Briefings in Bioinformatics, 2017, 1–12

doi: 10.1093/bib/bbx093
Paper

Computational resources for ribosome profiling: from database to Web server and software

Hongwei Wang*, Yan Wang* and Zhi Xie

Corresponding authors: Hongwei Wang, State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, China. Tel.: +86 20 87335131; E-mail: bioccwhw@126.com; Zhi Xie, State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, China. Tel.: +86 20 87335131; E-mail: xiezhi@gmail.com
*These authors contributed equally to the work as joint first authors.

Abstract

Ribosome profiling is emerging as a powerful technique that enables genome-wide investigation of *in vivo* translation at sub-codon resolution. The increasing application of ribosome profiling in recent years has achieved remarkable progress toward understanding the composition, regulation and mechanism of translation. This benefits from not only the awesome power of ribosome profiling but also an extensive range of computational resources available for ribosome profiling. At present, however, a comprehensive review on these resources is still lacking. Here, we survey the recent computational advances guided by ribosome profiling, with a focus on databases, Web servers and software tools for storing, visualizing and analyzing ribosome profiling data. This review is intended to provide experimental and computational biologists with a reference to make appropriate choices among existing resources for the question at hand.

Key words: ribosome profiling; translational regulation; database; Web server; software

Introduction

Precise control of gene expression can be directly modulated not only at the transcriptional level but also at the translational level. Translation, serving as an important step within the cascade of gene expression regulation, plays an essential role in controlling gene expression [1, 2]. Translational alterations have been found to underlie many human diseases and increase disease susceptibility [3]. Emerging evidence has shown that targeting components of translation apparatus may be a promising strategy for disease treatment [4, 5]. Although substantial progress has been made during the past years, our understanding of translational control in comparison with transcriptional control remains relatively limited.

With rapid advances in high-throughput sequencing technologies, a revolutionary technique, ribosome profiling (ribo-seq), was developed that uses deep sequencing to enable genome-wide, quantitative analysis of *in vivo* translation with sub-codon resolution [6, 7]. By precisely pinpointing ribosomes during translation, this technique has provided unprecedented insights into the complexity of translatome [8–12]. Previous studies using ribo-seq have shown: (1) pervasive translation inside and outside of annotated protein-coding regions, including 5′ untranslated regions (UTRs), 3′ UTRs and long noncoding RNA (lncRNA) regions such as upstream open reading frames (uORFs), downstream open reading frames (dORFs) and small open reading frames (sORFs) [13–18]; (2) dynamics of translational regulation, including translational initiation, elongation and termination in response to changing environments such as circadian rhythms, stress, viral infections and drugs [12, 19–21]; and (3) translational mechanisms, including non-canonical initiation, elongation and termination such as ribosome shunting, ribosomal frameshifting and stop codon readthrough [11, 22–24]. These findings not only illuminate the importance of translational control but also advance our understanding in the composition, regulation and mechanism of translation.

Hongwei Wang (PhD) is a research fellow in Zhongshan Ophthalmic Center at Sun Yat-sen University. His research is focused on translational regulation of gene expression.
Yan Wang (MS) is a research assistant in Zhongshan Ophthalmic Center at Sun Yat-sen University. Her research is focused on high-throughput omics data analysis.
Zhi Xie (MD, PhD) is a professor of Bioinformatics in Zhongshan Ophthalmic Center at Sun Yat-sen University. He is interested in understanding posttranscriptional and translational regulation through the integrative analysis of multi-omics data.
Submitted: 3 May 2017; **Received (in revised form):** 5 July 2017

To facilitate understanding of ribo-seq data, a large and diverse set of computational resources have been developed to date, spanning a broad spectrum from database to Web server and software. Some specialized databases, such as GWIPS-viz [25], RPFdb [26] and sORFs.org [27], have been built to exhaustively collate ribo-seq data and its related contents. Several Web applications and visualization toolkits, such as RiboGalaxy [28] and riboviz [29], have been designed to facilitate the exploration and analysis of ribo-seq data. Moreover, many stand-alone software tools have been specifically developed for ribo-seq data, such as Plastid [30], RiboProfiling [31] and systemPipeR [32] for data preprocessing, RUST [33] for data normalization, anota [34], Babel [35] and Xtail [36] for differential translation detection and so on [13–15, 37–51]. Undoubtedly, these computational resources have led to a great step forward in the storage, manipulation, analysis, presentation and interpretation of ribo-seq data. However, given the current landscape of computational resources available for ribo-seq data, it becomes increasingly difficult to make appropriate choices from tens of these resources to meet specific research needs. Therefore, it is necessary to conduct a comprehensive review on these resources.

Here, we present a systematic review of the existing computational resources for ribo-seq data. We begin with a brief summary of relevant databases, Web servers and software tools. Next, we compare these resources and discuss their strengths and limitations. Finally, we conclude with a discussion of current computational challenges and future directions on the topic.

Current status and prospects of computational resources

Since ribosome profiling was first described by Weissman and Ingolia in 2009 [6] (Figure 1A), a plethora of computational

