletion for different depletion strategies. Library composition is shown on the *left* and the amount of rRNA contamination relative to mRNA on the *right* (see also Table 2). (*B*) Read abundance in treated and untreated libraries expressed as a function of read length, GC composition, minimum free folding energy (MFE), and optimal binding energy to reverse-complemented rRNA (essentially the RiboZero probe) for reads mapping to mRNAs. (*C*) RPF densities based on all RPFs mapping to NCBI RefSeq mRNAs for treated and untreated samples, expressed as fragments per kilobase per million mapped reads (FPKM). (*D*) Abundances of distinct mRNA-derived read species in treated and untreated samples. The gray guideline indicates the expected relationship if there is no depletion of mRNA—the slope is the ratio of the total number of mapped mRNA-derived reads in each sample. A theoretical 95% envelope based on χ^2 statistics is shown in pink. R^2 is calculated for distinct RPF species that have $\geq$5 occurrences in the untreated sample, and is relative to the expected relationship indicated by the gray line.

maximum number of reads (the maximal phase) is not the expected phase for that putative coding region. If the ratio of reads in the expected phase to maximal phase does not significantly (χ^2 test with significance threshold of 0.05) exceed that ratio observed for all coding regions (Figs. 4B, 5B), the putative coding region is excluded from further analysis.

Using these selected coding sequences, riboSeqR constructs the weighted average number of n-nucleotide reads around the annotated coding start and stop sites (Fig. 4C). Contributions of reads from individual coding sequences are down-weighted by the total number of n-nucleotide reads per coding sequence length to avoid highly translated coding regions unduly influencing the profile. These plots further demonstrate the high level of triplet periodicity as a function of read size class. More importantly, they allow more detailed observations to be made concerning the behavior of the ribosome, especially at the sites of translation initiation and termination. During termination, the incorporation of release factors into the ribosomal pretermination complex induces a structural rearrangement that results in a footprint ∼1–2 nt larger relative to the footprint of the initiating or elongating ribosome (Wolin et al. 1988; Alkalaeva et al. 2006). This change is clearly apparent in the riboSeqR figures generated from our data sets. In Figure 4C, the majority of RPFs in interior regions of coding sequences (i.e., elongation-state RPFs) have a length of 27 nt, with the modest read peak corresponding to terminating ribosomes most likely reflecting ribosomes paused at the stop codon with an unoccupied A-site. Based on the positions of the maximum values near to the start and stop codons, we can infer that, for 27-nt

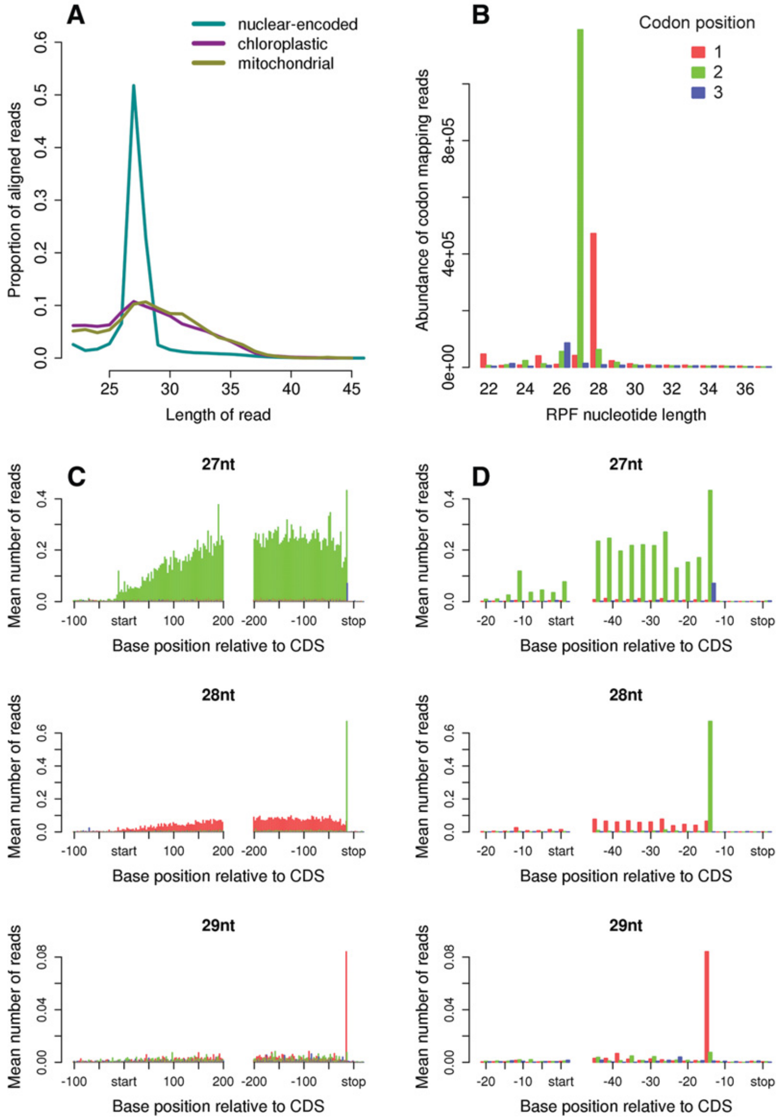

FIGURE 4. Ribosome profiling of *C. reinhardtii*. (*A*) Length distributions of RPFs mapping to the interior regions of nuclear-encoded, mitochondrial and chloroplastic coding ORFs. (*B*) Histogram of the codon positions (i.e., first [red], second [green] or third [blue] nucleotide of each $N_1N_2N_3$ codon) to which the 5′ ends of RPFs map, as a function of RPF size class, for RPFs mapping to the interior regions of nuclear-encoded coding ORFs. (*C*) Histograms of RPF 5′ end positions relative to start and stop codons, for 27-, 28-, and 29-nt RPFs mapping to nuclear-encoded mRNAs. Coloring indicates the codon positions of the 5′ ends of RPFs. (*D*) Enlarged view around the start and stop codons; "start" and "stop" indicates the first nucleotide of the start and stop codon positions.

RPFs, the ribosome protects 11 nt 5′ of the P-site codon and 10 nt 3′ of the A-site codon (e.g., Fig. 4D). Pausing during termination leads to a much greater density of 28-nt RPFs at stop codons compared with interior positions (Fig. 4C). The 28-nt RPFs that map to stop codons most likely derive from ribosomes that have bound release factor complex. Again, relative to the density in interior positions, a still higher termination peak is apparent for the 29-nt RPFs size class (Fig. 4C; enlarged view in Fig. 4D). Based on the position at which the 5′ ends of RPFs map, it is apparent that the 28-nt termination RPFs involve a single nucleotide addition at the 3′ end, the 28-nt elongation RPFs involve a single nucleotide addition at the 5′ end, and the 29-nt termination RPFs involve one nucleotide addition at each end, all relative to 27-nt elongation-state RPFs (Fig. 4C,D).

