

RESEARCH ARTICLE

riboWaltz: Optimization of ribosome P-site positioning in ribosome profiling data

Fabio Lauria¹*, Toma Tebaldi²✉, Paola Bernabò¹, Ewout J. N. Groen^{3,4}, Thomas H. Gillingwater^{3,4}, Gabriella Viero^{1*}

1 Institute of Biophysics, CNR Unit at Trento, Trento, Italy, **2** Centre for Integrative Biology, University of Trento, Trento, Italy, **3** Euan MacDonald Centre for Motor Neurone Disease Research, University of Edinburgh, Edinburgh, United Kingdom, **4** Centre for Integrative Physiology, University of Edinburgh, Edinburgh, United Kingdom

✉ These authors contributed equally to this work.

✉ Current address: Yale Cancer Center, Yale University School of Medicine, New Haven, CT, United States of America

* fabio.lauria@unitn.it (FL); gabriella.viero@cnr.it (GV)



OPEN ACCESS

Citation: Lauria F, Tebaldi T, Bernabò P, Groen E, Gillingwater TH, Viero G (2018) riboWaltz: Optimization of ribosome P-site positioning in ribosome profiling data. *PLoS Comput Biol* 14(8): e1006169. <https://doi.org/10.1371/journal.pcbi.1006169>

Editor: Aaron E. Darling, University of Technology Sydney, AUSTRALIA

Received: November 30, 2017

Accepted: April 30, 2018

Published: August 13, 2018

Copyright: © 2018 Lauria et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Two NGS datasets produced by the authors are available in GEO under accessions GSE102318 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE102318>) and GSE111866 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE111866>). All the remaining data used in the manuscript were obtained from GEO and have been already presented in the papers cited throughout the manuscript. The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper and its Supporting Information

Abstract

Ribosome profiling is a powerful technique used to study translation at the genome-wide level, generating unique information concerning ribosome positions along RNAs. Optimal localization of ribosomes requires the proper identification of the ribosome P-site in each ribosome protected fragment, a crucial step to determine the trinucleotide periodicity of translating ribosomes, and draw correct conclusions concerning where ribosomes are located. To determine the P-site within ribosome footprints at nucleotide resolution, the precise estimation of its offset with respect to the protected fragment is necessary. Here we present riboWaltz, an R package for calculation of optimal P-site offsets, diagnostic analysis and visual inspection of ribosome profiling data. Compared to existing tools, riboWaltz shows improved accuracies for P-site estimation and neat ribosome positioning in multiple case studies. riboWaltz was implemented in R and is available as an R package at <https://github.com/LabTranslationalArchitectomics/RiboWaltz>.

This is a *PLOS Computational Biology* Software paper.

Introduction

Ribosome profiling (RiboSeq) is an experimental technique used to investigate translation at single nucleotide resolution and genome-wide scale [1,2], through the identification of short RNA fragments protected by ribosomes from nuclease digestion [3,4]. The last few years have witnessed a rapid adoption of this technique and a consequent explosion in the volume of RiboSeq data [5,6]. In parallel, a number of dedicated computational algorithms were developed for extracting transcript-level information, including unannotated open reading frames (ORFs) [7–10], novel translation initiation sites and differentially translated genes [11,12], as well as positional

files, or available at the following GitHub repository: <https://github.com/LabTranslationalArchitectomics/riboWaltz>.

Funding: This work was supported by the Autonomous Province of Trento through the Axonomix project (to FL, TT, PB and GV), and the Wellcome Trust (106098/Z/14/Z; to EJNG and THG). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

information describing fluxes of ribosomes along the RNA at sub-codon resolution [13–15] and conformational changes in ribosomes during the elongation step of translation [16].

Much of this information relies on the ability to determine the exact localization of the P-site, i.e. the site holding the t-RNA associated to the growing polypeptide chain during translation, within ribosome protected fragments (RPF, also called reads hereinafter, following the notation adopted by [1]). This position can be specified by the distance of the P-site from both 5' and 3' ends of the reads, the so-called P-site Offset, PO (Fig 1A).

Accurate determination of the PO is a crucial step to verify the trinucleotide periodicity of ribosomes along coding regions [1,17], derive reliable translation initiation and elongation rates [18,19], accurately estimate codon usage bias and translation pauses [15,20–23], and reveal novel translated regions in known protein coding transcripts or ncRNAs [8,24,25].

Typically, the PO is defined as a constant number of nucleotides from either the 3' or 5' end of reads, independently from their length (Fig 1A) [26]. This approach may lead to an inaccurate detection of the P-site's position owing to potential offset variations associated with the length of the reads due to different ribosome conformations [16], non-translating ribosomes [27], nuclease digestion biases [15] and sequencing biases [2]. This problem is frequently resolved by selecting subsets of reads with defined length [28,29]. As such, this procedure removes from the analysis reads that are potentially derived from fragments associated to alternative conformations of the ribosome [30,31] and characterized by shorter or longer lengths [16]. Recently, computational tools have been developed to assist with RiboSeq analysis and P-site localization; examples are Plastid [32] and RiboProfiling [33]. Both tools compute the PO after stratifying the reads in bins, according to their length. However, each bin is treated independently, possibly leading to excessive variability of the offsets across bins.

Here, we describe the development of riboWaltz, an R package aimed at computing the PO for all reads from single or multiple RiboSeq samples. Taking advantage of a two-step algorithm, where offset information is passed through populations of reads with different length to maximize the offset coherence, riboWaltz computes with extraordinary precision the PO and shows higher accuracy and specificity of P-site positions than the other methods. riboWaltz provides the user with a variety of graphical representations, laying the foundations for further accurate RiboSeq analyses and better interpretation of positional information.

Design and implementation

Input acquisition and processing

riboWaltz is an R package that requires two mandatory input data files: 1) alignment files, in BAM format or as GAlignments objects in R, ideally from transcriptome alignments of RiboSeq reads, and; 2) transcript annotation files, in GTF/GFF3 format or provided as TxDb objects in R. Alternatively, annotation can also be provided as a tab separated text file containing minimal transcript annotation: the length of the transcripts and of their annotated coding sequences and UTRs (Fig 1B). Optionally, a third file containing transcript sequence information in FASTA format can be provided as input to perform P-site specific codon sequence analysis. The user is also free to specify a genome build and the corresponding BSgenome object in R will be used for sequence retrieval (Fig 1B).

riboWaltz acquires BAM files and converts them into BED files utilizing the *bamtobed* function of the BEDTools suite [34].

Selection of read lengths

Different lengths of RPFs may derive from alternative ribosome conformations [16,30,31]. Therefore, the researcher should be free to modify the tolerance for the selection of the read