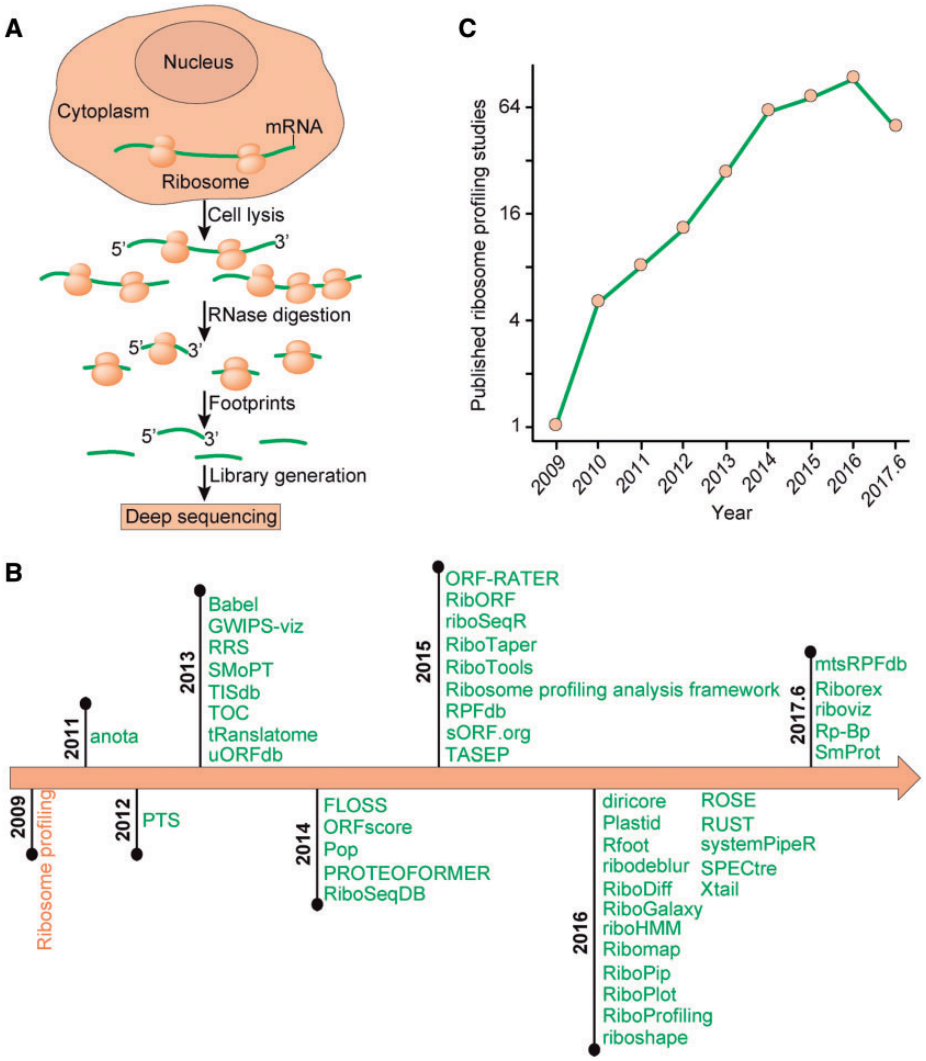

Figure 1. Ribosome profiling and associated computational resources. (**A**) A schematic representation of ribosome profiling. Cycloheximide-treated lysates from cultured cells are digested by nucleases to remove unprotected portions of mRNA. The resulting monosome translating complexes are purified by ultracentrifugation using either a sucrose gradient or a cushion. Ribosome-protected fragments are then recovered, converted to complementary DNA libraries and sequenced. (**B**) History of the development of computational resources related to ribosome profiling. (**C**) The number of ribosome profiling studies per year, based on literature searches with keyword 'ribosome profiling' in PubMed. diricore, differential ribosome codon reading; FLOSS, fragment length organization similarity score; ORF-RATER, ORF regression algorithm for translational evaluation of ribosome-protected fragments (RPFs); PTS, periodicity transition score; TASEP, totally asymmetric exclusion process; ROSE, ribosome stalling estimator; RRS, ribosome release score; RUST, ribo-seq unit step transformation; SMoPT, stochastic model of protein translation; TOC, translated ORF classifier. Pop refers to the statistical model presented by Pop *et al.*

Table 1. Databases for ribo-seq and its related contents

Name	Brief description	Collected data	Organisms	Statistics	Reference
GWIPS-viz	An online genome browser for viewing ribo-seq data	Footprint genome browser; mRNA-seq genome browser	*Arabidopsis thaliana*, *Bacillus subtili*, *Caulobacter crescentus*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Escherichia coli*, human, mouse, *Plasmodium falciparum*, rat, *Streptomyces coelicolor*, *Staphylococcus aureus*, *Trypanosoma brucei brucei*, virus, *Xenopus laevis*, yeast, *Zea mays* and zebrafish	1104 ribo-seq samples from 134 studies	[25]
RPFdb	A database for genome-wide information of translated mRNA generated from ribo-seq data	Expression measurements; footprint genome browser	*Arabidopsis thaliana*, *Caenorhabditis elegans*, *Drosophila*, *Escherichia coli*, human, mouse, yeast and zebrafish	777 ribo-seq samples from 82 studies	[26]
RiboSeqDB	A repository of selected ribo-seq and RNA-seq data on common cell lines	Expression measurements; translation start sites	*Arabidopsis thaliana*, human, mouse and rat	364 ribo-seq samples from 38 studies	[52]
mtsRPFdb	A database for tissue-specific translation information for both coding and noncoding genes in mice	Expression measurements; short ORF annotations	Mouse	5983 uORFs, 3039 dORFs and 1998 sORFs from 8 tissues	/
SmProt	A database of small proteins, especially encoded by noncoding RNAs	Small protein annotations	*Caenorhabditis elegans*, *Escherichia coli*, fruit fly, human, mouse, rat, yeast and zebrafish	29 331 uORFs, 2495 dORFs, 223 077 sORFs and 107 others	[53]
sORFs.org	A repository of sORFs (<100 amino acids) from both coding and noncoding genes identified by ribo-seq	Short ORF annotations	Fruit fly, human and mouse	3 221 654 sORFs from 64 studies	[27]
TISdb	A database for alternative translation initiation in mammalian cells	TISs	Human and mouse	6828 aTISs, 7760 dTISs and 12 147 uTISs	[54]
uORFdb	A comprehensive literature database on eukaryotic uORF biology	uORF-related properties	*Arabidopsis thaliana*, human, mouse, rat, virus, yeast and others	500 uORF-related references	[55]

resources have been developed with the increasing popularity of this technology. A time series plot depicting a historical view of the development of these resources is shown in Figure 1B. Obviously, during the past 2 years, there is a blowout trend. According to their functionality, these resources can be broadly categorized into three major types: database, Web server and software.

Databases

The application of ribo-seq in various organisms is enabling data to be generated at unprecedented scales (Figure 1C). As a result, efficient storage, retrieval and management of these ribo-seq data are crucially important to the research community. Not surprisingly, a variety of databases have emerged, including GWIPS-viz [25], mtsRPFdb, RiboSeqDB [52] and RPFdb [26] (Table 1). Notably, GWIPS-viz (http://gwips.ucc.ie/) is the first designed specifically for ribo-seq data. It provides an intuitive graphical interface of translation in the genomes for visualizing ribo-seq data. Ribo-seq tracks, coupled with RNA sequencing (RNA-seq) tracks, are currently available for 23 genomes from 9 clades, such as mammal, plant and virus. Thus, GWIPS-viz is a powerful resource for users seeking supporting ribo-seq evidence for alternative proteoforms. Subsequently, multiple relevant databases such as RiboSeqDB, RPFdb and mtsRPFdb have been built. The former, RiboSeqDB (http://micro.biouml.org/bioumlweb/), is a repository of selected human, mouse and rat ribo-seq and RNA-seq data on common cell lines. The latter two, RPFdb (http://sysbio.sysu.edu.cn/rpfdb/) and mtsRPFdb (http://sysbio.sysu.edu.cn/mtsRPFdb/), whose development is driven by our group, are intended to facilitate the exploration of these data and understanding translational regulation. RPFdb is a more comprehensive database for genome-wide translation information of protein-coding genes.

Tabulated ribo-seq data from 777 samples of 82 studies for eight species, processed by using a unified pipeline, have been compiled and made readily accessible for further data mining and comparison. mtsRPFdb is the first database providing tissue-specific translation information for both coding and noncoding genes.

The appearance of alternative translation products detected by ribo-seq gives birth to another class of databases, including SmProt [53], sORFs.org [27], TISdb [54] and uORFdb [55] (Table 1). Each is specifically tailored for different contents. For example, SmProt (http://bioinfo.ibp.ac.cn/SmProt/) contains records of small proteins encoded by coding and noncoding RNAs, which provides a user-friendly interface for users to search, browse, submit, blast, export and download these small proteins. sORFs.org (http://sorfs.org/) is a repository of sORFs from coding and noncoding RNAs that allows users to examine individual ORFs or to search based on several criteria for further large-scale studies. TISdb (http://tisdb.human.cornell.edu/) is a repository of alternative translation initiation sites (TISs) in mammalian cells that provides a simple browser interface for query of high-confidence TIS and the associated open reading frames (ORFs). uORFdb (http://www.compgen.uni-muenster.de/tools/uorfdb/) is a comprehensive literature database on eukaryotic uORF biology that provides a rapid, targeted and convenient overview of structural and functional uORF-related properties. Obviously, these databases, different from the ones abovementioned, are designed to fulfill additional requirements, such as facilitating the unveiling of principles governing alternative translation or aiding functional research in the micropeptide field.

Web servers

Ribo-seq, as one of the important applications for next-generation sequencing (NGS), also confronts great challenges in data analysis. The post-analysis of ribo-seq data demands not only appropriate bioinformatics know-how but also suitable computing infrastructures. User-friendly Web server for flexible online analysis and visualization of ribo-seq data is probably a perfect solution, but currently available Web servers are scarce. So far, there have been only two reported Web servers: RiboGalaxy [28] and riboviz [29]. Hereinto, RiboGalaxy (http://ribogalaxy.ucc.ie/) is a Galaxy-based Web server that provides a compact suite of tools specifically tailored for processing, aligning and analyzing ribo-seq and RNA-seq data. It consists of preprocessing tool suite, GWIPS-viz mapping suite, transcriptome mapping suite, riboplot suite, riboSeqR suite [56] and RUST suite [33]. A variety of embedded tools, such as RiboTools [57] and RUST, can be invoked for preprocessing, quality assessment, normalization and differential translation analysis. These tools will be introduced in more detail in the following section. Notably, RiboGalaxy currently implements some common functions but lacks built-in modules for performing more specialized analysis. Continuous expansion of the existing repertoire of toolkits will help to bridge the gap. In contrast, riboviz (www.riboviz.org) is a Web application for the convenient exploration and analysis of ribo-seq data. It consists of a comprehensive and flexible backend analysis pipeline that allows visualization of three-nucleotide periodicity along the ORFs, accumulation of footprints at the start and stop codons, distribution of ribosomal footprint lengths, position-specific nucleotide frequencies of mapped reads and as well as codon-specific densities of ribosomes.