We next used the same riboSeqR processing steps to analyze data from mouse cells (Fig. 5). Murine cytoplasmic RPFs had a length distribution typically peaking at 28–30 nt (Fig. 5A), ∼2 nt longer than the *C. reinhardtii* cytoplasmic RPFs. There is a precedent for such a difference because the wheat germ ribosome has an mRNA footprint some 2–4 nt smaller than the rabbit reticulocyte ribosome (Wolin and Walter 1981). The length distribution of mouse mitochondrial RPFs was broader than that of cytoplasmic RPFs, and shifted to longer length classes (Fig. 5A). We did not observe the bimodal peak for mitochondrial RPFs seen in a previous study (Rooijers et al. 2013), but this could be a consequence of size selection at the gel-purification stage that was not tailored specifically for organelle RPFs. For the murine sample, the 5′ ends of 87.4% of all RPFs mapping to interior regions of annotated coding sequences of nuclear-encoded mRNAs mapped to the first nucleotide of codons (Fig. 5B,C). For the most abundant length size class (29 nt), 94.0% of RPF 5′ ends mapped to this codon position. Such RPFs contain 12 nt 5′ of the P-site codon and 11 nt 3′ of the A-site codon (Fig. 5D). At termination codons, there was a noticeable decrease in the number of 28-nt RPFs and a substantial increase in the number of 30-nt RPFs (Fig. 5C), again illustrating the larger footprint of terminating ribosomes relative to elongating ribosomes. Similarly to *C. reinhardtii*, this increased RPF size corresponded to addition of a nucleotide at the 3′ end of RPFs (Fig. 5C).

De novo inference of coding regions

Ribosome profiling allows the de novo annotation of coding regions within a transcriptome. We used riboSeqR to identify regions of the transcriptome beginning and ending with canonical start (AUG) and stop (UAG, UAA, UGA) codons in frame. We subsequently filtered these putative coding

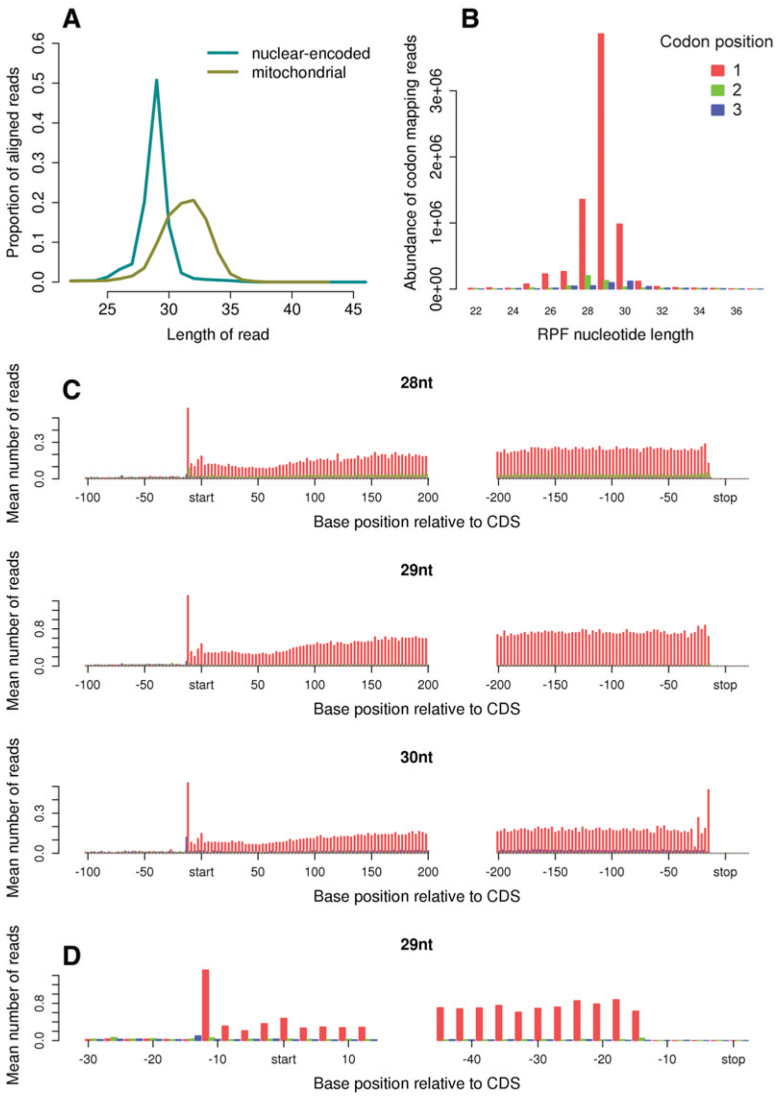

FIGURE 5. Ribosome profiling of mouse cells. (*A*) Length distributions of RPFs mapping to the interior regions of nuclear-encoded and mitochondrial coding ORFs. (*B*) Histogram of the codon positions to which the 5′ ends of RPFs map as a function of RPF length size class, as in Figure 4B. (*C*) Histograms of RPF 5′ end positions relative to start and stop codons, for 28-, 29-, and 30-nt RPFs mapping to nuclear-encoded mRNAs. Coloring indicates the codon positions of the 5′ ends of RPFs. (*D*) Enlarged view around the start and stop codons for the 29-nt size class. Deviations from the mean for peaks located 3–4 codons upstream (downstream) of the termination (initiation) site likely result from ligation and nuclease biases: RPFs whose 5′ ends align to these positions have constant nucleotides at or near to the 3′ (5′) end because of the conserved stop (start) codon; whereas, at other positions in the histogram, any ligation or nuclease biases deriving from the identity of the nucleotides at the termini of the RPFs are averaged out when averaging over different mRNA species.

regions using the same criteria as for RefSeq annotated sequences as described above. For *Chlamydomonas*, such de novo construction of coding sequences finds 97.5% of the known coding sequences that pass the same filtering criteria (50 reads mapping, at least 10 unique hits) and of these, 96.7% are exact matches to those in RefSeq. These hits correspond to 25.5% of the total coding sequences recorded for *Chlamydomonas* in RefSeq, but note that not all sequences are being translated in our single, wild-type sample. For mouse, the equivalent figures are 98.1%, 99.1%, and 46.9%, respectively.

Visualization of uORFs and overlapping ORFs using riboSeqR

Ribosome profiling permits the identification of short, translated ORFs upstream of "main" coding sequence (uORFs), some of which utilize near-cognate, non-AUG initiation codons (e.g., CUG) (Ingolia et al. 2011). It can also facilitate the discovery of short ORFs that lack a specific initiation codon but are instead accessed via noncanonical translation mechanisms (Michel et al. 2012; Gerashchenko et al. 2012). RPFs immediately after a canonical stop codon may derive from stop codon read-through (Dunn et al. 2013) or ribosomal frameshifting at or near the stop codon. Figure 6 shows a number of validated examples of such translation events. In each case, riboSeqR was used to identify and display the relevant reads from our data files following input of the chosen accession number. In each panel, RNA-seq reads for a particular gene are shown in gray with Ribo-seq reads superimposed in three colors, representing RPFs whose 5′ ends map to each of the three possible codon positions. Above each panel, the annotated NCBI Reference Sequence Database (RefSeq) ORF is displayed as a turquoise box. Colored lines above the panels show the coding sequences inferred with riboSeqR. We then identified cases in which the filtered putative coding sequence did not correspond to the annotated coding sequence. These cases may represent either misannotation of the transcriptome or transcriptomic sequence or more interestingly, alternative transcription/translation events.

