

# A systematic computational analysis of the rRNA–3' UTR sequence complementarity suggests a regulatory mechanism influencing post-termination events in metazoan translation

JOSEF PÁNEK,<sup>1</sup> MICHAL KOLÁŘ,<sup>2</sup> ANNA HERRMANNOVÁ,<sup>3</sup> and LEOŠ SHIVAYA VALÁŠEK<sup>3</sup>

<sup>1</sup>Laboratory of Bioinformatics, Institute of Microbiology of the Academy of Sciences of Czech Republic, 14220 Prague, Czech Republic

<sup>2</sup>Laboratory of Genomics and Bioinformatics, Institute of Molecular Genetics of the Academy of Sciences of Czech Republic, 14220 Prague, Czech Republic

<sup>3</sup>Laboratory of Regulation of Gene Expression, Institute of Microbiology of the Academy of Sciences of Czech Republic, 14220 Prague, Czech Republic

## ABSTRACT

Nucleic acid sequence complementarity underlies many fundamental biological processes. Although first noticed a long time ago, sequence complementarity between mRNAs and ribosomal RNAs still lacks a meaningful biological interpretation. Here we used statistical analysis of large-scale sequence data sets and high-throughput computing to explore complementarity between 18S and 28S rRNAs and mRNA 3' UTR sequences. By the analysis of 27,646 full-length 3' UTR sequences from 14 species covering both protozoans and metazoans, we show that the computed 18S rRNA complementarity creates an evolutionarily conserved localization pattern centered around the ribosomal mRNA entry channel, suggesting its biological relevance and functionality. Based on this specific pattern and earlier data showing that post-termination 80S ribosomes are not stably anchored at the stop codon and can migrate in both directions to codons that are cognate to the P-site deacylated tRNA, we propose that the 18S rRNA–mRNA complementarity selectively stabilizes post-termination ribosomal complexes to facilitate ribosome recycling. We thus demonstrate that the complementarity between 18S rRNA and 3' UTRs has a non-random nature and very likely carries information with a regulatory potential for translational control.

**Keywords:** metazoan 18S rRNA–mRNA 3' UTRs complementarity; large-scale data set; statistics; translation termination; ribosomal recycling

## INTRODUCTION

Complementarity between two DNA/RNA sequences, formed by nucleotide base-pairing, is the basic principle of the processes of DNA replication and transcription, which allows cells to copy genetic information from one generation to another, to find and repair damages of the genetic code, and to build living organisms based on their specific genetic information. Complementarity also mediates regulatory functions in cells by base-pairing between regulated (e.g., mRNAs) and regulatory RNA molecules such as noncoding RNAs, antisense RNAs, miRNAs, and siRNAs (Katayama et al. 2005; He et al. 2008; Faghihi et al. 2010; Kosaka et al. 2013), when the interaction influences the function of the regulated molecule. Regulatory effects are also achieved by base-pairing within single DNA/RNA molecules that creates double-strand-like structures such as hairpin loops, junctions, bulges, or internal loops that have specific regulatory

functions. Well-studied examples of such regulatory structures are transcription terminators (Richardson and Richardson 1996) and riboswitches (Nahvi et al. 2002).

However, there exists a particular type of RNA/RNA complementarity between ribosomal RNAs (rRNAs) and 5' and 3' untranslated regions (UTRs) of mRNAs, the function of which is still only poorly explored, with the exception of the bacteria and archaea-specific Shine–Dalgarno sequence (Shine and Dalgarno 1975). First it should be mentioned that the mRNA's UTRs have been shown to be important for translation regulation by themselves as they contain various regulatory elements that influence localization, stability, export, and translational efficiency of mRNAs. Elements in 5' UTRs with regulatory function include, for example, closed-loops (Kang and Han 2011), iron-response elements (Araujo et al. 2012), and upstream short open reading frames

Corresponding authors: [panek@biomed.cas.cz](mailto:panek@biomed.cas.cz), [valasek@biomed.cas.cz](mailto:valasek@biomed.cas.cz)  
Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.056119.116>.