Software tools

Currently, an increasing variety of software tools are available to aid in analyzing and interpreting ribo-seq data. There are literally dozens of such tools designed for different tasks, as summarized in Table 2. Individual description of relevant themes is presented below in detail.

Data preprocessing

In recent years, many efforts have been put forth to develop computational pipelines for preprocessing ribo-seq data. These tools include Plastid [30], PROTEOFORMER [58], ribodeblur [59], RiboPip, RiboPlot [28], RiboProfiling [31], riboSeqR [56], ribosome-profiling-analysis-framework [60], RiboTools [57] and systemPipeR [32]. Some of them only provide limited support for ribo-seq data analysis, such as ribodeblur only for recovering ribosome position signals; RiboPlot only for plotting and outputting read counts of each transcript; ribosome-profiling-analysis-framework for identifying TISs; and RiboTools for detecting translational ambiguities, stop codon readthrough events and codon occupancy of ribosomal A, P and E sites, but most provide a set of pipeline tools for facilitating quality assessment, alignment, read quantification, metagene analysis and others. For example, PROTEOFORMER provides a dedicated pipeline for automatically processing ribo-seq data, including quality control, mapping, assignment of transcripts with evidence of translation, identification of TIS and others. RiboPip provides a Ruby pipeline for preprocessing and analyzing ribo-seq data, including data preparation (such as adaptor clipping, trimming and length selection), removal of unwanted RNAs, alignment, read extraction for different genomic features and differential analysis. The workflow of systemPipeR includes read preprocessing, read alignments, read counting in different genomic features, gene ontology enrichment analysis, differential translation and differential translational efficiency analyses as well as a variety of genome-wide summary plots for visualizing expression trends. Moreover, Plastid also contains some specific analysis for ribo-seq data, such as ribosomal P-site offset determination, read phasing examination and translational efficiency measurement. RiboProfiling can be used to identify ribo-seq read offset, generate transcript and (multi-) codon coverage quantification data, characterize ribosome stalling at sequence motifs and perform principal component analysis as well as graphical representation. riboSeqR allows users to examine read length distributions, parse ribo-seq data from multiple samples, identify triplet periodicity, infer unannotated coding ORFs, plot average and transcript-specific behavior of these data as well as visualize a variety of translational control events. These software together provide sufficient programs to cover almost all the key steps for ribo-seq data preprocessing.

RNase footprint examination

During the biochemical isolation procedure of ribo-seq experiments, ribosomes are often not specifically selected and hence contaminating footprint-sized fragments, such as those from non-ribosomal complexes, should also exist. It is necessary to identify these false readouts of translation. To this end, Rfoot [37], which aims to systematically identify native, non-ribosomal RNA-protein complexes, was developed. Specifically, it allows users to precisely map RNase-protected regions within a variety of types of cytoplasmic and nuclear RNAs, including small nucleolar RNAs, spliceosomal RNAs, microRNAs, transfer RNAs (tRNAs), lncRNAs and UTRs. Through the use of Rfoot, many ribo-seq data can be repurposed for the study of

Table 2. A Summary of software tools with different tasks

Software	Brief description	Reference
Data preprocessing		
Plastid	A python library that is designed to retain the user-friendly pipeline tools for nucleotide-resolution analysis of NGS data, such as ribo-seq and RNA-seq	[30]
PROTEOFORMER	A proteogenomic pipeline for automatically processing ribo-seq data to allow genome-wide visualization of ribosome occupancy and delineating *in vivo* proteoforms to build an optimal protein sequence search database for peptide to Mass spectrometry/mass spectrometry (MS/MS) matching	[58]
ribodeblur	A python package for estimating the ribosome A-site position signals from ribo-seq data	[59]
RiboPip	A Ruby-based alignment and analysis pipeline for Ribo-seq and RNA-seq data	/
RiboPlot	A python package that provides riboplot and ribocount programs to plot and output read counts of ribo-seq data	[28]
RiboProfiling	An R/Bioconductor package that provides a full pipeline to cover all key steps for facilitating quality assessment and quantification of ribo-seq experiments	[31]
riboSeqR	An R/Bioconductor package that provides a set of programs for processing, analyzing and displaying ribo-seq data	[56]
ribosome-profiling-analysis-framework	A dedicated pipeline for identifying TISs	[60]
RiboTools	A Galaxy toolbox for the analysis of ribo-seq data that is dedicated to detecting translational ambiguities, stop codon readthrough events and codon occupancy of ribosomal A, P and E-sites	[57]
systemPipeR	An R/Bioconductor package for building and running automated end-to-end analysis workflows for a wide range of NGS applications, including ribo-seq	[32]
RNase footprint examination		
Rfoot	A computational pipeline for identifying RNA regions protected by nonribosomal RNA–protein complexes	[37]
Data normalization		
RUST	A smoothing transformation-based approach for ribo-seq normalization	[33]
Isoform-level footprint estimation		
Ribomap	An isoform-level ribosome occupancy estimation approach for quantifying isoform-level ribosome profiles	[38]
riboshape	A sparse regression model for predicting ribosome footprint profile shapes from transcript sequences	[39]
Translation or synthesis rate inference		
Pop	A flow conservation-based model for extracting codon translation rates and protein synthesis rates	[40]
SMoPT	A continuous-time, discrete-state Markov model for studying the dynamics of protein translation	[41]
TASEP	A totally asymmetric simple exclusion process-based approach for inferring translation kinetics	[42]
Ribosome stalling event prediction		
ROSE	A deep learning-based framework for predicting ribosome stalling events	[43]
ORF discovery and annotation		
FLOSS	A fragment length organization similarity metric for classifying the translation status of individual transcripts and subregions	[15]
ORF-RATER	A regression-based approach for annotating protein-coding sequences	[45]
ORFscore	A metric for exploring high-resolution footprinting with ribosome phasing to identify actively translated sORFs	[14]
PTS	A periodicity transition metric for detecting dual-coding regions by spotting frame preference changes	[46]
riboHMM	A mixed hidden Markov model-based approach for inferring translated sequences	[47]
RibORF	A support vector machine-based classifier for identifying actively translated ORFs	[48]
RiboTaper	A multitaper spectral-based approach for detecting actively translated ORFs	[50]
Rp-Bp	An unsupervised Bayesian approach for predicting translated ORFs	[61]
RRS	A metric for measuring ribosome release at *bona fide* stop codons to detect translation through ORFs	[49]
SPECtre	A spectral coherence-based classifier for identifying regions of active translation	[51]
TOC	A random forest classifier for distinguishing ORFs in annotated 5′ leaders, coding sequences (CDSs) and 3′ trailers	[13]
Differential translation detection		
anota	An R/Bioconductor package for applying per-gene analysis of partial variance coupled with variance shrinkage to identify differential translation	[34]
Babel	An errors-in-variables regression model-based framework for inferring genes with changes in translational regulation within cells and between conditions	[35]

Continued

Table 2. (continued)

Software	Brief description	Reference
diricore	A procedure for differential ribosome measurements of codon reading	[62]
RiboDiff	A generalized linear model-based framework for detecting genes with changes in translation efficiency across experimental conditions	[63]
Riborex	A generalized linear model-based tool for identifying genes exhibiting differential translation	[64]
tRanslatome	A complete platform for the analysis of differential profiles coming from transcriptome, translatome and proteome studies	[65]
Xtail	An analysis pipeline tailored for ribo-seq data to quantify the magnitudes and statistical significances of differential translations at the genome-wide scale	[36]

non-ribosomal complexes transcriptome-wide. For example, Rfoot analysis of ribo-seq data from two isogenic human cancer cell models revealed that 11.3% of the sequencing reads are derived from non-ribosomal complexes.