Figure 6A shows mouse initiation factor eIF4G2, the translation of which is initiated from a GUG codon (Takahashi et al. 2005). This is clearly evident in the plot, with abundant reads at the GUG followed by a continuum of reads in this frame up to the stop codon. We also noticed two other highly utilized AUG codons in the 5′ leader of eIF4G2 that would initiate translation of short uORFs (with 16 and seven codons, respectively) that could be regulatory. Indeed, the major peak in the eIF4G2 plot mapped to the most 5′ uORF (in green). The noticeable spike in red, however, does not correspond to a start codon (canonical or otherwise), but instead represents a strong termination peak from an overlapping ORF with generally low read counts. Another example of uORF identification is shown in Figure 6B. The 5′ leader of mouse ATF4 (Harding et al. 2000) harbors two uORFs, one very short (three codons) and a longer second (59 codons) uORF overlapping the main ATF4 coding sequence. As can be seen, the majority of reads mapped to the short uORF, consistent with its important regulatory role (Vattem and Wek 2004; Ait Ghezala et al. 2012), and translation of uORF2 and ATF4 was also detectable. Unfortunately, we could not illustrate ribosomal frameshifting or stop codon read-through (see Brierley et al. 2010; Dunn et al. 2013) because read coverage of the relevant mRNAs was too low. However, the transcript for mouse selenoprotein T (SelT) illustrates a specialized form of stop codon read-through (Heider et al. 1992; Kryukov et al. 2003). The efficiency of selenocysteine insertion is very high (Heider et al. 1992; Berry et al. 1993) and, consistent with this, there were similar levels of ribosome footprints in the coding sequences flanking the in-frame stop codon (Fig. 6C). With high levels of framing, the riboSeqR package also allows visualization of likely internal overlapping genes, for example, within the mouse complement component (3b/4b) receptor 1-like protein (Cr1l, also known as mCRY; Paul et al. 1989) (Fig. 6D).

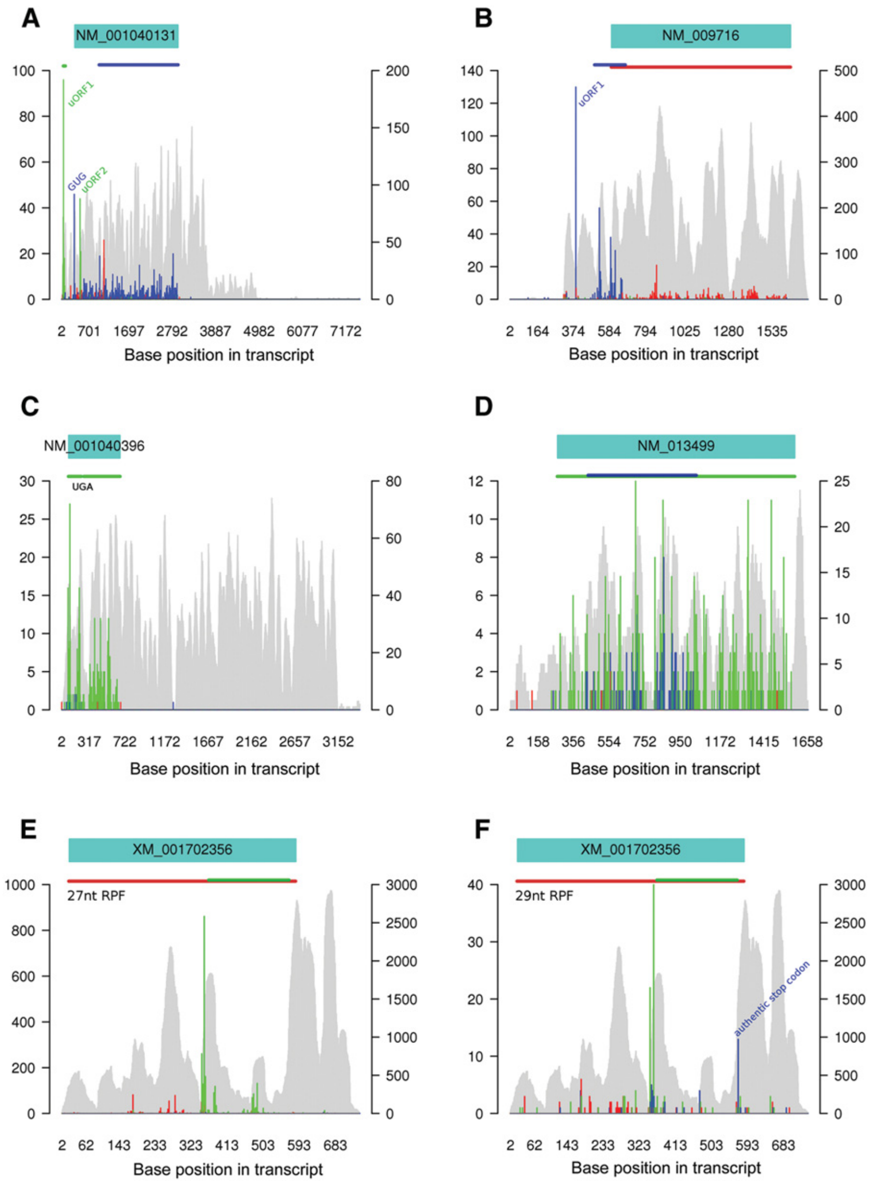

FIGURE 6. Translated uORFs and overlapping ORFs viewed using riboSeqR. (*A*) Analysis of RPFs mapping to NCBI RefSeq mRNA NM_001040131 (mouse eIF4G2), showing examples of uORFs. Histograms of the 5′ ends of RPFs (colored) and RNA-seq (gray) reads are shown. The three reading frames to which the 5′ ends of RPFs may map to (relative to nt 1 of the reference sequence) are indicated by different colors, as in Figure 4B. The positions of putative ORFs with at least 50 RPFs mapping to at least 10 locations are shown immediately *above* the histograms, color-coded as appropriate. A 15-codon sliding window mean of filtered (see Materials and Methods) Ribo-seq CHX RPF counts is shown *below* the transcript map. The main coding ORF for this transcript is in frame 2 (blue), but translation initiates upstream of the annotated ORF (at the indicated GUG codon). Two uORFs are apparent in frame 1 (green) and an overlapping ORF is also observed (red, see text). (*B*) Analysis of NM_009716 (mouse ATF4) reveals the two regulatory uORFs (see text), uORF1 and uORF2 (blue), upstream of the main coding sequence (red). (*C*) Analysis of NM_001040396 (mouse Sel T) showing an example of translation beyond a canonical stop codon by selenocysteine insertion at an in-frame UGA codon (indicated). Note this gene has proportionately a very long 3′ UTR. (*D*) An example of a likely internal overlapping gene from NM_013499 (Cr1l). (*E*,*F*) riboSeqR analysis of Rubisco expression as an example demonstrating a sequencing error in the RefSeq. This error results in an apparent change of reading frame of the 27-nt RPF ribosome profile. The lack of frame 1 (red) reads within the internal ORF and the clear detection of the de facto termination peak at position 568 (using the 29-nt data; *F*) confirms the RefSeq error.

The *Chlamydomonas* genome is less well-annotated than that of the mouse, and riboSeqR allowed us to identify and correct misannotations in RefSeq genes. In our initial analysis, Rubisco small subunit 2 was interpreted as having an internal overlapping gene (Fig. 6E, in green). Indeed, the RefSeq sequence, XM_001702356.1, contains a lengthy internal overlapping ORF. The position of this internal ORF coincided with a change in the phasing of the triplet periodicity of mapped 27-nt RPFs, indicating that translation is predominantly of the internal ORF in the region where it overlaps the annotated Rubisco ORF. However, there is an almost complete absence of RPFs mapping in phase with the main ORF in this region (absence of red spikes, presence of green spikes in the internal ORF of Fig. 6E). Subsequent investigation revealed that 8 nt are missing in the RefSeq, resulting in an incorrectly annotated amino acid sequence from amino acid position 110 onward. A peak in the 29-nt RPF plot (Fig. 6F) at the stop codon of the overlapping ORF in the +2 reading frame supports the idea that this stop codon, and not the annotated stop codon, is the major site of translation termination on the transcript. Although not a new discovery (the corrected amino acid is in agreement with the Uniprot sequence, P08475.1, as well as the latest *Chlamydomonas* transcriptome assembly Cre02.g120150), this example further illustrates the utility of riboSeqR.