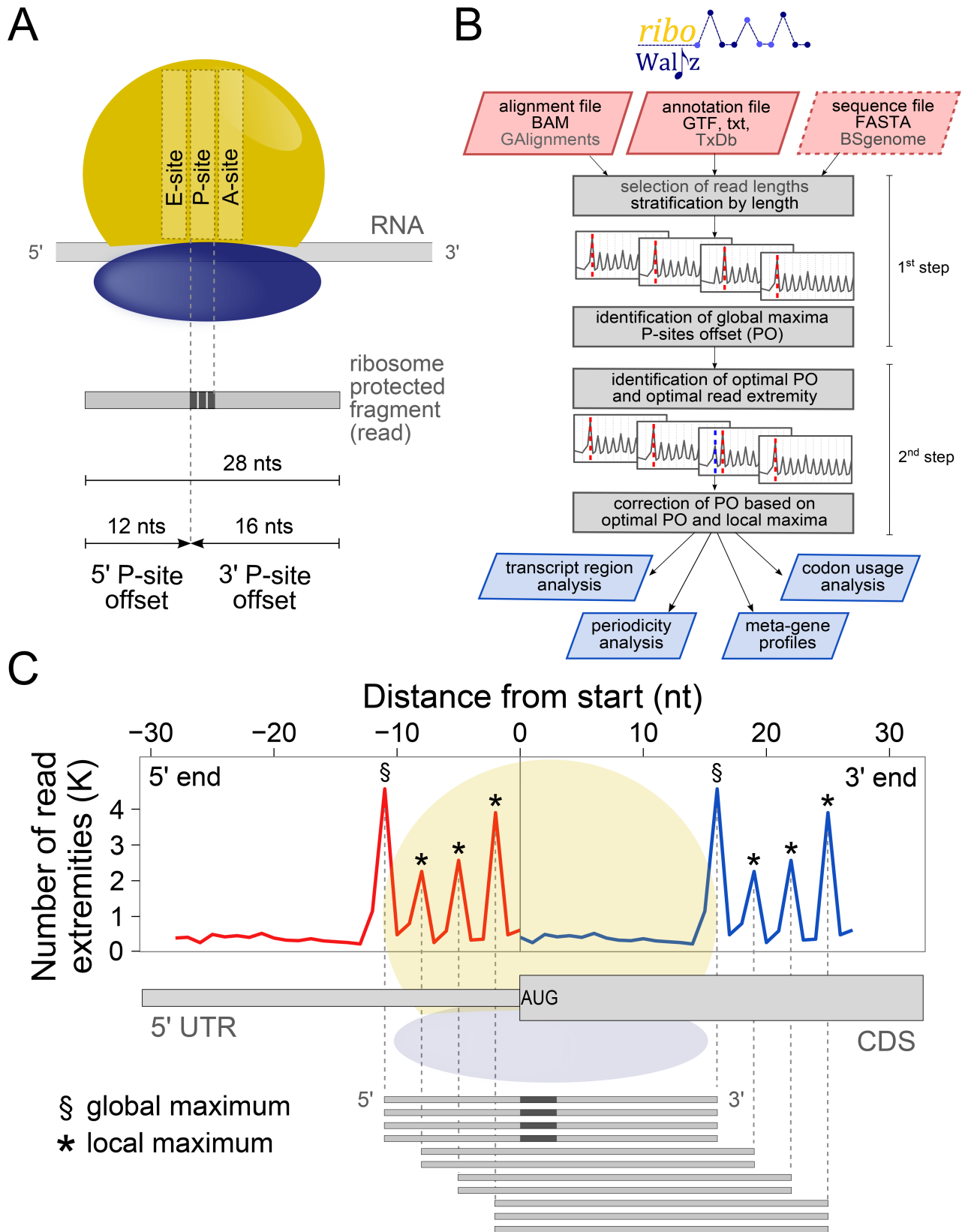


Fig 1. (A) Schematic representation of the P-site offset. Two offsets can be defined, one for each extremity of the read. (B) Flowchart representing the basic steps of riboWaltz, the input requirements and the outputs. (C) An example of ribosome occupancy profile obtained from the alignment of the 5' and the 3' end of reads around the start codon (reads length, 28 nucleotides) is superimposed to the schematic representations of a transcript, a ribosome positioned on the translation initiation site (TIS) and a set of reads used for generating the profiles.

<https://doi.org/10.1371/journal.pcbi.1006169.g001>

length according to the aim of the experiment. For this reason, riboWaltz has multiple options for treating read lengths: i) all read lengths are included in the analysis (all-inclusive mode) ii) only read lengths specified by the user are included (manual mode); iii) only read lengths satisfying a periodicity threshold are included in the analysis (periodicity threshold mode). The user can change the desired threshold (the default is 50%). This mode enables the removal of all the reads without periodicity, similarly to other approaches [10,35].

Identification of the P-site position

The identification of the P-site, defined by the position of its first nucleotide within the reads, is based on reads aligning across annotated translation initiation sites (TIS or start codon), as proposed by [1]. It is known that the P-site of the reads protected by ribosomes in translation initiation corresponds exactly to the start codon. Thus the P-site offset can be defined as the distance between the extremities of the reads and the start codon itself. After the identification of the P-site for the reads aligning on the TIS, the POs corresponding to each length are assigned to each read of the dataset.

riboWaltz specifically infers the PO in two-steps. First, riboWaltz groups the reads mapping on the TIS according to their length. Each group of reads with a specific length (L) corresponds to a bin. To avoid biases in PO calculation, reads whose extremities are too close to the start codon (9 nucleotides by default) are discarded from the computation of the PO. This parameter, called “flanking length” (FL), can be set by the user. Next, for each length bin, riboWaltz generates the occupancy profiles of read extremities, i.e. the number of 5' and 3' read ends in the region around the start codon (Fig 1C). For each bin, temporary 5' and 3' POs (tPO_L) are defined as the distances between the first nucleotide of the TIS and the nucleotide corresponding to the global maximum found in the profiles of the 5' and the 3' end at the left and at the right of the start codon, respectively (Fig 1C). Therefore, considering the occupancy profile as a function f of the nucleotide position x with respect to the TIS, the temporary 5' and 3' POs for each length bin are such that:

$$f(-5'tPO_L) \geq f(x) \forall x \in [-L + FL, -FL]$$

$$f(3'tPO_L) \geq f(x) \forall x \in [FL - 1, L - FL - 1]$$

The two sets of length-specific temporary POs are defined as:

$$5'tPO = \{5'tPO_{L_{min}}, \dots, 5'tPO_{L_{max}}\}$$

$$3'tPO = \{3'tPO_{L_{min}}, \dots, 3'tPO_{L_{max}}\}$$

where L_{min} and L_{max} are the minimum and the maximum length of the reads, respectively.

Next, to each read (R) mapping on the TIS the temporary POs corresponding to its length is assigned, obtaining two sets of read-specific tPOs:

$$5'tPO_R = \{5'tPO_{R_1}, \dots, 5'tPO_{R_N}\}$$

$$3'tPO_R = \{3'tPO_{R_1}, \dots, 3'tPO_{R_N}\}$$

where N is the number of reads mapping on the TIS.

Despite good estimation of P-site positions, artifacts may arise from either the small number of reads with a specific length or the presence of reads from ribosomes nearby the TIS, but not translating the first codon. In other words, the offset estimated independently from the global maximum of each read length is not necessarily always the best choice. In fact, while the most abundant population of reads are less subjected to the above mentioned biases and show consistent tPOs (see **S1–S12 Text**), this approach can produce high variability in tPO_L values of reads differing in only one nucleotide in length, especially across length bins with low number of reads.

To minimize this problem, riboWaltz exploits the most frequent tPO (optimal PO: oPO) associated to the predominant bins as a reference value for correcting the temporary POs of smaller bins. Briefly, the correction step defines for each length bin a new PO based on the local maximum, whose distance from the TIS is the closest to the oPO. The complete procedure is illustrated below.

The optimal PO at either 5' or 3' extremities (optimal extremity) are chosen as reference points to adjust the other tPOs. The optimal PO is selected between the two modes of read specific tPO sets ($Mode(5'tPO_R)$ and $Mode(3'tPO_R)$) as the one with the highest frequency.

$$oPO := \begin{cases} Mode(5'tPO_R) & \text{if } frequency(Mode(5'tPO_R)) \geq frequency(Mode(3'tPO_R)) \\ Mode(3'tPO_R) & \text{if } frequency(Mode(5'tPO_R)) < frequency(Mode(3'tPO_R)) \end{cases}$$

Note that this step also selects the optimal extremity to calculate the corrected PO.

The correction step is specific for each bin length and works as follows: if the offset associated to a bin is equal to the optimal PO, no changes are made. Otherwise, i) the local maxima of the occupancy profiles are extracted; ii) the distances between the first nucleotide of the TIS and each local maxima is computed; iii) the corrected PO is defined as the distance in point ii) that is closest to the optimal PO. Summarizing, given the set of local maxima positions (LMP) of the occupancy profile for the optimal extremity, the corrected PO for reads of length L (cPO_L) satisfies the following condition:

$$cPO_L - oPO = \min_{x \in LMP} (x - oPO)$$

Output

riboWaltz returns three data structures that can be used for multiple downstream analysis workflows (**Fig 1B**). The first is a list of sample-specific data frames containing for each read i) the position of the P-site (identified by the first nucleotide of the codon) with respect to the beginning of the transcript; ii) the distance between the P-site and both the start and the stop codon of the coding sequence; iii) the region of the transcript (5' UTR, CDS, 3' UTR) where the P-site is located and iv) the sequence of the triplet covered by the P-site, if a sequence file is provided as input. The second data structure is a data frame with the percentage of reads aligning across the start codon (if any) and along the whole transcriptome, stratified by sample and read length. Moreover, this file includes the P-site offsets from both the 5' and 3' extremities before and after the optimization ($5'tPO_L$, $3'tPO_L$, $5'cPO_L$, $3'cPO_L$ values). The third data structure is a data frame containing, for each transcript, the number of estimated in-frame P-sites on the CDS. This data frame can be used to estimate transcript-specific translation levels and to perform differential analysis comparing multiple samples in different conditions.

In addition, riboWaltz provides several graphical outputs based on the widely used “ggplot2” package. riboWaltz plots are described in more detail in the Results section. All