Data normalization

Accurate measurements of gene expression will depend heavily on a complicated set of reagents and hardware, along with highly trained personnel. When certain conditions vary during the course of experiments, measurements will have qualitatively different behaviors. Therefore, systematic artifacts in ribo-seq measurements need to be adjusted so that different samples can be more appropriately compared. RUST [33], which is based on a simple smoothing transformation of ribosome footprint densities into a binary step unit function, is designed specifically for ribo-seq data normalization. There are several implementations of RUST that are used to search for synergistic effects between adjacent amino acids, to record the Pearson's and Spearman's correlation coefficient between the observed and to predicted footprint densities for individual transcripts, and plot the observed and predicted footprint densities. As a standard step in the analysis of NGS data, normalization has proven essential to ensure accurate inference of expression levels, but each approach has trade-offs and makes different assumptions about the data. When these assumptions are violated, the normalization may even exacerbate technical artifacts.

Isoform-level footprint estimation

The short-read lengths produced by ribo-seq severely limit our ability to accurately characterize expression levels for alternatively spliced gene isoforms. Still, it remains challenging even for single-molecule real-time sequencing technology. Nonetheless, the efforts on formulating the isoform abundance estimation never stop. Ribomap [38] and riboshape [39] are just two examples that are dedicated to quantify isoform-level ribosome profiles. Using estimated transcript abundance of the candidate locations, Ribomap assigns ribo-seq reads to genomic locations. It can deal with both types of ambiguous mappings caused by either repetitive sequences along the genome or alternative splicing. riboshape is based on kernel smoothing to construct predictive features and builds a sparse model to infer isoform-specific ribosome footprints. In addition, it can also be used to design transcripts with fast translation speeds and to discover unknown modulation during translation. Taken together, these tools may serve as a starting point for downstream analysis of translational regulation from ribo-seq data.

Translation or synthesis rate inference

A critical consideration when interpreting ribo-seq data is to infer the rate of protein synthesis and at which each codon is translated (codon translation rate or elongation rate). There have been several such attempts, including Pop [40], SMoPT [41] and TASEP [42]. Pop is a statistical model for extracting codon translation rates and protein synthesis rates on individual transcripts from ribo-seq data. It allows users to identify causality in regulation of translation and to characterize the features associated with efficient elongation and translation. SMoPT uses a continuous-time, discrete-state Markov model to study translation dynamics for the entire transcriptome. It works on the following basic assumptions: (1) a fixed quantity of ribosomes, tRNA molecules and mRNA molecules in a cell; (2) gene-specific initiation probabilities; and (3) faster tRNA charging. Notably, SMoPT keeps track of every tRNA, mRNA and ribosome molecule in the cell simultaneously. TASEP applies two different fitting TASEP models ($TASEP^{init}$ and $TASEP^{elong}$) to simultaneously infer per-codon translation elongation and per-gene translation initiation rates under the assumption that translation elongation rates are not influenced by the codon context. It should be noted that the assumptions of these approaches are appropriate for a wide range of conditions, but it will not always be the case. For example, transiently translational pausing and abortion under stress conditions is one of the known exceptions [66].

Ribosome stalling event prediction

Substantial progress has been made in our understanding of ribosome stalling event during translation, but the impact of ribosome stalling on protein synthesis is variable, and the underlying reasons for this are largely unclear. To achieve a full understanding, it requires an approach to systematically characterize ribosome stalling events. The first attempt is ROSE [43] that formalizes the ribosome stalling modeling problem into a classification task and predicts ribosome stalling using a deep convolutional neural network with encoded sequence features. To capture the intrinsic contextual features of ribosome stalling, it relies on several motif detectors to scan the input sequence and integrates those stalling relevant motifs. Its predictions were shown to generally correlate with diverse putative regulatory factors of ribosome stalling, such as codon usage bias, tRNA adaptation, N6-methyladenosine modification, mRNA secondary structure and protein-nucleotide binding. Although the determination of the exact positions of stalling still needs to be improved, ROSE offers an opportunity to understand translation-related phenomena.

ORF discovery and annotation

Ribosome profiling has led to the discovery of pervasive translation of ORFs, including previously non-annotated ones,

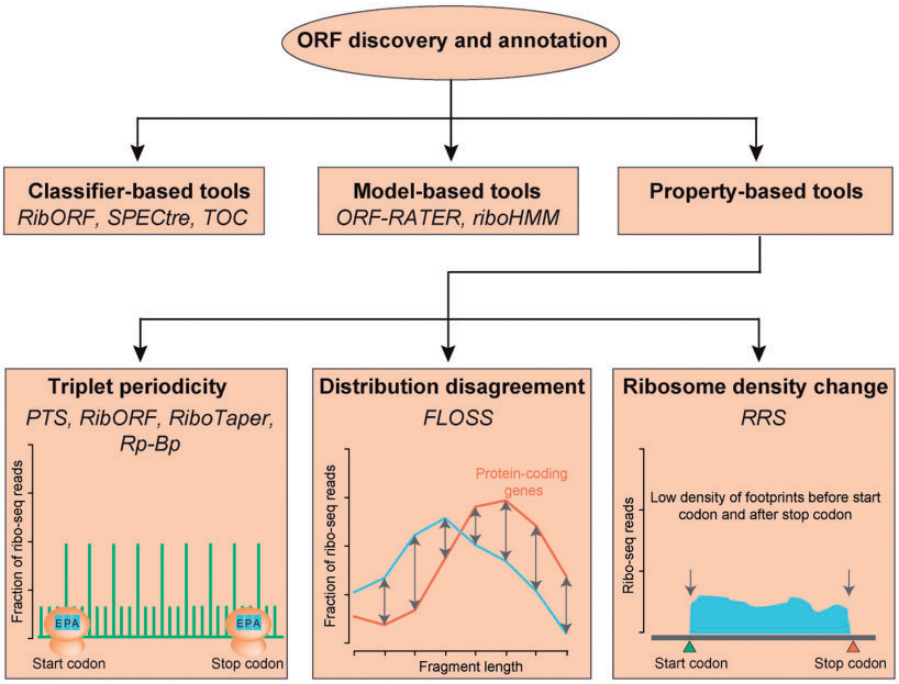

Figure 2. Classification of ORF discovery and annotation. The property-based tools can be further classified into three types: triplet periodicity, distribution disagreement and ribosome density change, shown by simplified diagrams on the figure below.

indicating that translation is far more pervasive than anticipated. To discover and characterize active translation events, multiple metrics and approaches have been proposed, including FLOSS [15], ORF-RATER [45], ORFscore [14], PTS [46], riboHMM [47], RibORF [48], RiboTaper [50], Rp-Bp [61], RRS [49], SPECtre [51] and TOC [13]. These tools can be classified into three types: classifier-based, model-based and property-based (Figure 2). TOC, RibORF and SPECtre extract characteristics from annotated protein-coding transcripts to train classifier, and which is then used to label translated and untranslated ORFs. Specifically, TOC integrates four features, including translational efficiency, inside-versus-outside, fraction length and disengagement score, to distinguish ORFs in annotated 5′ leaders, CDSs and 3′ trailers. In contrast, RibORF identifies translated ORFs based on the evaluation of phasing parameters obtained from canonical protein-coding genes. SPECtre leverages the overall triplet periodicity property of ribo-seq data to detect active translation by calculating the spectral coherence over sliding windows across a given region. ORF-RATER and riboHMM are model-based approaches. ORF-RATER identifies translated ORFs with a linear regression model by comparing the patterns of ribosome occupancy to those of coding ORFs. riboHMM infers translated coding sequences with a mixture of hidden Markov models by leveraging both the total abundance and triplet periodicity of ribo-seq reads. Notably, most of current tools are based on translation properties, and hence can be grouped to property-based approaches. For example, FLOSS exploits the magnitude of disagreement between the overall size distributions of fragments derived from 80 S footprints and non-ribosomal sources; RRS exploits the changes in ribosome density to detect translated ORFs; and others such as PTS, RibORF, RiboTaper and Rp-Bp exploit the triplet periodicity of translating ribosome movement on the transcript. Overall, these tools will greatly enhance our knowledge about protein-coding potential of the genome.