DISCUSSION

Ribosomal RNA contamination of Ribo-seq libraries derives largely from RNase 1 cleavage of surface-exposed regions of the ribosome, which can generate rRNA fragments similar in size to RPFs and leads to their subsequent acquisition during gel size selection. Rigorous RNase 1 digestion to improve triplet periodicity further increases the likelihood of rRNA contamination. The level of contamination varies considerably between different species and experimental protocols. In *Saccharomyces cerevisiae*, ∼90% of the reads derive from one fragment of 28S rRNA, and a single round of hybridization-subtraction using an antisense oligonucleotide targeting this sequence is sufficient to remove much of it (Gerashchenko et al. 2012). In other organisms, however, such as human and mouse embryonic stem cells, the contaminants are more complex and libraries substantially enriched in specific RPFs can only be obtained if 10–20 rRNA fragments are subtracted by hybridization (Ingolia et al. 2012, 2013). Recently, commercial rRNA depletion kits (i.e., RiboZero) have also been used to deplete Ribo-seq libraries with some success (Bazzini et al. 2014; Smith et al. 2014). The use of DSN outlined in the present study offers an alternative approach to enrich for mRNA-specific RPFs in profiling libraries.

In a side-by-side comparison, we found DSN was able to deplete rRNA to a level comparable to RiboZero treatment (nine- to 11-fold versus 18-fold enrichment of mRNA, respectively; Fig. 3A). DSN was found to bias the most highly abundant individual mRNA read species (Fig. 3D), but, on the other hand, RiboZero treatment was found to bias individual mRNA read species that showed complementarity to the RiboZero probes (Fig. 3B). On a whole-transcript level, neither DSN nor RiboZero led to appreciable bias in the mRNA population (Fig. 3C), presumably due to biased individual RPF species being a relatively small fraction of the total number of RPF species even in the most highly expressed mRNAs. DSN treatment may be particularly useful for non-model organisms as it obviates the necessity to identify major rRNA contaminants in advance for hybridization-subtraction approaches. Furthermore, it will also be useful in situations where the rRNA contamination is diverse. With *C. reinhardtii*, for example, the complexity of contaminating sequences (derived mostly from the highly abundant cytoplasmic ribosomes, but including mitochondrial and chloroplastic rRNA sequences) rendered hybridization-subtraction methods inadequate (data not shown). DSN could also deplete other highly abundant contaminants that may be present in some samples, such as fragments of tRNAs, U2 snRNA, or snoRNAs. Furthermore, DSN is simple to use and can also be used in conjunction with other enrichment methods.

The riboSeqR R package was developed to enable non-specialists to parse aligned Ribo-seq data, to identify the predominant lengths of ribosomal fragments and the codon positions to which they map, and to thus identify coding sequences undergoing translation. A plethora of visualization methods are provided to facilitate this task and summarize the behavior of the ribosome, both on average (i.e., summed over many mRNAs; Figs. 4, 5) and for individual transcripts (Fig. 6). We used the mouse and *Chlamydomonas* data sets here to show the versatility of the package in the identification and visualization of established examples of noncanonical translation. Good triplet periodicity in Ribo-seq data sets is advantageous in the identification by riboSeqR of previously unannotated ORFs (Gerashchenko et al. 2012; Michel et al. 2014) and of overlapping ORFs. The quality of framing is influenced by the extent of RNase 1 digestion and for *Chlamydomonas*, over a range of RNase 1 concentrations between 600 and 1600 units per mL of lysate ($A_{254} = 4$), we found that the most abundant RPFs showed framing between 85% and 96% (data not shown). The digestion conditions utilized for the mouse lysates were identical to those described by Ingolia et al. (2012) and generated good framing without further optimization.

Identification and assessment of differential translation, in which the ratio of translation as assessed by Ribo-seq to transcription as assessed by RNA-seq varies between biological conditions, present the next challenges in understanding translation regulation. In this paper, we have considered data from a single biological sample and so differential expression analyses are not relevant. However, methods are provided in the riboSeqR package for extracting count data for translated coding sequences for analysis of differential translation by summing over specific size class/frame combinations. These data may be paired with counts from RNA-seq data from the same samples for analyses of differential translation efficiency, which we suggest may be achieved through methods for analysis of paired high-throughput sequencing data, as implemented in the BaySeq R package Hardcastle and Kelly 2013.

MATERIALS AND METHODS

Ribosomal profiling

The profiling methodologies used were based largely on those described by Ingolia and colleagues (Ingolia et al. 2009, 2012), except library amplicons were constructed using a small RNA cloning strategy (Guo et al. 2010) adapted to Illumina smallRNA v2 to allow multiplexing.

Cell culture and lysis

Chlamydomonas reinhardtii cells (CC-4350 cw[15] Arg 7–8 mt$^+$) (Chlamydomonas Resource Center: http://chlamycollection.org/strains/) were maintained in 750 mL Tris–acetate–phosphate medium (Harris 1989) at 23°C on a rotatory shaker (140 rpm) under constant illumination with white light (70 µE m^2 sec^{-1}) to mid-log phase (OD_{750}~0.6). Cultures were harvested by filtering off the media, the cell paste was flash frozen and pulverized in liquid nitrogen with 5 mL of prefrozen lysis buffer (20 mM Tris–Cl pH7.5, 140 mM KCl, 5 mM $MgCl_2$, 100 µg/mL cycloheximide, 100 µg/mL chloramphenicol, 0.05 mM DTT, 0.1% NP40 and 5% sucrose). The frozen powder was thawed on ice and clarified by centrifugation for 30 min at 4700 rpm at 4°C followed by adjustment of A_{254} to ~4 before snap-freezing in liquid nitrogen and storage at −80°C.

Murine 17 clone 1 cells were maintained in Dulbecco's modification of Eagle's medium supplemented with 10% (vol/vol) fetal calf serum. Cells (10^7) were plated in a 10 cm dish and upon reaching 100% confluence, cycloheximide (Sigma-Aldrich) was added to 100 µg/mL. After 2 min, cells were rinsed with 5 mL of ice-cold PBS, the dishes submerged in a reservoir of liquid nitrogen for 10 sec, transferred to dry ice and 400 µL of lysis buffer (20 mM Tris–HCl pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 1% Triton X-100, 100 µg/mL cycloheximide and 25 units/mL TURBO DNase [Life Technologies]) dripped on. The cells were scraped extensively to ensure lysis, collected and triturated with a 26-G needle 10 times. Lysates were clarified by centrifugation for 20 min at 13,000*g* at 4°C, the supernatants recovered and stored in liquid nitrogen.