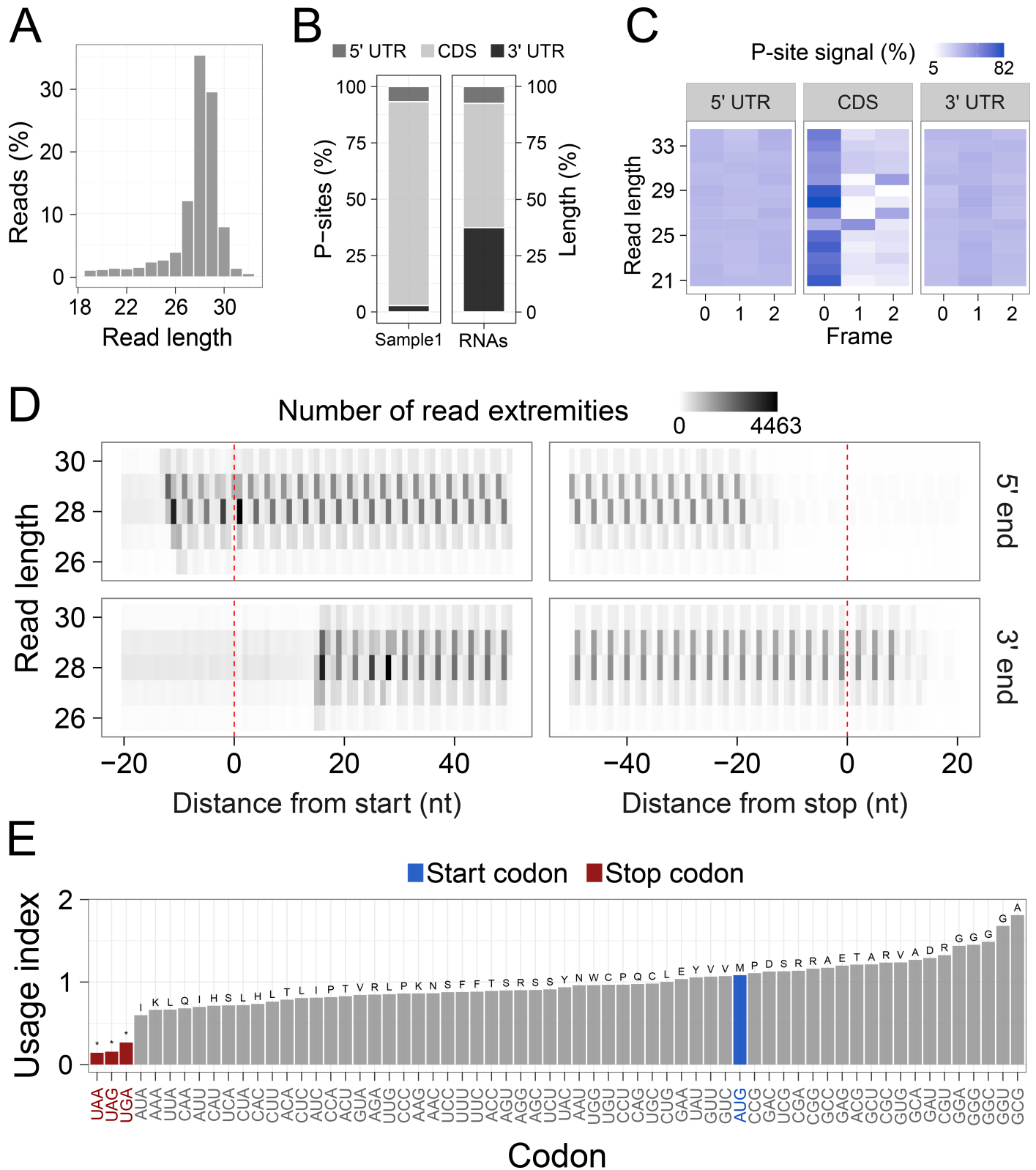


Fig 2. (A) Distribution of the read lengths. (B) Left, percentage of P-sites in the 5' UTR, CDS and 3' UTR of mRNAs from ribosome profiling data. Right, percentage of region lengths in mRNAs sequences. (C) Percentage of P-sites in the three frames along the 5' UTR, CDS and 3' UTR, stratified for read length. (D) Example of meta-gene heatmap reporting the signal associated to the 5' end (upper panel) and 3' end (lower panel) of the reads aligning around the start and the stop codon for

different read lengths. (E) Codon usage analysis based on in-frame P-sites. The codon usage index is calculated as the frequency of in-frame P-sites along the coding sequence associated to each codon, normalized for codon frequency in sequences. The amino-acids corresponding to the codons are displayed above each bar. All panels were obtained from ribosome profiling of whole mouse brain (GSE102318).

<https://doi.org/10.1371/journal.pcbi.1006169.g002>

graphical outputs are returned as lists containing objects of class “ggplot”, further customizable by the user, and data frames containing the source data for the plots.

Results

riboWaltz overview

To illustrate the functionalities of riboWaltz, we analyzed seven ribosome profiling datasets in yeast, mouse and human samples (see **Figs 2 and 3** and **S1–S13 Figs**).

riboWaltz integrates several graphical functions that provide multiple types of output results. First, the distribution of the length of the reads (**Fig 2A**): this is a useful preliminary inspection tool to understand the contribution of each bin to the final P-site determination, and eventually decide to remove certain bin from further analyses. Second, the percentage of P-sites located in the 5' UTR, CDS and 3' UTR regions of mRNAs compared to a uniform distribution weighted on region lengths, which simulates random P-site positioning along mRNAs (**Fig 2B**). This analysis is a good way to verify the expected enrichment of ribosome signal in the CDS. Third, to understand to which extent the obtained P-sites result in codon periodicity in the CDS, riboWaltz produces for every read group a plot with the percentage of P-sites in the three possible translation reading frames (periodicity analysis) for 5' UTR, CDS and 3' UTR (**Fig 2C**). Fourth, riboWaltz returns for every read length the meta-gene read density heatmap for both the 5' and 3' extremities of the reads (**Fig 2D**). This plot provides an overview of the occupancy profiles used for P-site determination and allows the visual inspection of PO values reliability. Fifth, to understand what codons display higher or lower ribosome density, riboWaltz provides the user with the analysis of the empirical codon usage, i.e. the frequency of in-frame P-sites along the coding sequence codon by codon, normalized for the frequency of each codon in the sequences (**Fig 2E**). Indeed, the comparison of these values in different biological conditions can be of great help to unravel possible defects in ribosome elongation at specific codons or aa-tRNAs use. Finally, single transcripts profiles and meta-gene profiles based on P-site position can be generated (**Fig 3B, top row** see **S1–S13 Figs** for examples) with multiple options: i) combining multiple replicates applying convenient scale factors provided by the user, ii) considering each replicate separately, or iii) selecting a subsets of reads with defined length.

Comparison with other tools

We tested riboWaltz on multiple ribosome profiling datasets in different model organisms: yeast (*S. cerevisiae*, [16,36]), mouse (mESC, [37]; whole brain, GSE102318) and human samples (Hek-293 [26]; MCF-7, GSE111866) and compared riboWaltz, RiboProfiling (v1.2.2, [33]) and Plastid (v0.4.5, [32]). Both Plastid and RiboProfiling compute the P-site offset considering the highest peak in the profile of reads mapping around the translation initiation site (TIS). Differently from RiboProfiling, Plastid considers only the signal from the 5' end of the read and imposes a default threshold for the minimum number of reads required for the computation. If this requirement is not met, Plastid will use a "default" constant offset value. **Table 1** and **S1–S6 Texts** contain the P-site offset comparison between the three tools, while **Table 2** and **S7–S12 Texts** provide additional details on the offsets computed by riboWaltz. The three tools were run using default settings. The