Differential translation detection

The unique characteristics of ribo-seq data impel the development of a rich variety of new algorithms for differential translation analysis. These tools include anota [34], Babel [35], diricore [62], RiboDiff [63], Riborex [64], tRanslatome [65] and Xtail [36]. To identify differential translation, anota applies per-gene analysis of partial variance coupled with variance shrinkage. Babel uses an errors-in-variables regression model. RiboDiff and Riborex use a generalized linear model. Xtail establishes fold-change probability distributions for translational efficiencies and a joint probability matrix which is used to assess gene-specific *P*-values. Moreover, to estimate the background variations and statistical significance properly, each approach makes some important restrictive assumptions. For example, anota makes several model assumptions: highly influential data points, a common slope between sample classes, homoscedasticity of residuals and normally distributed for residuals per gene. diricore assumes that global occupancy of ribosome positions in the cytosol can be used as a cellular detector of alterations in the availability of amino acid for protein synthesis. RiboDiff assumes that transcription and translation are successive cellular processes, and their abundances are linearly related. Xtail assumes that translation and transcription do not coordinate the response to a specific experimental condition, when a gene undergoes translational deregulation. Although these approaches differ in details of their assumptions and model implementation, they are all found with the same objective for differential translation problem: devoting to capturing true differential alterations.

Comparison of computational resources

Indubitably, with increasing amount of large-scale ribo-seq data available, more sophisticated analyses will become particularly critical for pushing the field forward. Owing to the number and

Table 3. Database comparison

Feature	GWIPS-viz	mtsRPFdb	RiboSeqDB	RPFdb
Data type	ribo-seq and RNA-seq	ribo-seq and RNA-seq	ribo-seq and RNA-seq	ribo-seq only
Number of studies	134	1	38	45
Data processing	The adaptor sequence or poly-(A) tails trimmed from 3′ ends of reads; three mismatches allowed	The adaptor sequence trimmed from 3′ ends of reads; two mismatches allowed	The adaptor sequence trimmed from 3′ ends of reads; two mismatches allowed	The first 26 nt kept; one mismatch allowed
Aligner	Bowtie	TopHat	TopHat	STAR
Genome browser	UCSC genome browser	Unsupported	Unsupported	Jbrowse
Meta information	Not searchable	Searchable	Searchable	Searchable
Main characteristics	Visualization of ribo-seq and RNA-seq data; comparison of profiles from different studies or organisms	Reads per kilobase per normalized library (RPKN) values for different genomic features; comparison of the expression from different tissues	Read counts for each transcript; ribo-seq translation start sites	RPKM values for different genomic features; visualization of ribo-seq data

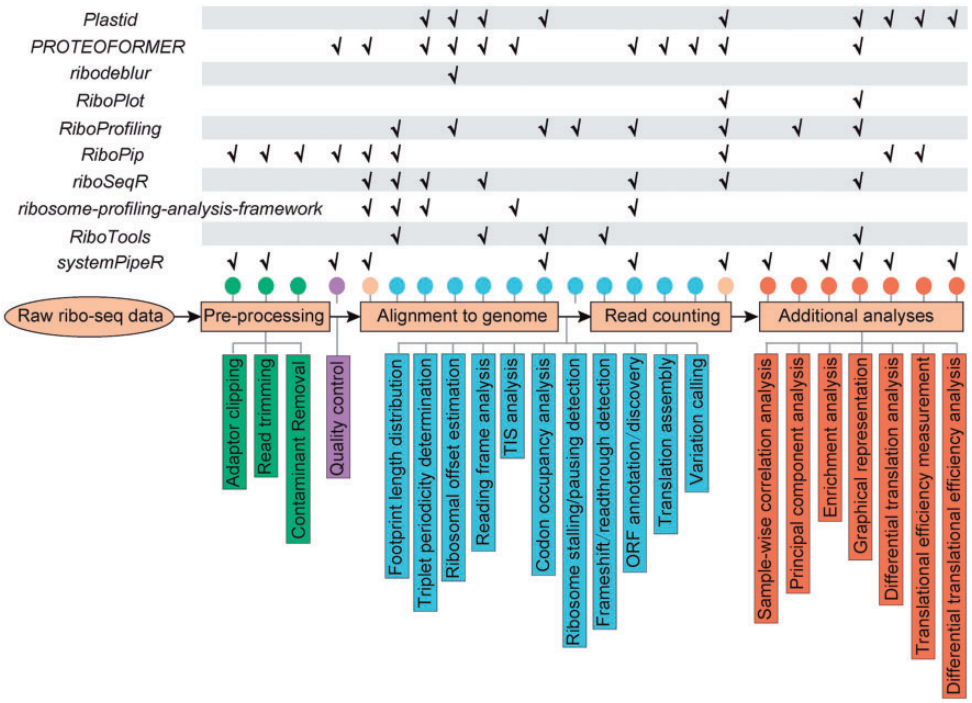

Figure 3. Functionalities of software for processing ribo-seq data. The top-down workflow displays the task information at each step, showing concordance and discordance among different tools. TIS, translation initiation site; ORF, open reading frame.

variety of current resources, it is sometimes difficult to find appropriate resources for a given task. Considering that most of the resources described above are developed to accomplish specific tasks relevant to a research problem at hand and there is no direct comparison between them, more emphasis will be put on the following research areas that are currently the focus of much resource development.

Mining the ribo-seq data for discovery crucially requires managing and organizing large-scale data sets. Multiple databases now exist, and novel databases for alternative translation products continue to appear. We here compare GWIPS-viz, mtsRPFdb, RiboSeqDB and RPFdb, and point out their differences in concreteness (Table 3). The primary difference lies in their objectives, such as GWIPS-viz for readily viewing ribo-seq information, mtsRPFdb for providing tissue-specific translation information and RPFdb for providing translational meta-information. The secondary difference lies in data collection; data preprocessing, including different adapter removal, aligner and mismatch handling; and data presentation. RPFdb is specifically limited to ribo-seq data, while other databases host both ribo-seq and RNA-seq data simultaneously. Notably, GWIPS-viz continuously integrates publicly available ribo-seq data, making it the largest and most comprehensive repository of ribo-seq tracks. Moreover, different from other databases, RPFdb applies different data preprocessing by keeping the first 26 nucleotides only. In addition, all except for GWIPS-viz provide a searching interface and present a summary of the results on its search page, so that users can quickly locate the needed information. Given these, the choice can become easier as often only one database will have the features a user would want.