Nuclease footprinting

For *Chlamydomonas*, lysates were slowly thawed on ice and a 200 µL aliquot (A_{254} = 4) treated with 300 units RNase 1 (100 units/µL, Life Technologies cat. no. AM2294) in a thermo-mixer at 28°C, 400 rpm for 30 min. The tube was placed on ice, 2 µL of SUPERase-In RNase inhibitor (20 units/mL, Life Technologies) added, and the reaction was layered onto a 1 M sucrose cushion prepared in *Chlamydomonas* polysome buffer (20 mM Tris–HCl pH 7.5, 140 mM KCl, 5 mM $MgCl_2$, 0.5 mM DTT, 100 µg/mL cycloheximide, 100 µg/mL chloramphenicol, and 0.5 µg/mL SUPERase-In). The cushion was ultracentrifuged at 38,000 rpm (5 h, 4°C) in a Beckman Sw41Ti rotor. For mouse samples, lysates were slowly thawed on ice and 300 µL treated with 7.5 µL RNase 1 followed by incubation for 45 min at room temperature on a rotating wheel. Ten microliters of SUPERase-In RNase inhibitor was added, the sample was layered onto a 1 M sucrose cushion in mammalian polysome buffer (20 mM Tris–HCl pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 100 µg/mL cycloheximide) and ultracentrifuged at 28,000 rpm (16 h, 4°C) in a Beckman SW55Ti rotor. Subsequently, all ribosome pellets were resuspended in 200 µL of the corresponding polysome buffer and digested with proteinase K (10 mM Tris–HCl pH 7.5, 10% SDS, 200 µg/mL Proteinase K [New England BioLabs]) for 30 min at 42°C. RPFs were recovered by extracting twice with prewarmed (65°C) acidic phenol:chloroform (Life Technologies) and once with chloroform (1:1, vol/vol, buffered with 10 mM Tris pH 7.5, 0.1 mM EDTA) followed by ethanol precipitation. RPFs were resuspended in 10 mM Tris–HCl pH 7.5 and quantified by spectrophotometry.

Purification of total RNA and fragmentation by alkaline hydrolysis

Two hundred microliters of each cell lysate was digested with proteinase K and cellular RNA extracted using acidic phenol:chloroform as above. Ribosomal RNA was depleted from 5 µg of total RNA using a RiboZero rRNA removal kit targeting the appropriate species following the manufacturer's instructions (Human/Mouse/Rat: Epicentre, cat. no. RZH1046, Plant Seed/root: Epicentre, cat. no. MRZSR116). Depleted RNA was resuspended in 10 mM Tris–HCl pH 7.5 and quantified by spectrophotometry. A measure of 1–2 µg of total RNA in 20 µL was mixed with an equal volume of 2× alkaline fragmentation solution (2 mM EDTA, 10 mM Na_2CO_3, and 90 mM $NaHCO_3$) and incubated for 15 min at 95°C. The reaction was diluted by addition of 280 µL stop/precipitation solution (300 mM NaOAc pH 5.5, GlycoBlue Co-precipitant [Ambion, 15 mg/mL]), and fragmented RNA recovered by ethanol precipitation.

RNA size selection

Fragmented total RNA or RPFs (1–2 µg) were separated on 15% (wt/vol) denaturing polyacrylamide gels and RNA species migrating between 28 and 34 nt were harvested. RNA was eluted from the gel slices on a rotator overnight at 4°C, in 600 µL RNA gel extraction buffer (300 mM NaOAc pH 5.5, 1 mM EDTA, and 0.25% SDS). Eluted RNA was ethanol precipitated as described above.

Generation of RNA libraries

The RNA samples from above were heated at 80°C for 2 min, cooled and the 3′ phosphate group removed using T4 polynucleotide kinase (T4 PNK, New England BioLabs) for 2 h at 37°C in a 20 µL reaction lacking ATP. The RNA was concentrated by ethanol precipitation, resuspended in 10 mM Tris–HCl pH 7.5 and ligated in a 20 µL reaction overnight at 14°C to a preadenylated 3′-adaptor (5′-rATGGAATTCTCGGGTGCCAAGG-3′) using T4 RNA Ligase 2 truncated K227Q (New England BioLabs). This 3′ adaptor was adenylated using a 5′ DNA adenylation kit (New England BioLabs) following the manufacturer's instructions. RNA was precipitated, loaded into a 15% (wt/vol) denaturing polyacrylamide gel and ligated RNA fragments migrating between 49 and 53 nt excised. The RNA was eluted, precipitated, resuspended as above, and 5′ phosphorylated using T4 PNK in the presence of 1 mM ATP for 2 h at 37°C. RNA was concentrated by ethanol precipitation, resuspended in 10 mM Tris–HCl pH 7.5 and ligated to a 5′ RNA adaptor (5′-rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrC-3′) in a 20 µL reaction overnight at 14°C using T4 RNA Ligase (Promega). The fully adapted RNAs were recovered by ethanol precipitation, dissolved in 6 µL 10 mM Tris–HCl pH 7.5, and 3 µL of this ligated product annealed to an RT primer (5′-GCCTTGGCACCCGAGAATTCCA, 50 pmol) for 5 min at 65°C. The RNA was subsequently reverse transcribed for 50 min at 55°C in a 20 µL reaction following addition of first strand buffer (Invitrogen, to 1×), 2.5 mM $MgCl_2$, 10 mM DTT, 0.5 µL SUPERase-In, and 1 µL SuperScript III (Invitrogen), followed by heat inactivation for 5 min at 85°C.

PCR amplification and barcode addition

Standard PCR reactions were used to prepare amplicons using forward primer RP1 (5′-AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGA-3′) and 5′-CAAGCAGAAGACGGCATACGAGATN$_6$GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA-3′ (RPIX) as reverse primer, where X is primer number (1–24) and N_6 the reverse complement of the respective hexanucleotide index sequence detected during Illumina sequencing. *Chlamydomonas* PCR amplification used New England BioLabs (NEB) Q5 2X master mix (because of high GC content) and was performed using a ramp-rate of 2.2°C/sec with the following cycling conditions: one cycle of 98°C for 3 min, 13 cycles of 98°C for 1 min, 65°C for 30 sec, and 72°C for 30 sec, followed by an elongation step of 72°C for 5 min. PCR of mouse cDNA used Phusion High-Fidelity PCR Master Mix (NEB) and comprised one cycle of 98°C for 30 sec, 13 cycles of 98°C for 10 sec, 60°C for 30 sec, and 72° C for 15 sec, followed by an elongation step of 72°C for 10 min. PCR reactions were loaded onto 10% nondenaturing polyacrylamide-TBE gels and run for 45 min at 12 W. Products of ∼150 bp were excised from the gel and eluted at 4°C overnight on a rotator following addition of 600 µL DNA gel extraction buffer (300 mM NaCl, 10 mM Tris–HCl pH 8 and 1 mM EDTA). These amplicon libraries were ethanol precipitated and resuspended in 15 µL 10 mM Tris–HCl pH 7.5. Libraries were sequenced using Illumina HiSeq2000, NextSeq500, or MiSeq platforms.

Ribosomal RNA depletion

RiboZero-based rRNA subtraction

Following nuclease footprinting, RPFs (2 µg) were subjected to RiboZero treatment as detailed above for total cellular RNA. Subsequent gel purification of appropriately sized RPFs and library amplicon generation was carried out in parallel with an undepleted library to allow unambiguous band identification.

Antisense oligonucleotide (AON)-based rRNA subtraction

Following adaptor-ligation and reverse transcription, major rRNA sequence contaminants in library cDNAs were targeted by annealing to a pool of 14 biotinylated AONs and subsequently removed using streptavidin beads as previously described (Ingolia et al. 2012).