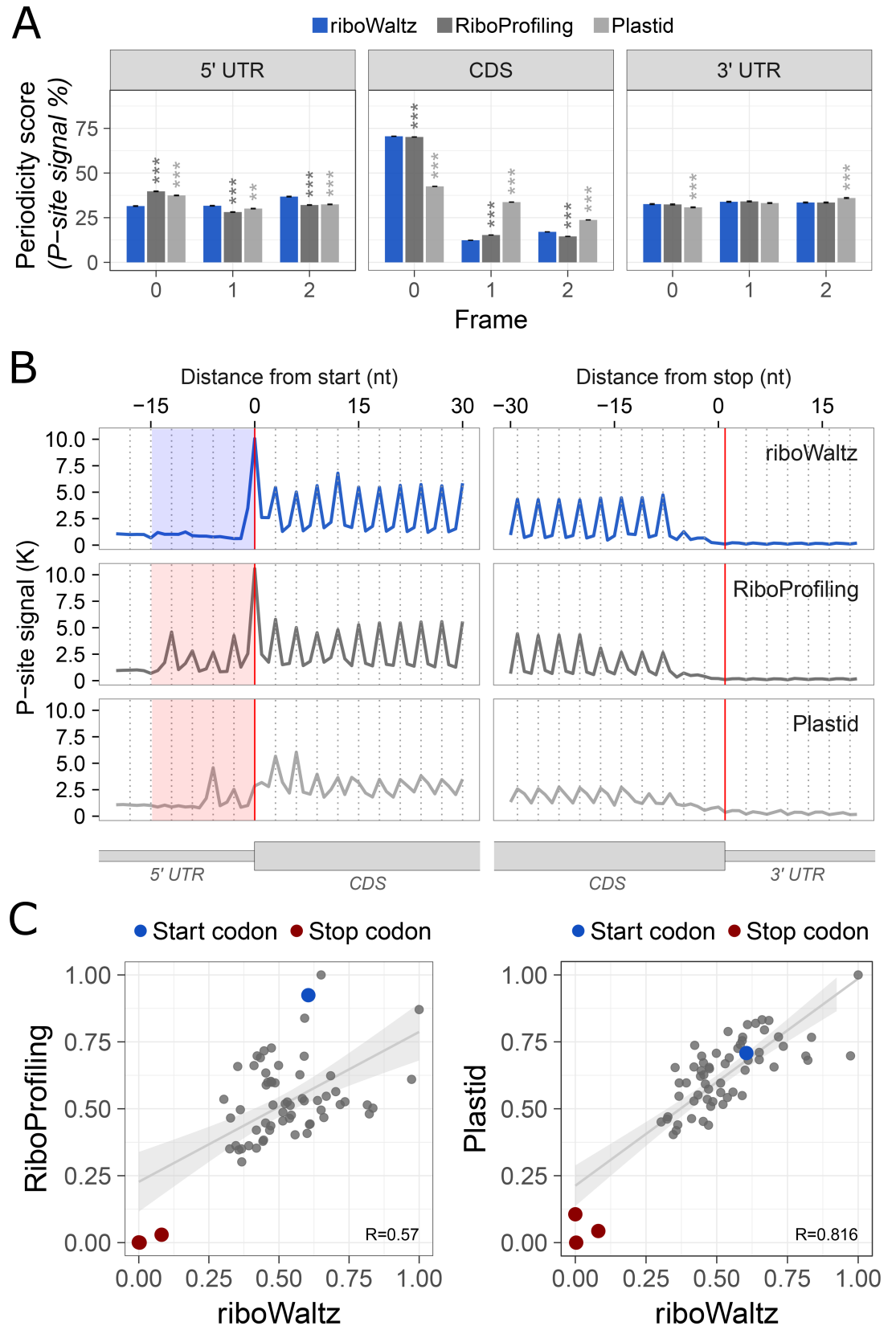


Fig 3. (A) Percentage of P-sites in the three frames (Periodicity score) along the 5' UTR, CDS and 3' UTR from ribosome profiling performed in mouse brain (GSE102318). The statistical significances from two-tailed Wilcoxon–Mann–Whitney test comparing RiboProfiling and Plastid with respect to riboWaltz are reported (P-value: ** < 0.01, *** < 0.001). (B) Meta-profiles showing the periodicity of ribosomes along the transcripts at the genome-wide scale. The three metaprofiles are based on the P-site identification obtained by using riboWaltz, RiboProfiling and Plastid. The shaded areas to the left of the start codon highlight the shift of the periodicity toward the 5' UTR that is absent in the case of data analysed using riboWaltz. (C) Comparison between the codon usage index based on in-frame P-sites from riboWaltz and RiboProfiling (left panel) and between the codon usage index based on in-frame P-sites from riboWaltz and Plastid (right panel). The length of the reads ranges from 19 up to 38 nucleotides (see Table 1) with the optimal PO used in the correction step of riboWaltz being 16 nucleotides from the 3' end.

<https://doi.org/10.1371/journal.pcbi.1006169.g003>

comparisons for single datasets are displayed in Fig 3 and in S1–S6 Figs, while the summary and the evaluation of the comparisons for all the datasets are displayed in Fig 4.

To evaluate the three methods, we considered two performance scores. First, we estimated the percentage of P-sites with correct frame within the CDS region (Periodicity score). The higher this measure, the better the performance. For RiboWaltz and RiboProfiling, this measure was comparable in almost all datasets, while Plastid performed worse (see Fig 3A and S1–S6A Figs for individual examples, Fig 4A and Table 3 for a resume. The median values are: riboWaltz: 57.07; RiboProfiling: 51.45; Plastid: 39.04).

Next, we took into consideration the meta-profiles. In all datasets riboWaltz displayed a neat periodicity uniquely in the CDS (Fig 3B and S1–S6B Figs), with almost no signal along the UTRs, neither in the proximity of the start nor of the stop codons. By contrast, both Plastid and RiboProfiling generated a shift toward the 5' UTR in the beginning of the periodic region

Table 1. Comparison of the P-site offsets identified for each read length by riboWaltz, RiboProfiling and Plastid in mouse (GSE102318).

Read length	riboWaltz		RiboProfiling		Plastid	
	from 5' end	from 3' end	from 5' end	from 3' end	from 5' end	from 3' end
19	2	16	2	16	13	5
20	4	15	4	15	13	6
21	4	16	4	16	13	7
22	5	16	5	16	13	8
23	6	16	6	16	13	9
24	7	16	7	16	13	10
25	8	16	1	25	13	11
26	10	15	10	15	13	12
27	10	16	10	16	13	13
28	11	16	1	28	5	22
29	12	16	12	16	13	15
30	12	17	10	19	35	6
31	13	17	20	50	13	17
32	15	16	15	16	13	18
33	16	16	17	15	13	19
34	17	16	17	16	13	20
35	18	16	18	16	13	21
36	16	19	19	16	13	22
37	20	16	22	58	13	23
38	21	16	15	22	13	24

The POs computed from both read extremities are reported. The optimal PO used in the correction step of riboWaltz corresponds to 16 nucleotides from the 3' end.

<https://doi.org/10.1371/journal.pcbi.1006169.t001>

Table 2. Comparison between temporary and corrected P-site offsets identified by riboWaltz in mouse (GSE102318).

Read length	Number of reads (%)	Temporary P-site offset		Corrected P-site offset	
		from 5'	from 3'	from 5'	from 3'
19	0.888	2	16	2	16
20	0.986	4	15	4	15
21	1.203	4	16	4	16
22	1.113	5	16	5	16
23	1.335	6	16	6	16
24	2.191	7	16	7	16
25	2.494	8	16	8	16
26	3.743	10	15	10	15
27	11.891	10	16	10	16
28	34.943	11	16	11	16
29	29.125	12	16	12	16
30	7.771	12	17	12	17
31	1.194	11	19	13	17
32	0.365	15	16	15	16
33	0.235	16	16	16	16
34	0.164	17	16	17	16
35	0.115	18	16	18	16
36	0.087	10	25	16	19
37	0.057	20	16	20	16
38	0.034	21	16	21	16

The POs computed from both read extremities are reported. The optimal PO used in the correction step correspond to 16 nucleotides from the 3' end.

<https://doi.org/10.1371/journal.pcbi.1006169.t002>

(Fig 3B and S1–S6B Figs). The presence of periodic peaks in the 5'UTR is undoubtedly a source of biological inaccuracy, conflicting with basic concepts in translation. In fact, outside the coding sequence, ribosomes are generally in non-translating mode. Translation can indeed occur outside the CDS, with upstream ORFs being the most documented examples. Nonetheless, occasional translation outside the CDS is unlikely to affect the codon periodicity in 5' UTR regions, especially when metagene plots are anchored on the annotated AUG start codons. The presence of prominent codon periodicity in the 5'UTR in this latter case most likely results from a technical mistake, such as the inaccurate computation of the P-site offset. To quantify this effect, we determined a “TIS accuracy score”, comparing the amount of periodic signal in a local window before and after the translation initiation site. Considering the occupancy profile as a function f of the nucleotide position x with respect to the TIS, the TIS accuracy score is defined as follows:

$$TIS\ accuracy\ score := \frac{\sum_{\{x \in [0,14]:3|x\}} f(x)}{\sum_{\{x \in [-15,14]:3|x\}} f(x)}$$

In the ideal scenario, this score should be equal to 1, meaning that the periodicity can be detected only within the CDS region. Lower scores are associated with a progressive increase of periodicity in the 5'UTR, indicative of ribosome mislocalization. Importantly, riboWaltz shows significantly higher TIS accuracy scores with respect to both RiboProfiling and Plastid (median values: 0.84, 0.62, 0.71 respectively. See Fig 4B and Table 4 for a resume).

The correct localization of ribosomes is a crucial step for obtaining estimations of the codon usage and for any downstream analyses. Empirical codon usage determination is a