Table 4. A summary of advantages and limitations of different tools for ORF annotation and discovery

Software	Advantages	Limitations
FLOSS	The ability to aid in separating true footprints from background RNAs; allows an easy incorporation into ribo-seq analysis workflow	No stand-alone package is available and must be implemented by the user; specific length distribution and FLOSS cutoff for each individual data set need to be determined empirically
ORF-RATER	Robustly identifies and quantifies *bona fide* and physiologically relevant translation; flexibility of the linear regression model facilitates its expansion to include additional features	Unsuited to the identification of translation events with non-canonical ribosome density, such as programmed frameshifts and stop codon readthrough; intractability for the application of nonlinear genomes
ORFscore	Leverages the periodicity of high-quality ribosome-protected fragments to define translated ORFs independent of the surrounding sequence context	No stand-alone package is available and must be implemented by the user; overlapping ORFs that are translated on different reading frames can be missed
PTS	Increases the size of the informative region, while the effect of false signals generated by isolated RPFs is reduced	Allows dually coded regions to be detected only if the alternative frame has RPF coverage comparable with, or higher than, that of the standard frame; heavily depends on features of dual coding, such as the length of the overlapping region and the density of RPFs
riboHMM	The ability to infer novel translated sequences that may be missed by annotation pipelines that focus on long CDSs (>100 amino acids), conservation-based approaches that require long-term purifying selection or direct proteomics measurements that are biased toward highly expressed, stable proteins	Unable to resolve overlapping translated sequences from multiple coding isoforms of a gene; has a relatively high error rate for identifying the precise TIS
RibORF	Effectively distinguish in-frame ORFs from overlapping off-frame ORFs; can distinguish reads arising from RNAs that are not associated with ribosomes	Manually check the offset distance between 5′ end and ribosomal A-site; manually check whether read distribution shows clear 3-nt periodicity
RiboTaper	Reconstructs the full set of ORFs in annotated coding and non-coding transcripts; general applicability and excellent performance	Requires matched ribo-seq and RNA-seq data; may take multiple days to analyze a single sample; can miss cases of non-AUG starting ORFs
Rp-Bp	The ability to incorporate and propagate uncertainty in the prediction process; automatic Bayesian selection of read lengths and ribosome P-site offsets	Unsuited to compare estimates from two different data sets because of its unnormalized parameter estimates; unable to distinguish overlapping ORFs and identify programmed frameshifts
RRS	Suited to distinguish real translation from non-ribosomal contamination; robustness to other potential sources of bias that lead to higher levels of contaminating non-ribosomal reads on specific classes of RNA	The size of the 3′ UTR have to be larger than the fragment length generated in the ribosome profiling assay; unable to provide a perfect classification for those transcripts that fall into overlapping regions between the distributions for coding and classical noncoding transcripts
SPECtre	The ability to report and visualize signals of periodicity in the context of surrounding genomic features; its ability to chunk experiments and parallelize analyses over multiple threads	Multiple analytical arguments, such as scoring method, the length of the windows, minimum fragments per kilobase of exon per million reads mapped (FPKM) cutoffs and false discovery rate (FDR) thresholds, need to be specified by the user
TOC	Effectively integrates intrinsic transcript information, such as sequence and location of ORFs, and external data, such as ribo-seq and RNA-seq	Unable to define the translated ORF, but rather classify whole transcripts as coding, leader- or trailer-like

The second area focuses on data preprocessing. Data preprocessing is a key step toward understanding of the basis of many biological questions, such as predicting actively translated ORF, inferring translational efficiency or detecting differential translation. To facilitate data preprocessing, more than a dozen of software tools are available. When having too many options, a major challenge is to pick out which tool to use because of each with its own feature sets. The functionality of each tool is illustrated in Figure 3. Frequently, these tools exhibit substantial feature overlap, where similar principles are often adopted. The decision of which tool to use may have little influence on the outcome of the analysis. Nevertheless, some important limitations must also be taken into account when making a choice between these tools. For example, half of these software tools take a BAM format alignment file as input, and PROTEOFORMER, RiboPip, riboSeqR, ribosome-profiling-analysis-framework and systemPipeR start with raw sequence data, supporting the alignment of ribo-seq data. ribosome-profiling-analysis-framework is based on *Mus musculus* mm10 genome assembly and is also suitable for *Homo sapiens* hg19 and hg38 genome assemblies, but it currently does not support other assemblies. In practice, these tools have not gained widespread use, mainly because of the availability of state-of-the-art tools for individual steps. Perhaps, the top-down analysis frameworks with an emphasis on simplicity for users, such as Plastid and systemPipeR, can become more popular.

The third area focuses on ORF discovery and annotation. Our current understanding of the coding potential of genomes has benefited greatly from many recently emerging software tools. However, these tools often face a daunting challenge of distinguishing ORFs from chance in-frame start and stop codons. Given pervasive non-canonical translation in the genome, this makes assignment of protein-coding ORFs even more difficult. To address this challenge, different strategies, such as

the requirement of triplet periodicity and adjustment of imperfect RNase trimming of footprints, were proposed to reduce false-positive predictions. Nevertheless, when interpreting results, the benefits of each tool and computational limitations must also be considered (Table 4). For example, RiboTaper only considers AUG start codons, and thus probably misses the cases of non-AUG starting ORFs. TOC is not used to define the translated ORFs, but rather classify whole transcripts as coding, leader- or trailer-like. Choosing between tools, especially for a basic user without a strong background, often comes down to personal preferences, such as familiar interface or ease of use. To assist the choice, several studies have attempted to provide some general guidelines [14, 50, 51]. For example, comparative analysis of FLOSS, ORFscore, RiboTaper and SPECtre showed that SPECtre conforms with high fidelity to RiboTaper and both outperform FLOSS and ORFscore [51]. RiboTaper seems to reflect the current state of the art for the discovery and annotation of ORFs. Although a thorough comparison between these tools is lacking, the tools integrating the key feature of protein synthesis: the triplet periodicity of translating ribosome movement, as reflected in the profiles of canonical protein-coding regions, and rigorous statistical model should bring us closer to full decryption of *bona fide* translated ORFs.

The fourth area focuses on differential translation detection. Different from differential transcription analysis, the detection of differences in translational efficiency demands specialized statistical approaches, considering that the measurements on translation depend on the translation rate and transcription abundance as well. Extensive efforts have been made for this. Although different tuning models are adopted to obtain optimal performance, they often grapple with technical and biological variations, requiring users to trade-off between false positives and false negatives. Several studies have already been done in this direction for comparison of these approaches. Yi *et al.* [63] showed that RiboDiff on simulated data with large difference between dispersions of between ribo-seq and RNA-seq counts exhibits a superior detection accuracy compared with Babel, but which is less pronounced when the dispersions are more similar. Xiao *et al.* [36] showed that Xtail on both simulated and experimental data sets exhibits high sensitivity with minimal false positives, outperforming the existing tools, including TE, anota, Babel, RiboDiff and baySeq [67], in the accuracy of quantifying differential translations. More recently, Wenzheng *et al.* [64] compared Riborex with a representative set of existing approaches, including Babel, RiboDiff and Xtail, on simulated data, showing that Riborex and Xtail have a slight advantage over Babel and RiboDiff. However, running time evaluation based on four published data sets [68–70] differing in numbers of genes and numbers of replicates showed that Riborex finishes in seconds, and all the other tools take substantially longer, with Xtail requiring >4 h. Collectively, Riborex has accuracy on par with the most accurate existing approaches but is significantly faster. Notably, some tools originally developed for RNA-seq data, such as DESeq2 and edgeR, can potentially be applied by modeling ratios of translational efficiencies for the two conditions. However, to ensure the accuracy of the attained results, it may be informative to run the analyses with more than one software.

Conclusions and perspectives

The scale and complexity of ribo-seq data pose considerable challenges, requiring experimental and computational biologists to apply increasingly more sophisticated computational resources in the storage, analysis and interpretation of these data. Here, we present, to our knowledge, the first review of available computational resources for ribosome profiling. We systematically described the types of software tools that are required at various stages of ribo-seq analysis, and also discussed their distinct functionality and applications to help users choose appropriate tools and resources.