Treatment with duplex-specific nuclease (DSN)

Ribosomal RNA was depleted from Ribo-seq samples at the library amplicon stage; 12 µL of the relevant Ribo-seq library was mixed with 4 µL of 4× hybridization buffer (200 mM HEPES pH 7.5 and 2 M NaCl) and denatured at 98°C for 2 min. DNA was re-annealed for 5 h at 68°C prior to addition of 2 µL of 10× DSN master buffer and 2 µL of DSN (4 units, Evrogen). Digestion was allowed to proceed for 25 min at 68°C (mouse) or up to 90 min (*Chlamydomonas*), before addition of 20 µL 10 mM EDTA and incubation for a further 5 min at 68°C. DNA was recovered by a single extraction with phenol–chloroform (1:1, vol/vol) followed by ethanol precipitation and resuspended in 4 µL 10 mM Tris–HCl pH 7.5. The treated amplicon library was subjected to another round of PCR (as above) and the resulting library sequenced or subjected to a second round of DSN treatment.

Bioinformatic analysis

Adaptor sequences were trimmed using the FastX-toolkit. To remove remaining post-DSN contamination, trimmed reads were first mapped to *C. reinhardtii* and *Mus musculus* (as appropriate) databases of rRNA and common noncoding RNAs (ncRNAs), using Bowtie version 1 (Langmead et al. 2009) with seed length 23. In order to select good-quality samples of nuclear-encoded, mitochondrial and chloroplastic mRNA-derived RPFs, remaining reads were mapped to *C. reinhardtii* and *M. musculus* NCBI RefSeq mRNAs, and organellar coding ORFs derived from NCBI RefSeq *C. reinhardtii* and *M. musculus* organellar genomes. No specific consideration was given to the presence of multiple isoforms within the RefSeq database: Each read that could be mapped to multiple transcripts was assigned at random to one of these transcripts. The Bowtie alignment was used as input for riboSeqR, along with a FASTA file containing the transcriptome of interest, to generate Figures 4–6.

RPF framing distributions produced with riboSeqR (Figs. 4B, 5B) were derived from reads mapping to the "interior" regions of annotated coding ORFs; specifically the entire read had to be contained within the ORF, thus, in general, excluding RPFs of initiating or terminating ribosomes. Histograms of 5′ end positions of RPFs relative to start and stop codons were derived from reads mapping to RefSeq mRNAs with at least 50 reads of the most abundant read-length size class mapping in frame in at least 10 separate locations in the ORF. For a given length size class, the values shown represent a weighted average of the abundance of reads mapping on all selected transcripts.

For the comparison of rRNA depletion strategies (Figs. 2, 3), library composition was determined by mapping reads to rRNA, mRNA, and genomic (gDNA) databases. Remaining reads that mapped to gDNA presumably derived from ncRNAs and unannotated transcripts (both ncRNA and mRNA). Minimum free folding energies of individual reads were calculated using RNAfold from the ViennaRNA package (Hofacker 2003). Since DSN is applied post-adaptor-ligation and post RT-PCR, the minimum free energy of reads was calculated using DNA energy parameters (dna_mathews2004.par). The untreated and test libraries were sequenced together using unique multiplex tags at nt 34–39 of the 63 nt 3′ adaptor sequence. Since the multiplex tag produces a systematic bias in the calculated MFE that varies from one sample to another, the MFE was calculated for each read in the context of the 5′ adaptor and just the first 33 nt of the 3′ adaptor. Optimal RNA:RNA binding energies between reads and reverse-complemented rRNA (RiboZero) were calculated using RNAduplex from the ViennaRNA package with default RNA energy parameters. For both RNAfold and RNAduplex, the temperature parameter was set to 68°C, the annealing temperature used in the DSN and RiboZero protocols.

DATA DEPOSITION

The Ribo-seq data have been deposited in the ArrayExpress database (http://www.ebi.ac.uk/arrayexpress) under accession numbers E-MTAB-2934 and E-MTAB-3583.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

This work was supported by an EMBL long-term postdoctoral fellowship to B.Y.C., Sir Henry Wellcome Fellowships to B.Y.C. and N.I., a Wellcome Trust PhD scholarship to J.D.J., a Wellcome Trust Fellowship to A.E.F. (088789), and UK Biotechnology and Biological Sciences Research Council grants to I.B. (BB/L000334/1) and A.E.F. (BB/J007072/1). Work in the Baulcombe laboratory is supported by The Gatsby Charitable Foundation and the European Research Council Advanced Investigator grant TRIBE. D.C.B. is the Royal Society Edward Penley Abraham Research Professor. We wish to thank Professor Stuart G. Siddell, University of Bristol, for providing the murine 17 clone 1 cells.

Received May 15, 2015; accepted June 23, 2015.

REFERENCES

Ait Ghezala H, Jolles B, Salhi S, Castrillo K, Carpentier W, Cagnard N, Bruhat A, Fafournoux P, Jean-Jean O. 2012. Translation termination efficiency modulates ATF4 response by regulating ATF4 mRNA translation at 5′ short ORFs. *Nucleic Acids Res* **40:** 9557–9570.

Alkalaeva EZ, Pisarev AV, Frolova LY, Kisselev LL, Pestova TV. 2006. In vitro reconstitution of eukaryotic translation reveals cooperativity between release factors eRF1 and eRF3. *Cell* **125:** 1125–1136.

Arias C, Weisburd B, Stern-Ginossar N, Mercier A, Madrid AS, Bellare P, Holdorf M, Weissman JS, Ganem D. 2014. KSHV 2.0: a comprehensive annotation of the Kaposi's sarcoma-associated herpesvirus genome using next-generation sequencing reveals novel genomic and functional features. *PLoS Pathog* **10:** e1003847.

Artieri CG, Fraser HB. 2014. Evolution at two levels of gene expression in yeast. *Genome Res* **24:** 411–421.

Baudin-Baillieu A, Legendre R, Kuchly C, Hatin I, Demais S, Mestdagh C, Gautheret D, Namy O. 2014. Genome-wide translational changes induced by the prion [PSI+]. *Cell Rep* **8:** 439–448.

Bazzini AA, Johnstone TG, Christiano R, Mackowiak SD, Obermayer B, Fleming ES, Vejnar CE, Lee MT, Rajewsky N, Walther TC, et al. 2014. Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO J* **33:** 981–993.

Berry MJ, Banu L, Harney JW, Larsen PR. 1993. Functional characterization of the eukaryotic SECIS elements which direct selenocysteine insertion at UGA codons. *EMBO J* **12:** 3315–3322.

Brar GA, Yassour M, Friedman N, Regev A, Ingolia NT, Weissman JS. 2012. High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science* **335:** 552–557.

Brierley I, Gilbert RJC, Pennell S. 2010. Pseudoknot-dependent programmed-1 ribosomal frameshifting: structures, mechanisms and models. In *Recoding: expansion of decoding rules enriches gene expression* (ed. Atkins JF, Gesteland RF), pp. 149–174. Springer, New York.

Brubaker SW, Gauthier AE, Mills EW, Ingolia NT, Kagan JC. 2014. A bicistronic MAVS transcript highlights a class of truncated variants in antiviral immunity. *Cell* **156:** 800–811.

Caro F, Ahyong V, Betegon M, DeRisi JL. 2014. Genome-wide regulatory dynamics of translation in the *Plasmodium falciparum* asexual blood stages. *Elife* **3:** e04106.