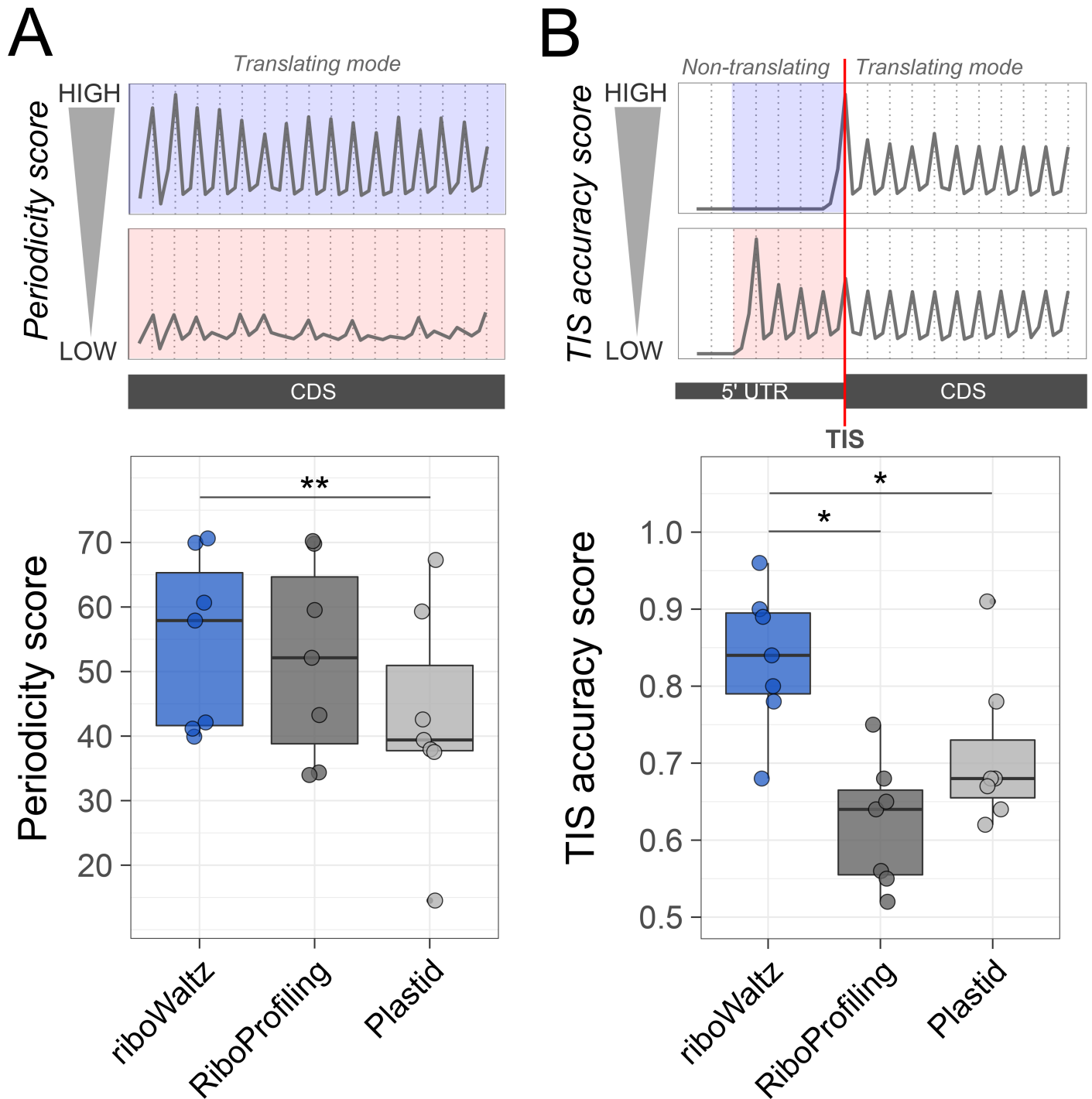


Fig 4. (A) Comparison of the percentage of P-sites in frame 0 (Periodicity score) along the coding sequence and (B) comparison of the average TIS accuracy score based on P-sites identification by riboWaltz, RiboProfiling and Plastid. Both panels display the results obtained from 7 datasets (2 yeast, 3 mouse and 2 human), each dataset represented by a dot. Statistical significances from paired one-tailed Wilcoxon–Mann–Whitney test are shown (* $P < 0.05$, ** $P < 0.01$).

<https://doi.org/10.1371/journal.pcbi.1006169.g004>

popular analysis for ribosome profiling data, and it is equally important for the biological interpretation of results and for the development of reliable mathematical models of translation [20–22,38–40]. To highlight the differences arising in codon usage after the identification

Table 3. Summary and comparison of the percentage of P-sites in frame 0 along the coding sequence (Periodicity score) based on P-sites identification by riboWaltz, RiboProfiling and Plastid.

Organism	Reference	Mean % of P-site in frame 0			Statistical significance	
		riboWaltz	Ribo Profiling	Plastid	riboWaltz vs RiboProfiling	riboWaltz vs Plastid
Yeast	Lareau et al., 2014 [16]	42.11	43.26	39.40	$5.90 \cdot 10^{-4}$ ***	$8.99 \cdot 10^{-21}$ ***
Yeast	Beaupere et al., 2017 [36]	69.95	69.80	67.29	0.0046 **	$5.40 \cdot 10^{-124}$ ***
Mouse	This publication (GSE102318)	70.63	70.21	42.58	$1.12 \cdot 10^{-7}$ ***	$< 1 \cdot 10^{-324}$ ***
Mouse (IP RPL10)	Shi et al., 2017 [37]	39.91	34.37	37.94	$< 1 \cdot 10^{-324}$ ***	$2.15 \cdot 10^{-125}$ ***
Mouse (IP RPL22)	Shi et al., 2017 [37]	41.15	33.97	37.54	$< 1 \cdot 10^{-324}$ ***	$4.39 \cdot 10^{-277}$ ***
Human	Gao et al., 2015 [26]	60.67	59.53	59.31	$2.37 \cdot 10^{-15}$ ***	$1.27 \cdot 10^{-15}$ ***
Human	This publication (GSE111866)	57.90	52.13	14.52	$5.89 \cdot 10^{-191}$ ***	$< 1 \cdot 10^{-324}$ ***

The values obtained from 7 datasets (2 yeast, 3 mouse and 2 human) are shown, together with the statistical significances from two-tailed Wilcoxon–Mann–Whitney test (P-value:

* < 0.05

** < 0.01

*** < 0.001).

<https://doi.org/10.1371/journal.pcbi.1006169.t003>

Table 4. Summary and comparison of the average TIS accuracy score based on P-sites identification by riboWaltz, RiboProfiling and Plastid.

Organism	Reference	Average TIS accuracy score			Statistical significance	
		riboWaltz	Ribo Profiling	Plastid	riboWaltz vs RiboProfiling	riboWaltz vs Plastid
Yeast	Lareau et al., 2014 [16]	0.90	0.75	0.91	$6.0 \cdot 10^{-45}$ ***	0.6817
Yeast	Beaupere et al., 2017 [36]	0.96	0.56	0.68	$< 1 \cdot 10^{-324}$ ***	$< 1 \cdot 10^{-324}$ ***
Mouse	This publication (GSE102318)	0.89	0.65	0.68	$< 1 \cdot 10^{-324}$ ***	$< 1 \cdot 10^{-324}$ ***
Mouse (IP RPL10)	Shi et al., 2017 [37]	0.68	0.56	0.67	$1.5 \cdot 10^{-98}$ ***	0.9015
Mouse (IP RPL22)	Shi et al., 2017 [37]	0.78	0.52	0.79	$< 1 \cdot 10^{-324}$ ***	0.0013 **
Human	Gao et al., 2015 [26]	0.84	0.68	0.62	$3.4 \cdot 10^{-221}$ ***	$< 1 \cdot 10^{-324}$ ***
Human	This publication (GSE111866)	0.80	0.65	0.64	$3.2 \cdot 10^{-78}$ ***	$1.1 \cdot 10^{-50}$ ***

The values obtained from 7 datasets (2 yeast, 3 mouse and 2 human) are shown, together with the statistical significances from two-tailed Wilcoxon–Mann–Whitney test (P-value:

* < 0.05

** < 0.01

*** < 0.001).

<https://doi.org/10.1371/journal.pcbi.1006169.t004>

of the P-site using different approaches, we compared codon usage values across each dataset analysed using riboWaltz, RiboProfiling and Plastid (Fig 3C and S1–S6C Figs). The results show correlation values ranging from 0.075 to 0.999. This analysis is a descriptive evaluation of the difference between riboWaltz and the other tools in computing the codon usage, depending on the different approach used for the P-site determination.

In summary we show that the choice of the strategy for P-site positioning has a strong impact on downstream analyses and that riboWaltz is a more reliable tool for the identification of P-site offsets and the positional analysis of ribosome profiling data.

Availability and future directions

riboWaltz identifies with high precision the position of ribosome P-sites from ribosome profiling data. By improving on other currently-available approaches, riboWaltz can assist with the detailed interrogation of ribosome profiling data, providing precise information that may lay the groundwork for further positional analyses and new biological discoveries.

riboWaltz is written in the R programming language, and is compatible with Linux, Mac, or Windows PCs. riboWaltz depends on multiple R packages such as GenomicFeatures for handling GTF/GFF3 files, Biostrings, BSgenome and GenomicAlignments for dealing with sequence data and ggplot2 for data visualization. Furthermore, to easily handle datasets with several millions of reads preserving a high efficiency in terms of RAM usage and running-time, riboWaltz employs an enhanced version of data frames provided by the data.table package. Installation instructions for the dependencies are provided in the manual.

riboWaltz is an Open-Source software package that can be extended in future releases to include other analysis methods as they are developed. Source code for riboWaltz is distributed under the MIT license and is available at the following GitHub repository: <https://github.com/LabTranslationalArchitectomics/riboWaltz>. The package includes the R implementation of riboWaltz, data used in this article, extensive documentation and a stable release.