Innovative computational resources underlie many emerging applications of ribosome profiling, which is becoming increasingly important to foster novel discovery in translatome analysis. However, an important prerequisite for successful discovery is that the data itself must be of sufficient quality. Currently, ribo-seq experiment is still technically challenging in some aspects, such as the requirement of a rapid inhibition of translation, an effective nuclease treatment and a rigorous selection of the desired fragments. In general, each experimental step, from cell harvesting to nuclease digestion to library generation, may potentially cause distortions in the data output [9]. Moreover, ribosome footprints are typically short, and hence, the determination of correct alignment position has remained a challenge, even though several tools are available for handling ambiguity in read mapping. How to effectively remove the contaminating footprint-sized readouts is also a considerable challenge. This is particularly important in light of recent findings that the monosome fraction isolated by sucrose sedimentation contains a large quantity of inactive ribosomes that do not engage on mRNAs to direct translation [71]. Besides technological innovation of ribo-seq, novel tools for filtering such readouts will be required to serve as an adjunct control at the preprocessing step. Like all genome-wide expression profiling, ribo-seq data that may be subject to technical and biological variations also need to be adjusted. Relevant study has shown that the best practice is the use of spike-in normalization [72]. Furthermore, comprehensive identification of ORFs has been challenging, but given pervasive actively translated ORFs defined by ribo-seq, deciphering the functions of the newly identified ORFs is now an outstanding challenge. A plethora of chemical biology methods can potentially offer opportunities for the molecular characterization of their functions.

In addition to the availability, software usability, such as the ease of installation and good documentation, has often lagged behind, but its optimization is anticipated because of an increasing recognition of its value to the research community. Further optimization to decrease computational cost and increase statistical efficiency will be required, which can yield significant performance benefits. Additional effort should be made to build comprehensive, flexible and integrated cloud-computing solutions that allow users to specify their needs and to aggregate diverse resources (e.g. private, public or a mix of them) into their analysis workflow. This will enable researchers to enjoy greater benefits from ribosome profiling.

Key Points

- The present study is, to our knowledge, the first comprehensive review of available computational resources for ribosome profiling to assist in the appropriate selection and use of relevant tools.
- These resources spanning a broad spectrum from database to Web server and software have led to a great step forward in the storage, manipulation, analysis, display and interpretation of ribo-seq data.
- The performance and usability comparison provides readers with guidelines on an informed selection of the

most appropriate tool for the problem at hand.
- It also provides an outlook regarding current computational challenges and future directions towards cloud-supported data computing and analytics solutions.

Funding

The National Natural Science Foundation of China (grant number 31471232), Science and Technology Planning Projects of Guangdong Province (grant number 2014B030301040), Joint Research Fund for Overseas Natural Science of China (grant number 3030901001222) and Center for Precision Medicine to Z.X.; the China Postdoctoral Science Foundation (grant number 2015M582459 to H.W.W.).

References

1. Hershey JW, Sonenberg N, Mathews MB. Principles of translational control: an overview. *Cold Spring Harb Perspect Biol* 2012; **4**:a011528.
2. Van Der Kelen K, Beyaert R, Inze D, *et al*. Translational control of eukaryotic gene expression. *Crit Rev Biochem Mol Biol* 2009; **44**:143–68.
3. Calkhoven CF, Muller C, Leutz A. Translational control of gene expression and disease. *Trends Mol Med* 2002;**8**:577–83.
4. Bhat M, Robichaud N, Hulea L, *et al*. Targeting the translation machinery in cancer. *Nat Rev Drug Discov* 2015;**14**:261–78.
5. Grzmil M, Hemmings BA. Translation regulation as a therapeutic target in cancer. *Cancer Res* 2012;**72**:3891–900.
6. Ingolia NT, Ghaemmaghami S, Newman JR, *et al*. Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science* 2009;**324**:218–23.
7. Ingolia NT, Brar GA, Rouskin S, *et al*. The ribosome profiling strategy for monitoring translation *in vivo* by deep sequencing of ribosome-protected mRNA fragments. *Nat Protoc* 2012;**7**:1534–50.
8. Ingolia NT. Ribosome profiling: new views of translation, from single codons to genome scale. *Nat Rev Genet* 2014;**15**: 205–13.
9. Brar GA, Weissman JS. Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nat Rev Mol Cell Biol* 2015;**16**:651–64.
10. Ingolia NT. Ribosome footprint profiling of translation throughout the genome. *Cell* 2016;**165**:22–33.
11. Andreev DE, O'Connor PB, Loughran G, *et al*. Insights into the mechanisms of eukaryotic translation gained with ribosome profiling. *Nucleic Acids Res* 2017;**45**:513–26.
12. Gobet C, Naef F. Ribosome profiling and dynamic regulation of translation in mammals. *Curr Opin Genet Dev* 2017;**43**:120–7.
13. Chew GL, Pauli A, Rinn JL, *et al*. Ribosome profiling reveals resemblance between long non-coding RNAs and 5' leaders of coding RNAs. *Development* 2013;**140**:2828–34.
14. Bazzini AA, Johnstone TG, Christiano R, *et al*. Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO J* 2014;**33**:981–93.
15. Ingolia NT, Brar GA, Stern-Ginossar N, *et al*. Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell Rep* 2014;**8**:1365–79.
16. Ruiz-Orera J, Messeguer X, Subirana JA, *et al*. Long non-coding RNAs as a source of new peptides. *Elife* 2014;**3**:e03523.
17. Mackowiak SD, Zauber H, Bielow C, *et al*. Extensive identification and analysis of conserved small ORFs in animals. *Genome Biol* 2015;**16**:179.
18. Wang H, Wang Y, Xie S, *et al*. Global and cell-type specific properties of lincRNAs with ribosome occupancy. *Nucleic Acids Res* 2017;**45**:2786–96.
19. Atger F, Gobet C, Marquis J, *et al*. Circadian and feeding rhythms differentially affect rhythmic mRNA transcription and translation in mouse liver. *Proc Natl Acad Sci USA* 2015; **112**:E6579–88.
20. Park JE, Yi H, Kim Y, *et al*. Regulation of Poly(A) tail and translation during the somatic cell cycle. *Mol Cell* 2016;**62**:462–71.
21. Spriggs KA, Bushell M, Willis AE. Translational regulation of gene expression during conditions of cell stress. *Mol Cell* 2010; **40**:228–37.
22. Jackson RJ, Hellen CU, Pestova TV. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* 2010;**11**:113–27.
23. Mitchell SF, Parker R. Modifications on translation initiation. *Cell* 2015;**163**:796–8.
24. Hinnebusch AG, Lorsch JR. The mechanism of eukaryotic translation initiation: new insights and challenges. *Cold Spring Harb Perspect Biol* 2012;**4**:a011544.
25. Michel AM, Fox G, M Kiran A, *et al*. GWIPS-viz: development of a ribo-seq genome browser. *Nucleic Acids Res* 2014;**42**:D859–64.
26. Xie SQ, Nie P, Wang Y, *et al*. RPFdb: a database for genome wide information of translated mRNA generated from ribosome profiling. *Nucleic Acids Res* 2016;**44**:D254–8.
27. Olexiouk V, Crappe J, Verbruggen S, *et al*. sORFs.org: a repository of small ORFs identified by ribosome profiling. *Nucleic Acids Res* 2016;**44**:D324–9.
28. Michel AM, Mullan JP, Velayudhan V, *et al*. RiboGalaxy: a browser based platform for the alignment, analysis and visualization of ribosome profiling data. *RNA Biol* 2016;**13**:316–19.
29. Carja O, Xing T, Plotkin J, *et al*. riboviz: analysis and visualization of ribosome profiling datasets. *bioRxiv* 2017. doi: https://doi.org/10.1101/100032.
30. Dunn JG, Weissman JS. Plastid: nucleotide-resolution analysis of next-generation sequencing and genomics data. *BMC Genomics* 2016;**17**:958.
31. Popa A, Lebrigand K, Paquet A, *et al*. RiboProfiling: a bioconductor package for standard Ribo-seq pipeline processing. *F1000Res* 2016;**5**:1309.
32. Backman TWH, Girke T. systemPipeR: NGS workflow and report generation environment. *BMC Bioinformatics* 2016;**17**:388.
33. O'Connor PB, Andreev DE, Baranov PV. Comparative survey of the relative impact of mRNA features on local ribosome profiling read density. *Nat Commun* 2016;**7**:12915.
34. Larsson O, Sonenberg N, Nadon R. anota: analysis of differential translation in genome-wide studies. *Bioinformatics* 2011; **27**:1440–1.
35. Olshen AB, Hsieh AC, Stumpf CR, *et al*. Assessing gene-level translational control from ribosome profiling. *Bioinformatics* 2013;**29**:2995–3002.
36. Xiao Z, Zou Q, Liu Y, *et al*. Genome-wide assessment of differential translations with ribosome profiling data. *Nat Commun* 2016;**7**:11194.
37. Ji Z, Song R, Huang H, *et al*. Transcriptome-scale RNase-footprinting of RNA-protein complexes. *Nat Biotechnol* 2016; **34**:410–13.
38. Wang H, McManus J, Kingsford C. Isoform-level ribosome occupancy estimation guided by transcript abundance with Ribomap. *Bioinformatics* 2016;**32**:1880–2.
39. Liu TY, Song YS. Prediction of ribosome footprint profile shapes from transcript sequences. *Bioinformatics* 2016;**32**: i183–91.