Christodoulou DC, Gorham JM, Herman DS, Seidman JG. 2011. Construction of normalized RNA-seq libraries for next-generation sequencing using the crab duplex-specific nuclease. *Curr Protoc Mol Biol* **94:** 4.12.1–4.12.11.

Crappé J, Ndah E, Koch A, Steyaert S, Gawron D, De Keulenaer S, De Meester E, De Meyer T, Van Criekinge W, Van Damme P, et al. 2015. PROTEOFORMER: deep proteome coverage through ribosome profiling and MS integration. *Nucleic Acids Res* **43:** e29.

de Klerk E, Fokkema IF, Thiadens KA, Goeman JJ, Palmblad M, den Dunnen JT, von Lindern M, 't Hoen PA. 2015. Assessing the translational landscape of myogenic differentiation by ribosome profiling. *Nucleic Acids Res* **43:** 4408–4428.

Duncan CD, Mata J. 2014. The translational landscape of fission-yeast meiosis and sporulation. *Nat Struct Mol Biol* **21:** 641–647.

Dunn JG, Foo CK, Belletier NG, Gavis ER, Weissman JS. 2013. Ribosome profiling reveals pervasive and regulated stop codon readthrough in *Drosophila melanogaster*. *Elife* **2:** e01179.

Gao X, Wan J, Liu B, Ma M, Shen B, Qian SB. 2015. Quantitative profiling of initiating ribosomes in vivo. *Nat Methods* **12:** 147–153.

Gerashchenko MV, Gladyshev VN. 2014. Translation inhibitors cause abnormalities in ribosome profiling experiments. *Nucleic Acids Res* **42:** e134.

Gerashchenko MV, Lobanov AV, Gladyshev VN. 2012. Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress. *Proc Natl Acad Sci* **109:** 17394–17399.

Gonzalez C, Sims JS, Hornstein N, Mela A, Garcia F, Lei L, Gass DA, Amendolara B, Bruce JN, Canoll P, Sims PA. 2014. Ribosome profiling reveals a cell-type-specific translational landscape in brain tumors. *J Neurosci* **34:** 10924–10936.

Guo H, Ingolia NT, Weissman JS, Bartel DP. 2010. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466:** 835–840.

Haft RJ, Keating DH, Schwaegler T, Schwalbach MS, Vinokur J, Tremaine M, Peters JM, Kotlajich MV, Pohlmann EL, Ong IM, et al. 2014. Correcting direct effects of ethanol on translation and transcription machinery confers ethanol tolerance in bacteria. *Proc Natl Acad Sci* **111:** E2576–E2585.

Hardcastle TJ, Kelly KA. 2013. Empirical Bayesian analysis of paired high-throughput sequencing data with a beta-binomial distribution. *BMC Bioinformatics* **14:** 135.

Harding HP, Novoa I, Zhang Y, Zeng H, Wek R, Schapira M, Ron D. 2000. Regulated translation initiation controls stress-induced gene expression in mammalian cells. *Mol Cell* **6:** 1099–1108.

Harris EH. 1989. *The Chlamydomonas sourcebook: a comprehensive guide to biology and laboratory use*. Academic Press, San Diego.

Heider J, Baron C, Böck A. 1992. Coding from a distance: dissection of the mRNA determinants required for the incorporation of selenocysteine into protein. *EMBO J.* **11:** 3759–3766.

Hendriks GJ, Gaidatzis D, Aeschimann F, Großhans H. 2014. Extensive oscillatory gene expression during *C. elegans* larval development. *Mol Cell* **53:** 380–392.

Hofacker IL. 2003. Vienna RNA secondary structure server. *Nucleic Acids Res* **31:** 3429–3431.

Huang Y, Ainsley JA, Reijmers LG, Jackson FR. 2013. Translational profiling of clock cells reveals circadianly synchronized protein synthesis. *PLoS Biol* **11:** e1001703.

Ingolia NT. 2010. Genome-wide translational profiling by ribosome footprinting. *Methods Enzymol* **470:** 119–142.

Ingolia NT. 2014. Ribosome profiling: new views of translation, from single codons to genome scale. *Nat Rev Genet* **15:** 205–213.

Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. 2009. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324:** 218–223.

Ingolia NT, Lareau LF, Weissman JS. 2011. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147:** 789–802.

Ingolia NT, Brar GA, Rouskin S, McGeachy AM, Weissman JS. 2012. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat Protoc* **7:** 1534–1550.

Ingolia NT, Brar GA, Rouskin S, McGeachy AM, Weissman JS. 2013. Genome-wide annotation and quantitation of translation by ribosome profiling. *Curr Protoc Mol Biol* **103:** 4.18.1–4.18.19.

Jackson R, Standart N. 2015. The awesome power of ribosome profiling. *RNA* **21:** 652–654.

Jensen BC, Ramasamy G, Vasconcelos EJ, Ingolia NT, Myler PJ, Parsons M. 2014. Extensive stage-regulation of translation revealed by ribosome profiling of *Trypanosoma brucei*. *BMC Genomics* **15:** 911.

Katz Y, Li F, Lambert NJ, Sokol ES, Tam WL, Cheng AW, Airoldi EM, Lengner CJ, Gupta PB, Yu Z, et al. 2014. Musashi proteins are post-transcriptional regulators of the epithelial-luminal cell state. *Elife* **7:** e03915.

Kronja I, Yuan B, Eichhorn SW, Dzeyk K, Krijgsveld J, Bartel DP, Orr-Weaver TL. 2014. Widespread changes in the posttranscriptional landscape at the *Drosophila* oocyte-to-embryo transition. *Cell Rep* **7:** 1495–1508.

Kryukov GV, Castellano S, Novoselov SV, Lobanov AV, Zehtab O, Guigó R, Gladyshev VN. 2003. Characterization of mammalian selenoproteomes. *Science* **300:** 1439–1443.

Labunskyy VM, Gerashchenko MV, Delaney JR, Kaya A, Kennedy BK, Kaeberlein M, Gladyshev VN. 2014. Lifespan extension conferred by endoplasmic reticulum secretory pathway deficiency requires induction of the unfolded protein response. *PLoS Genet* **10:** e1004019.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10:** R25.

Lareau LF, Hite DH, Hogan GJ, Brown PO. 2014. Distinct stages of the translation elongation cycle revealed by sequencing ribosome-protected mRNA fragments. *Elife* **3:** e01257.

Lee MT, Bonneau AR, Takacs CM, Bazzini AA, DiVito KR, Fleming ES, Giraldez AJ. 2013. Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. *Nature* **503:** 360–364.

Legendre R, Baudin-Baillieu A, Hatin I, Namy O. 2015. RiboTools: a Galaxy toolbox for qualitative ribosome profiling analysis. *Bioinformatics* **31:** 2586–2588.

Liu X, Jiang H, Gu Z, Roberts JW. 2013. High-resolution view of bacteriophage lambda gene expression by ribosome profiling. *Proc Natl Acad Sci* **110:** 11928–11933.

Manuell AL, Quispe J, Mayfield SP. 2007. Structure of the chloroplast ribosome: novel domains for translation regulation. *PLoS Biol* **5:** e209.

Matvienko M, Kozik A, Froenicke L, Lavelle D, Martineau B, Perroud B, Michelmore R. 2013. Consequences of normalizing transcriptomic and genomic libraries of plant genomes using a duplex-specific nuclease and tetramethylammonium chloride. *PLoS One* **8:** e55913.

McManus CJ, May GE, Spealman P, Shteyman A. 2014. Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. *Genome Res* **24:** 422–430.

Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, Witman GB, Terry A, Salamov A, Fritz-Laylin LK, Maréchal-Drouard L, et al. 2007. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* **318:** 245–250.

Michel AM, Baranov PV. 2013. Ribosome profiling: a Hi-Def monitor for protein synthesis at the genome-wide scale. *Wiley Interdiscip Rev RNA* **4:** 473–490.

Michel AM, Choudhury KR, Firth AE, Ingolia NT, Atkins JF, Baranov PV. 2012. Observation of dually decoded regions of the human genome using ribosome profiling data. *Genome Res* **22:** 2219–2229.

Michel AM, Fox G, M Kiran A, De Bo C, O'Connor PB, Heaphy SM, Mullan JP, Donohue CA, Higgins DG, Baranov PV. 2014. GWIPS-viz: development of a ribo-seq genome browser. *Nucleic Acids Res* **42:** D859–D864.

Miller DF, Yan PS, Buechlein A, Rodriguez BA, Yilmaz AS, Goel S, Lin H, Collins-Burow B, Rhodes LV, Braun C, et al. 2013. A new method for stranded whole transcriptome RNA-seq. *Methods* **63:** 126–134.

Morris DR. 2009. Ribosomal footprints on a transcriptome landscape. *Genome Biol* **10:** 215.

Muzzey D, Sherlock G, Weissman JS. 2014. Extensive and coordinated control of allele-specific expression by both transcription and translation in *Candida albicans*. *GenomeRes* **24:** 963–973.

O'Connor PB, Li GW, Weissman JS, Atkins JF, Baranov PV. 2013. rRNA:mRNA pairing alters the length and the symmetry of mRNA-protected fragments in ribosome profiling experiments. *Bioinformatics* **29:** 1488–1491.

Paul MS, Aegerter M, O'Brien SE, Kurtz CB, Weis JH. 1989. The murine complement receptor gene family. Analysis of mCRY gene products and their homology to human CR1. *J Immunol* **142:** 582–589.

Rooijers K, Loayza-Puch F, Nijtmans LG, Agami R. 2013. Ribosome profiling reveals features of normal and disease-associated mitochondrial translation. *Nat Commun* **4:** 2886.

Schrader JM, Zhou B, Li GW, Lasker K, Childers WS, Williams B, Long T, Crosson S, McAdams HH, Weissman JS, et al. 2014. The coding and noncoding architecture of the *Caulobacter crescentus* genome. *PLoS Genet* **10:** e1004463.

Shagin DA, Rebrikov DV, Kozhemyako VB, Altshuler IM, Shcheglov AS, Zhulidov PA, Bogdanova EA, Staroverov DB, Rasskazov VA, Lukyanov S. 2002. A novel method for SNP detection using a new duplex-specific nuclease from crab hepatopancreas. *Genome Res* **12:** 1935–1942.

Shagina I, Bogdanova E, Mamedov IZ, Lebedev Y, Lukyanov S, Shagin D. 2010. Normalization of genomic DNA using duplex-specific nuclease. *Biotechniques* **48:** 455–459.

Sidrauski C, McGeachy AM, Ingolia NT, Walter P. 2015. The small molecule ISRIB reverses the effects of eIF2α phosphorylation on translation and stress granule assembly. *Elife* **6:** 4.

Smith JE, Alvarez-Dominguez JR, Kline N, Huynh NJ, Geisler S, Hu W, Coller J, Baker KE. 2014. Translation of small open reading frames within unannotated RNA transcripts in *Saccharomyces cerevisiae*. *Cell Rep* **7:** 1858–1866.

Stadler M, Fire A. 2013. Conserved translatome remodeling in nematode species executing a shared developmental transition. *PLoS Genet* **9:** e1003739.

Stern-Ginossar N. 2015. Decoding viral infection by ribosome profiling. *J Virol* doi: 10.1128/JVI.02528-14.

Stern-Ginossar N, Weisburd B, Michalski A, Le VT, Hein MY, Huang SX, Ma M, Shen B, Qian SB, Hengel H, et al. 2012. Decoding human cytomegalovirus. *Science* **338:** 1088–1093.

Stumpf CR, Moreno MV, Olshen AB, Taylor BS, Ruggero D. 2013. The translational landscape of the mammalian cell cycle. *Mol Cell* **52:** 574–582.

Subramaniam AR, Deloughery A, Bradshaw N, Chen Y, O'Shea E, Losick R, Chai Y. 2013. A serine sensor for multicellularity in a bacterium. *Elife* **2:** e01501.

Subtelny AO, Eichhorn SW, Chen GR, Sive H, Bartel DP. 2014. Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* **508:** 66–71.

Takahashi K, Maruyama M, Tokuzawa Y, Murakami M, Oda Y, Yoshikane N, Makabe KW, Ichisaka T, Yamanaka S. 2005. Evolutionarily conserved non-AUG translation initiation in *NAT1/p97/DAP5 (EIF4G2)*. *Genomics* **85:** 360–371.

Vaidyanathan PP, Zinshteyn B, Thompson MK, Gilbert WV. 2014. Protein kinase A regulates gene-specific translational adaptation in differentiating yeast. *RNA* **20:** 912–922.

Vasquez JJ, Hon CC, Vanselow JT, Schlosser A, Siegel TN. 2014. Comparative ribosome profiling reveals extensive translational complexity in different *Trypanosoma brucei* life cycle stages. *Nucleic Acids Res* **42:** 3623–3637.

Vattem KM, Wek RC. 2004. Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. *Proc Natl Acad Sci* **101:** 11269–11274.

Weiss RB, Atkins JF. 2011. Molecular biology. Translation goes global. *Science* **334:** 1509–1510.

Wickersheim ML, Blumenstiel JP. 2013. Terminator oligo blocking efficiently eliminates rRNA from *Drosophila* small RNA sequencing libraries. *Biotechniques* **55:** 269–272.

Wiita AP, Ziv E, Wiita PJ, Urisman A, Julien O, Burlingame AL, Weissman JS, Wells JA. 2013. Global cellular response to chemotherapy-induced apoptosis. *Elife* **2:** e01236.

Williams CC, Jan CH, Weissman JS. 2014. Targeting and plasticity of mitochondrial proteins revealed by proximity-specific ribosome profiling. *Science* **346:** 748–751.

Wolin SL, Walter P. 1988. Ribosome pausing and stacking during translation of a eukaryotic mRNA. *EMBO J* **7:** 3559–3569.

Yang Z, Cao S, Martens CA, Porcella SF, Xie Z, Ma M, Shen B, Moss B. 2015. Deciphering poxvirus gene expression by RNA sequencing and ribosome profiling. *J Virol* **89:** 6874–6886.

Yi H, Cho YJ, Won S, Lee JE, Jin Yu H, Kim S, Schroth GP, Luo S, Chun J. 2011. Duplex-specific nuclease efficiently removes rRNA for prokaryotic RNA-seq. *Nucleic Acids Res* **39:** e140.

Zid BM, O'Shea EK. 2014. Promoter sequences direct cytoplasmic localization and translation of mRNAs during starvation in yeast. *Nature* **514:** 117–121.

RNA SOCIETY

RNA

A PUBLICATION OF THE RNA SOCIETY

The use of duplex-specific nuclease in ribosome profiling and a user-friendly software package for Ribo-seq data analysis

Betty Y. Chung, Thomas J. Hardcastle, Joshua D. Jones, et al.

RNA published online August 18, 2015

Silencing lncRNA reduces metastasis
– watch the new webinar now
EXIQON