© 2016 Pánek et al. This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see <http://rnajournal.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

(for review, see Wethmar 2014). In the case of the 3' UTRs, the regulatory elements are, for example, microRNA response elements (MREs), AU-rich elements (AREs), and the poly(A) tail (for review, see Barrett et al. 2012). On top of these regulatory roles, another level of the UTR-mediated regulation can be achieved by their direct base-pairing with rRNAs, which was observed already a long time ago, due to their abundance and ubiquity found in both 5' and 3' UTRs (Matveeva and Shabalina 1993; Mauro and Edelman 1997).

From the historical perspective, the rRNA–5' UTR complementarity was originally hypothesized to promote the selective recruitment of diverse mRNAs to initiating ribosomes in order to influence the efficiency of their translation (Tranque et al. 1998; Mauro and Edelman 2002, 2007). This hypothesis was supported experimentally for a 9-nucleotide (nt) element in the mouse *Gtx* homeodomain mRNA (Dresios et al. 2006). Later, the idea of the regulatory role of the mRNA–rRNA complementarity was, based on statistical analysis of a large number of eukaryotic mRNAs, extended by a suggestion that the rRNA–5' UTR complementarity modulates the rate and processivity of scanning of the 40S subunit through the mRNA leader (Panek et al. 2013). Among experimentally well-documented examples of base-pairing between mRNA UTRs and rRNA are, for example, the TURBS motif in the Calicivirus mRNA, the base-pairing of which with 18S rRNA is absolutely essential for the termination/reinitiation mechanism governing the expression of this mRNA in an eukaryotic cell (Luttermann and Meyers 2009), and the hexanucleotide CAR-NBA consensus sequence immediately following the stop codon of several viral and cellular genes, mediating so-called programmed stop codon readthrough by base-pairing with 18S rRNA to extend the C-termini of the encoded proteins (Namy et al. 2001, 2004).

Unlike the 5' UTR–rRNA complementarity, the 3' UTR–rRNA complementarity has never been systematically explored. The only attempt of its characterization, to our knowledge, was a computational search that reported an occurrence of 18S rRNA sequences complementary to 3' UTRs of several mRNAs in a few eukaryotic species (Mauro and Edelman 1997); as this search focused primarily on 5' UTR–rRNA complementarity, a subsequent analysis of this complementary sequence in 18S rRNA and mRNAs either computational or experimental was not carried out.

Since the 3' UTR–rRNA complementarity has been vastly unexplored, we subjected it to a robust, systematic search for using large-scale data sets of the full-length annotated 3' UTR sequences of 14 eukaryotes. We found statistically significant and evolutionarily conserved patterns of complementarity specifically occurring between 18S rRNAs and only the first 50 nt of the 3' UTRs of only metazoan mRNAs. In addition, this conserved complementarity pattern was also found to be spatially restricted around the mRNA entry pore on the 40S ribosome. Based on these results, we hypothesize that the 3' UTR base-pairing with 18S rRNA selectively stabilizes

the post-termination 80S ribosome at the stop codon, perhaps to expedite ribosomal recycling to stimulate overall translational rates.