Supporting information

S1 Fig. (A) Percentage of P-sites in the three frames along the 5' UTR, CDS and 3' UTR from ribosome profiling in Hek-293 (Gao et al., 2015). The statistical significances from two-tailed Wilcoxon–Mann–Whitney test comparing RiboProfiling and Plastid with respect to riboWaltz are reported (P-value: *** < 0.001). (B) Meta-profiles showing the periodicity of ribosomes along the transcripts at the genome-wide scale. The three metaprofiles are based on the P-site identification obtained by using riboWaltz, RiboProfiling and Plastid. The shaded areas to the left of the start codon highlight the shift of the periodicity toward the 5' UTR that is absent in the case of data analysed using riboWaltz. (C) Comparison between the codon usage index based on in-frame P-sites from riboWaltz and RiboProfiling (left panel) and between the codon usage index based on in-frame P-sites from riboWaltz and Plastid (right panel). The length of the reads ranges from 25 up to 34 nucleotides (see Table 1) with the optimal PO used in the correction step of riboWaltz being 12 nucleotides from the 5' end. (TIF)

S2 Fig. (A) Percentage of P-sites in the three frames along the 5' UTR, CDS and 3' UTR from ribosome profiling in MCF-7 (GSE111866). The statistical significances from two-tailed Wilcoxon–Mann–Whitney test comparing RiboProfiling and Plastid with respect to riboWaltz are reported (P-value: * < 0.05, ** < 0.01, *** < 0.001). (B) Meta-profiles showing the periodicity of ribosomes along the transcripts at the genome-wide scale. The three metaprofiles are based on the P-site identification obtained by using riboWaltz, RiboProfiling and Plastid. The shaded

areas to the left of the start codon highlight the shift of the periodicity toward the 5' UTR that is absent in the case of data analysed using riboWaltz. (C) Comparison between the codon usage index based on in-frame P-sites from riboWaltz and RiboProfiling (left panel) and between the codon usage index based on in-frame P-sites from riboWaltz and Plastid (right panel). The length of the reads ranges from 20 to 45 nucleotides (see [S2 Text](#)) with the optimal PO used in the correction step of riboWaltz being 11 nucleotides from the 5' end. (TIF)

S3 Fig. (A) Percentage of P-sites in the three frames along the 5' UTR, CDS and 3' UTR from ribosome profiling in mouse after immunoprecipitation of ribosomes using the ribosomal protein RPL10 as tag (Shi et al. 2017). The statistical significances from two-tailed Wilcoxon–Mann–Whitney test comparing RiboProfiling and Plastid with respect to riboWaltz are reported (P-value: * < 0.05, ** < 0.01, *** < 0.001). (B) Meta-profiles showing the periodicity of ribosomes along the transcripts at the genome-wide scale. The three metaprofiles are based on the P-site identification obtained by using riboWaltz, RiboProfiling and Plastid. The shaded areas to the left of the start codon highlight the shift of the periodicity toward the 5' UTR that is absent in the case of data analysed using riboWaltz. (C) Comparison between the codon usage index based on in-frame P-sites from riboWaltz and RiboProfiling (left panel) and between the codon usage index based on in-frame P-sites from riboWaltz and Plastid (right panel). The length of the reads ranges from 19 up to 50 nucleotides (see [S3 Text](#)) with the optimal PO used in the correction step of riboWaltz being 11 nucleotides from the 5' end. (TIF)

S4 Fig. (A) Percentage of P-sites in the three frames along the 5' UTR, CDS and 3' UTR from ribosome profiling in mouse after immunoprecipitation of ribosomes using the ribosomal protein RPL22 as tag (Shi et al. 2017). The statistical significances from two-tailed Wilcoxon–Mann–Whitney test comparing RiboProfiling and Plastid with respect to riboWaltz are reported (P-value: * < 0.05, *** < 0.001). (B) Meta-profiles showing the periodicity of ribosomes along the transcripts at the genome-wide scale. The three metaprofiles are based on the P-site identification obtained by using riboWaltz, RiboProfiling and Plastid. The shaded areas to the left of the start codon highlight the shift of the periodicity toward the 5' UTR that is absent in the case of data analysed using riboWaltz. (C) Comparison between the codon usage index based on in-frame P-sites from riboWaltz and RiboProfiling (left panel) and between the codon usage index based on in-frame P-sites from riboWaltz and Plastid (right panel). The length of the reads ranges from 19 up to 50 nucleotides (see [S2 Text](#)) with the optimal PO used in the correction step of riboWaltz being 11 nucleotides from the 5' end. (TIF)

S5 Fig. (A) Percentage of P-sites in the three frames along the 5' UTR, CDS and 3' UTR from ribosome profiling in yeast (Beaupere et al., 2017). The statistical significances from two-tailed Wilcoxon–Mann–Whitney test comparing RiboProfiling and Plastid with respect to riboWaltz are reported (P-value: * < 0.05, ** < 0.01, *** < 0.001). (B) Meta-profiles showing the periodicity of ribosomes along the transcripts at the genome-wide scale. The three metaprofiles are based on the P-site identification obtained by using riboWaltz, RiboProfiling and Plastid. The shaded areas to the left of the start codon highlight the shift of the periodicity toward the 5' UTR that is absent in the case of data analysed using riboWaltz. (C) Comparison between the codon usage index based on in-frame P-sites from riboWaltz and RiboProfiling (left panel) and between the codon usage index based on in-frame P-sites from riboWaltz and Plastid (right panel). The length of the reads ranges from 20 to 46 nucleotides (see [S5 Text](#)), with the

optimal PO used in the correction step of riboWaltz being 15 nucleotides from the 3' end. (TIF)

S6 Fig. (A) Percentage of P-sites in the three frames along the 5' UTR, CDS and 3' UTR from ribosome profiling in yeast (Lareau et al., 2014). The statistical significances from two-tailed Wilcoxon–Mann–Whitney test comparing RiboProfiling and Plastid with respect to riboWaltz are reported (P-value: * < 0.05, *** < 0.001). (B) Meta-profiles showing the periodicity of ribosomes along the transcripts at the genome-wide scale. The three metaprofiles are based on the P-site identification obtained by using riboWaltz, RiboProfiling and Plastid. The shaded areas to the left of the start codon highlight the shift of the periodicity toward the 5' UTR that is absent in the case of data analysed using riboWaltz. (C) Comparison between the codon usage index based on in-frame P-sites from riboWaltz and RiboProfiling (left panel) and between the codon usage index based on in-frame P-sites from riboWaltz and Plastid (right panel). The length of the reads ranges from 21 to 40 nucleotides (see S6 Text) with the optimal PO used in the correction step of riboWaltz being 13 nucleotides from the 5' end. (TIF)

S7 Fig. Performance of riboWaltz compared with RiboProfiling and Plastid in mouse brain tissue (GSE102318) using reads with a length of 27, 28 and 29 nucleotides. (A) Percentage of P-sites in the three frames along the 5' UTR, CDS and 3' UTR. The statistical significances from two-tailed Wilcoxon–Mann–Whitney test comparing RiboProfiling and Plastid with respect to riboWaltz are reported (P-value: * < 0.05, ** < 0.01, *** < 0.001). (B) Meta-profiles showing the periodicity of ribosomes along the transcripts at the genome-wide scale. The three metaprofiles are based on the P-site identification obtained using riboWaltz, RiboProfiling and Plastid. The shaded areas to the left of the start codon highlight the shift of the periodicity toward the 5' UTR that is absent in the case of data analysed using riboWaltz. (TIF)

S8 Fig. Performance of riboWaltz compared with RiboProfiling and Plastid in Hek-293 cells (Gao et al., 2015) using reads with a length of 27, 28 and 29 nucleotides. (A) Percentage of P-sites in the three frames along the 5' UTR, CDS and 3' UTR. The statistical significances from two-tailed Wilcoxon–Mann–Whitney test comparing RiboProfiling and Plastid with respect to riboWaltz are reported (P-value: *** < 0.001). (B) Meta-profiles showing the periodicity of ribosomes along the transcripts at the genome-wide scale. The three metaprofiles are based on the P-site identification obtained using riboWaltz, RiboProfiling and Plastid. The shaded areas to the left of the start codon highlight the shift of the periodicity toward the 5' UTR that is absent in the case of data analysed using riboWaltz. (TIF)

S9 Fig. Performance of riboWaltz compared with RiboProfiling and Plastid in MCF-7 cells (GSE111866) using reads with length of 28, 29 and 30 nucleotides. (A) Percentage of P-sites in the three frames along the 5' UTR, CDS and 3' UTR. The statistical significances from two-tailed Wilcoxon–Mann–Whitney test comparing RiboProfiling and Plastid with respect to riboWaltz are reported (P-value: * < 0.05, *** < 0.001). (B) Meta-profiles showing the periodicity of ribosomes along the transcripts at the genome-wide scale. The three metaprofiles are based on the P-site identification obtained using riboWaltz, RiboProfiling and Plastid. The shaded areas to the left of the start codon highlight the shift of the periodicity toward the 5' UTR that is absent in the case of data analysed using riboWaltz. (TIF)

S10 Fig. Performance of riboWaltz compared with RiboProfiling and Plastid in mouse after immunoprecipitation of ribosomes using the ribosomal protein RPL10 as tag (Shi et al. 2017) using reads with length of 29, 30 and 31 nucleotides. (A) Percentage of P-sites in the three frames along the 5' UTR, CDS and 3' UTR. The statistical significances from two-tailed Wilcoxon–Mann–Whitney test comparing RiboProfiling and Plastid with respect to riboWaltz are reported (P-value: * < 0.05, *** < 0.001). (B) Meta-profiles showing the periodicity of ribosomes along the transcripts at the genome-wide scale. The three metaprofiles are based on the P-site identification obtained using riboWaltz, RiboProfiling and Plastid. The shaded areas to the left of the start codon highlight the shift of the periodicity toward the 5' UTR that is absent in the case of data analysed using riboWaltz.
(TIF)