40. Pop C, Rouskin S, Ingolia NT, *et al*. Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation. *Mol Syst Biol* 2014;**10**:770.
41. Shah P, Ding Y, Niemczyk M, *et al*. Rate-limiting steps in yeast protein translation. *Cell* 2013;**153**:1589–601.
42. Gritsenko AA, Hulsman M, Reinders MJ, *et al*. Unbiased quantitative models of protein translation derived from ribosome profiling data. *PLoS Comput Biol* 2015;**11**:e1004336.
43. Zhang S, Hu H, Zhou J, *et al*. ROSE: a deep learning based framework for predicting ribosome stalling. *bioRxiv* 2016. doi: https://doi.org/10.1101/067108.
44. Zupanic A, Meplan C, Grellscheid SN, *et al*. Detecting translational regulation by change point analysis of ribosome profiling data sets. *RNA* 2014;**20**:1507–18.
45. Fields AP, Rodriguez EH, Jovanovic M, *et al*. A regression-based analysis of ribosome-profiling data reveals a conserved complexity to mammalian translation. *Mol Cell* 2015;**60**:816–27.
46. Michel AM, Choudhury KR, Firth AE, *et al*. Observation of dually decoded regions of the human genome using ribosome profiling data. *Genome Res* 2012;**22**:2219–29.
47. Raj A, Wang SH, Shim H, *et al*. Thousands of novel translated open reading frames in humans inferred by ribosome footprint profiling. *Elife* 2016;**5**:e13328.
48. Ji Z, Song R, Regev A, *et al*. Many lncRNAs, 5'UTRs, and pseudogenes are translated and some are likely to express functional proteins. *Elife* 2015;**4**:e08890.
49. Guttman M, Russell P, Ingolia NT, *et al*. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* 2013;**154**:240–51.
50. Calviello L, Mukherjee N, Wyler E, *et al*. Detecting actively translated open reading frames in ribosome profiling data. *Nat Methods* 2016;**13**:165–70.
51. Chun SY, Rodriguez CM, Todd PK, *et al*. SPECtre: a spectral coherence–based classifier of actively translated transcripts from ribosome profiling sequence data. *BMC Bioinformatics* 2016;**17**:482.
52. Sharipov RN, Yevshin IS, Kondrakhin YV, *et al*. RiboSeqDB-a repository of selected human and mouse ribosome footprint and rna-seq data. *Virtual Biology* 2014;**1**:37–46.
53. Hao Y, Zhang L, Niu Y, *et al*. SmProt: a database of small proteins encoded by annotated coding and non-coding RNA loci. *Brief Bioinform* 2017, in press. pii:bbx005.
54. Wan J, Qian SB. TISdb: a database for alternative translation initiation in mammalian cells. *Nucleic Acids Res* 2014;**42**: D845–50.
55. Wethmar K, Barbosa-Silva A, Andrade-Navarro MA, *et al*. uORFdb–a comprehensive literature database on eukaryotic uORF biology. *Nucleic Acids Res* 2014;**42**:D60–7.
56. Chung BY, Hardcastle TJ, Jones JD, *et al*. The use of duplex-specific nuclease in ribosome profiling and a user-friendly software package for Ribo-seq data analysis. *RNA* 2015;**21**: 1731–45.
57. Legendre R, Baudin-Baillieu A, Hatin I, *et al*. RiboTools: a Galaxy toolbox for qualitative ribosome profiling analysis. *Bioinformatics* 2015;**31**:2586–8.
58. Crappe J, Ndah E, Koch A, *et al*. PROTEOFORMER: deep proteome coverage through ribosome profiling and MS integration. *Nucleic Acids Res* 2015;**43**:e29.
59. Wang H, McManus J, Kingsford C. Accurate recovery of ribosome positions reveals slow translation of wobble-pairing codons in yeast. *J Comput Biol* 2017;**23**:486–500.
60. de Klerk E, Fokkema IF, Thiadens KA, *et al*. Assessing the translational landscape of myogenic differentiation by ribosome profiling. *Nucleic Acids Res* 2015;**43**:4408–28.
61. Malone B, Atanassov I, Aeschimann F, *et al*. Bayesian prediction of RNA translation from ribosome profiling. *Nucleic Acids Res* 2017;**45**:2960–72.
62. Loayza-Puch F, Rooijers K, Buil LC, *et al*. Tumour-specific proline vulnerability uncovered by differential ribosome codon reading. *Nature* 2016;**530**:490–4.
63. Zhong Y, Karaletsos T, Drewe P, *et al*. RiboDiff: detecting changes of mRNA translation efficiency from ribosome footprints. *Bioinformatics* 2017;**33**:139–41.
64. Li W, Wang W, Uren PJ, *et al*. Riborex: fast and flexible identification of differential translation from Ribo-seq data. *Bioinformatics* 2017;**33**:1735–7.
65. Tebaldi T, Dassi E, Kostoska G, *et al*. tRanslatome: an R/Bioconductor package to portray translational control. *Bioinformatics* 2014;**30**:289–91.
66. Subramaniam AR, Zid BM, O'Shea EK. An integrated approach reveals regulatory controls on bacterial translation elongation. *Cell* 2014;**159**:1200–11.
67. Hardcastle TJ, Kelly KA. baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinformatics* 2010;**11**:422.
68. Bennett CG, Riemondy K, Chapnick DA, *et al*. Genome-wide analysis of Musashi-2 targets reveals novel functions in governing epithelial cell migration. *Nucleic Acids Res* 2016;**44**: 3788–800.
69. Zur H, Aviner R, Tuller T. Complementary post transcriptional regulatory information is detected by PUNCH-P and ribosome profiling. *Sci Rep* 2016;**6**:21635.
70. Hsieh AC, Liu Y, Edlind MP, *et al*. The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature* 2012;**485**:55–61.
71. Liu B, Qian SB. Characterizing inactive ribosomes in translational profiling. *Translation* 2016;**4**:e1138018.
72. Loven J, Orlando DA, Sigova AA, *et al*. Revisiting global gene expression analysis. *Cell* 2012;**151**:476–82.