S11 Fig. Performance of riboWaltz compared with RiboProfiling and Plastid in mouse after immunoprecipitation of ribosomes using the ribosomal protein RPL22 as tag (Shi et al. 2017) using reads with length of 28, 29 and 30 nucleotides. (A) Percentage of P-sites in the three frames along the 5' UTR, CDS and 3' UTR. The statistical significances from two-tailed Wilcoxon–Mann–Whitney test comparing RiboProfiling and Plastid with respect to riboWaltz are reported (P-value: *** < 0.001). (B) Meta-profiles showing the periodicity of ribosomes along the transcripts at the genome-wide scale. The three metaprofiles are based on the P-site identification obtained using riboWaltz, RiboProfiling and Plastid. The shaded areas to the left of the start codon highlight the shift of the periodicity toward the 5' UTR that is absent in the case of data analysed using riboWaltz.
(TIF)

S12 Fig. Performance of riboWaltz compared with RiboProfiling and Plastid in yeast (Beaupere et al., 2017) using reads with a length of 27, 28 and 29 nucleotides. (A) Percentage of P-sites in the three frames along the 5' UTR, CDS and 3' UTR. The statistical significances from two-tailed Wilcoxon–Mann–Whitney test comparing RiboProfiling and Plastid with respect to riboWaltz are reported (P-value: ** < 0.01, *** < 0.001). (B) Meta-profiles showing the periodicity of ribosomes along the transcripts at the genome-wide scale. The three metaprofiles are based on the P-site identification obtained using riboWaltz, RiboProfiling and Plastid. The shaded areas to the left of the start codon highlight the shift of the periodicity toward the 5' UTR that is absent in the case of data analysed using riboWaltz.
(TIF)

S13 Fig. Performance of riboWaltz compared with RiboProfiling and Plastid in yeast (Larreau et al., 2014) using reads with a length of 28, 29 and 30 nucleotides. (A) Percentage of P-sites in the three frames along the 5' UTR, CDS and 3' UTR. The statistical significances from two-tailed Wilcoxon–Mann–Whitney test comparing RiboProfiling and Plastid with respect to riboWaltz are reported (P-value: ** < 0.01, *** < 0.001). (B) Meta-profiles showing the periodicity of ribosomes along the transcripts at the genome-wide scale. The three metaprofiles are based on the P-site identification obtained using riboWaltz, RiboProfiling and Plastid. The shaded areas to the left of the start codon highlight the shift of the periodicity toward the 5' UTR that is absent in the case of data analysed using riboWaltz.
(TIF)

S1 Text. Comparison of P-site offsets identified for each read length by riboWaltz, RiboProfiling and Plastid in human (Hek-293, Gao et al., 2015). The PO computed from both read extremities are reported. The optimal PO used in the correction step of riboWaltz corresponds to 12 nucleotides from the 5' end.
(DOCX)

S2 Text. Comparison of P-site offsets identified for each read length by riboWaltz, Ribo-Profiling and Plastid in human (MCF-7, GSE111866). The PO computed from both read extremities are reported. The optimal PO used in the correction step of riboWaltz corresponds to 11 nucleotides from the 5' end.

(DOCX)

S3 Text. Comparison of P-site offsets identified for each read length by riboWaltz, Ribo-Profiling and Plastid in mouse (after pull-down of RLP10, Shi et al. 2017). The PO computed from both read extremities are reported. The optimal PO used in the correction step of riboWaltz corresponds to 11 nucleotides from the 5' end.

(DOCX)

S4 Text. Comparison of P-site offsets identified for each read length by riboWaltz, Ribo-Profiling and Plastid in mouse (after pull-down of RLP22, Shi et al. 2017). The PO computed from both read extremities are reported. The optimal PO used in the correction step of riboWaltz corresponds to 11 nucleotides from the 5' end.

(DOCX)

S5 Text. Comparison of P-site offsets identified for each read length by riboWaltz, Ribo-Profiling and Plastid in yeast (Beaupere et al., 2017). The PO computed from both read extremities are reported. The optimal PO used in the correction step of riboWaltz corresponds to 15 nucleotides from the 3' end.

(DOCX)

S6 Text. Comparison of P-site offsets identified for each read length by riboWaltz, Ribo-Profiling and Plastid in yeast (Lareau et al., 2014). The PO computed from both read extremities are reported. The optimal PO used in the correction step of riboWaltz corresponds to 13 nucleotides from the 5' end.

(DOCX)

S7 Text. Comparison between temporary and corrected P-site offsets identified by riboWaltz in human (Hek-293, Gao et al., 2015). The PO computed from both read extremities are reported. The optimal PO used in the correction step corresponds to 12 nucleotides from the 5' end.

(DOCX)

S8 Text. Comparison between temporary and corrected P-site offsets identified by riboWaltz in human (MCF-7, GEO111866). The PO computed from both read extremities are reported. The optimal PO used in the correction step corresponds to 11 nucleotides from the 5' end.

(DOCX)

S9 Text. Comparison between temporary and corrected P-site offsets identified by riboWaltz in mouse (after pull-down of RLP10, Shi et al. 2017). The PO computed from both read extremities are reported. The optimal PO used in the correction step corresponds to 11 nucleotides from the 5' end.

(DOCX)

S10 Text. Comparison between temporary and corrected P-site offsets identified by riboWaltz in mouse (after pull-down of RLP22, Shi et al. 2017). The PO computed from both read extremities are reported. The optimal PO used in the correction step corresponds to

11 nucleotides from the 5' end.
(DOCX)

S11 Text. Comparison between temporary and corrected P-site offsets identified by riboWaltz in yeast (Beaupere et al., 2017). The PO computed from both read extremities are reported. The optimal PO used in the correction step corresponds to 15 nucleotides from the 3' end.
(DOCX)

S12 Text. Comparison between temporary and corrected P-site offsets identified by riboWaltz in yeast (Lareau et al., 2014). The PO computed from both read extremities are reported. The optimal PO used in the correction step corresponds to 13 nucleotides from the 5' end.
(DOCX)

S13 Text. Supplementary methods.
(DOCX)

Acknowledgments

We thank the Core Facility, Next Generation Sequencing Facility (HTS) CIBIO, University of Trento (Italy) for technical support.

Author Contributions

Conceptualization: Toma Tebaldi, Gabriella Viero.

Data curation: Fabio Lauria, Toma Tebaldi.

Formal analysis: Fabio Lauria, Toma Tebaldi.

Funding acquisition: Thomas H. Gillingwater, Gabriella Viero.

Investigation: Fabio Lauria, Toma Tebaldi, Gabriella Viero.

Methodology: Fabio Lauria, Toma Tebaldi, Paola Bernabò, Ewout J. N. Groen, Thomas H. Gillingwater.

Project administration: Gabriella Viero.

Software: Fabio Lauria, Toma Tebaldi.

Supervision: Fabio Lauria, Toma Tebaldi, Gabriella Viero.

Visualization: Fabio Lauria, Toma Tebaldi, Gabriella Viero.

Writing – original draft: Fabio Lauria, Toma Tebaldi, Gabriella Viero.

Writing – review & editing: Fabio Lauria, Toma Tebaldi, Gabriella Viero.

References

1. Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*. 2009; 324: 218–23. <https://doi.org/10.1126/science.1168978> PMID: 19213877
2. Ingolia NT, Brar GA, Rouskin S, McGeachy AM, Weissman JS. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat Protoc*. 2012; 7: 1534–50. <https://doi.org/10.1038/nprot.2012.086> PMID: 22836135

3. Steitz JA. Polypeptide chain initiation: nucleotide sequences of the three ribosomal binding sites in bacteriophage R17 RNA. *Nature*. 1969; 224: 957–64. Available: <http://www.ncbi.nlm.nih.gov/pubmed/5360547> PMID: 5360547
4. Wolin SL, Walter P. Ribosome pausing and stacking during translation of a eukaryotic mRNA. *EMBO J*. 1988; 7: 3559–69. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=454858&tool=pmcentrez&rendertype=abstract> PMID: 2850168
5. Michel AM, Baranov P V. Ribosome profiling: a Hi-Def monitor for protein synthesis at the genome-wide scale. *Wiley Interdiscip Rev RNA*. 2013; 4: 473–90. <https://doi.org/10.1002/wrna.1172> PMID: 23696005
6. 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. <https://doi.org/10.1038/nrm4069> PMID: 26465719
7. Fields AP, Rodriguez EH, Jovanovic M, Stern-Ginossar N, Haas BJ, Mertins P, et al. A Regression-Based Analysis of Ribosome-Profiling Data Reveals a Conserved Complexity to Mammalian Translation. *Mol Cell*. 2015; 60: 816–827. <https://doi.org/10.1016/j.molcel.2015.11.013> PMID: 26638175
8. Raj A, Wang SH, Shim H, Harpak A, Li YI, Engelmann B, et al. Thousands of novel translated open reading frames in humans inferred by ribosome footprint profiling. *Elife*. 2016; 5. <https://doi.org/10.7554/eLife.13328> PMID: 27232982
9. Calviello L, Mukherjee N, Wyler E, Zauber H, Hirsekorn A, Selbach M, et al. Detecting actively translated open reading frames in ribosome profiling data. *Nat Methods*. 2016; 13: 165–70. <https://doi.org/10.1038/nmeth.3688> PMID: 26657557
10. Malone B, Atanassov I, Aeschmann F, Li X, Großhans H, Dieterich C. Bayesian prediction of RNA translation from ribosome profiling. *Nucleic Acids Res*. 2016; 45: 2960–2972. <https://doi.org/10.1093/nar/gkw1350> PMID: 28126919
11. Xiao Z, Zou Q, Liu Y, Yang X. Genome-wide assessment of differential translations with ribosome profiling data. *Nat Commun*. 2016; 7: 11194. <https://doi.org/10.1038/ncomms11194> PMID: 27041671
12. Zhong Y, Karaletsos T, Drewe P, Sreedharan VT, Kuo D, Singh K, et al. RiboDiff: Detecting changes of mRNA translation efficiency from ribosome footprints. *Bioinformatics*. 2017; 33: 139–141. <https://doi.org/10.1093/bioinformatics/btw585> PMID: 27634950
13. Martens AT, Taylor J, Hilser VJ. Ribosome A and P sites revealed by length analysis of ribosome profiling data. *Nucleic Acids Res*. 2015; 43: 3680–7. <https://doi.org/10.1093/nar/gkv200> PMID: 25805170
14. Legendre R, Baudin-Baillieu A, Hatin I, Namy O. RiboTools: a Galaxy toolbox for qualitative ribosome profiling analysis. *Bioinformatics*. 2015; 31: 2586–8. <https://doi.org/10.1093/bioinformatics/btv174> PMID: 25812744
15. Wang H, McManus J, Kingsford C. Accurate Recovery of Ribosome Positions Reveals Slow Translation of Wobble-Pairing Codons in Yeast. *J Comput Biol*. 2017; 24: 486–500. <https://doi.org/10.1089/cmb.2016.0147> PMID: 27726445
16. Lareau LF, Hite DH, Hogan GJ, Brown PO. Distinct stages of the translation elongation cycle revealed by sequencing ribosome-protected mRNA fragments. *Elife*. 2014; 3: e01257. <https://doi.org/10.7554/eLife.01257> PMID: 24842990
17. Guo H, Ingolia NT, Weissman JS, Bartel DP. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*. 2010; 466: 835–40. <https://doi.org/10.1038/nature09267> PMID: 20703300
18. Gritsenko AA, Hulsman M, Reinders MJT, de Ridder D. Unbiased Quantitative Models of Protein Translation Derived from Ribosome Profiling Data. *PLoS Comput Biol*. 2015; 11: e1004336. <https://doi.org/10.1371/journal.pcbi.1004336> PMID: 26275099
19. Michel AM, Andreev DE, Baranov P V. Computational approach for calculating the probability of eukaryotic translation initiation from ribo-seq data that takes into account leaky scanning. *BMC Bioinformatics*. 2014; 15: 380. <https://doi.org/10.1186/s12859-014-0380-4> PMID: 25413677
20. Renana S, Tamir T. Modelling the efficiency of codon-tRNA interactions based on codon usage bias. *DNA Res*. 2014; 21: 511–525. <https://doi.org/10.1093/dnares/dsu017> PMID: 24906480
21. Dana A, Tuller T. Mean of the typical decoding rates: a new translation efficiency index based on the analysis of ribosome profiling data. *G3 (Bethesda)*. 2014; 5: 73–80. <https://doi.org/10.1534/g3.114.015099> PMID: 25452418
22. Pop C, Rouskin S, Ingolia NT, Han L, Phizicky EM, Weissman JS, et al. Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation. *Mol Syst Biol*. 2014; 10: 770. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4300493&tool=pmcentrez&rendertype=abstract> PMID: 25538139
23. Weinberg DE, Shah P, Eichhorn SW, Hussmann JA, Plotkin JB, Bartel DP. Improved Ribosome-Footprint and mRNA Measurements Provide Insights into Dynamics and Regulation of Yeast Translation. *Cell Rep*. 2016; 14: 1787–1799. <https://doi.org/10.1016/j.celrep.2016.01.043> PMID: 26876183

24. Hsu PY, Calviello L, Wu H-YL, Li F-W, Rothfels CJ, Ohler U, et al. Super-resolution ribosome profiling reveals unannotated translation events in *Arabidopsis*. *Proc Natl Acad Sci*. 2016; 113: E7126–E7135. <https://doi.org/10.1073/pnas.1614788113> PMID: 27791167
25. Kochetov A V., Allmer J, Klimenko AI, Zuraev BS, Matushkin YG, Lashin SA, et al. AltORFev facilitates the prediction of alternative open reading frames in eukaryotic mRNAs. *Bioinformatics*. 2017; 33: 923–925. <https://doi.org/10.1093/bioinformatics/btw736> PMID: 28039164
26. Gao X, Wan J, Liu B, Ma M, Shen B, Qian SB. Quantitative profiling of initiating ribosomes in vivo. *Nat Methods*. 2015; 12: 147–153. <https://doi.org/10.1038/nmeth.3208> PMID: 25486063
27. Archer SK, Shirokikh NE, Beilharz TH, Preiss T. Dynamics of ribosome scanning and recycling revealed by translation complex profiling. *Nature*. 2016; 535: 570–574. <https://doi.org/10.1038/nature18647> PMID: 27437580
28. Bazzini AA, Johnstone TG, Christiano R, Mackowiak SD, Obermayer B, Fleming ES, et al. Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO J*. 2014; 33: 981–993. <https://doi.org/10.1002/embj.201488411> PMID: 24705786
29. Han Y, Gao X, Liu B, Wan J, Zhang X, Qian SB. Ribosome profiling reveals sequence-independent post-initiation pausing as a signature of translation. *Cell Res*. 2014; 24: 842–851. <https://doi.org/10.1038/cr.2014.74> PMID: 24903108
30. Chen J, Tsai A, O’Leary SE, Petrov A, Puglisi JD. Unraveling the dynamics of ribosome translocation. *Curr Opin Struct Biol*. 2012; 22: 804–814. <https://doi.org/10.1016/j.sbi.2012.09.004> PMID: 23142574
31. Budkevich T V., Giesebrecht J, Behrmann E, Loerke J, Ramrath DJF, Mielke T, et al. Regulation of the mammalian elongation cycle by subunit rolling: A eukaryotic-specific ribosome rearrangement. *Cell*. 2014; 158: 121–131. <https://doi.org/10.1016/j.cell.2014.04.044> PMID: 24995983
32. Dunn JG, Weissman JS. Plastid: Nucleotide-resolution analysis of next-generation sequencing and genomics data. *BMC Genomics*. 2016; 17. <https://doi.org/10.1186/s12864-016-3278-x> PMID: 27875984
33. Popa A, Lebrigand K, Paquet A, Nottet N, Robbe-Sermesant K, Waldmann R, et al. RiboProfiling: a Bio-conductor package for standard Ribo-seq pipeline processing. *F1000Research*. 2016; 5: 1309. <https://doi.org/10.12688/f1000research.8964.1> PMID: 27347386
34. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010; 26: 841–842. <https://doi.org/10.1093/bioinformatics/btq033> PMID: 20110278
35. Zhang P, He D, Xu Y, Hou J, Pan BF, Wang Y, et al. Genome-wide identification and differential analysis of translational initiation. *Nat Commun*. 2017; 8. <https://doi.org/10.1038/s41467-017-01981-8> PMID: 29170441
36. Beaupere C, Wasko BM, Lorusso J, Kennedy BK, Kaerberlein M, Labunskyy VM. CAN1 Arginine Permease Deficiency Extends Yeast Replicative Lifespan via Translational Activation of Stress Response Genes. *Cell Rep*. 2017; 18: 1884–1892. <https://doi.org/10.1016/j.celrep.2017.01.077> PMID: 28228255
37. Shi Z, Fujii K, Kovary KM, Genuth NR, Röst HL, Teruel MN, et al. Heterogeneous Ribosomes Preferentially Translate Distinct Subpools of mRNAs Genome-wide. *Mol Cell*. 2017; 67: 71–83.e7. <https://doi.org/10.1016/j.molcel.2017.05.021> PMID: 28625553
38. Hanson G, Collier J. Codon optimality, bias and usage in translation and mRNA decay. *Nat Rev Mol Cell Biol*. 2018; 19: 20–30. <https://doi.org/10.1038/nrm.2017.91> PMID: 29018283
39. Lauria F, Tebaldi T, Lunelli L, Struffi P, Gatto P, Pugliese A, et al. RiboAbacus: a model trained on polyribosome images predicts ribosome density and translational efficiency from mammalian transcriptomes. *Nucleic Acids Res*. 2015; <https://doi.org/10.1093/nar/gkv781> PMID: 26240374
40. Raveh A, Margaliot M, Sontag ED, Tuller T. A model for competition for ribosomes in the cell. *J R Soc Interface*. 2016; 13: 20151062. <https://doi.org/10.1098/rsif.2015.1062> PMID: 26962028