## RESULTS AND DISCUSSION

### Distribution of rRNA–mRNA 3' UTR sequence complementarity in eukaryotic 3' UTRs

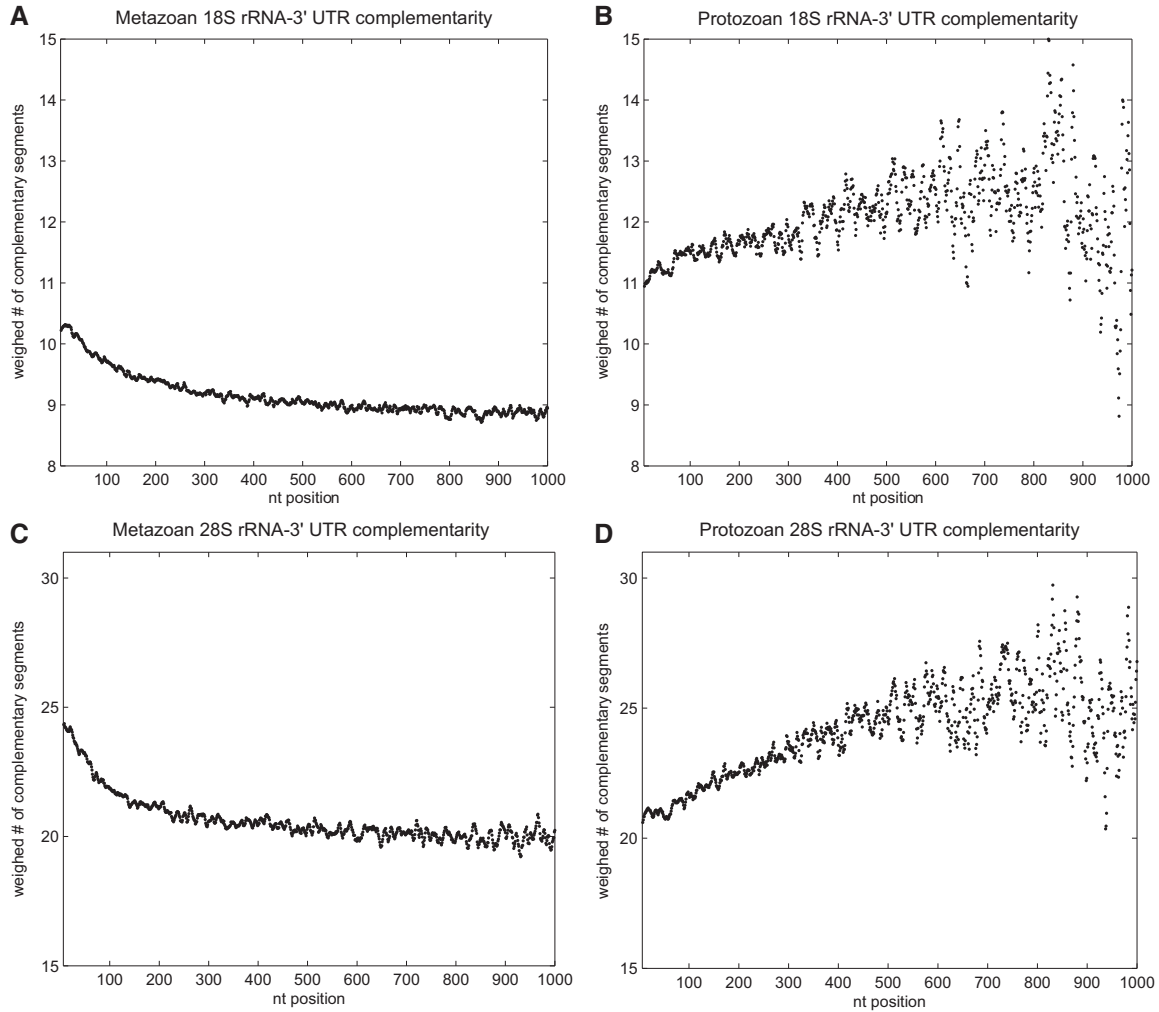
We first collected 27,646 sequences of the full-length 3' UTRs and sequences of both 18S and 28S rRNAs of 14 eukaryotic species and computationally determined all segments within 3' UTR sequences that (i) exhibited reverse complementarity to both 18S and 28S rRNAs without gaps and (ii) were at least 5 nt long. We plotted counts of segments complementary between 3' UTRs and rRNA at each nucleotide position in all 3' UTRs (Fig. 1). The counts were normalized to the number of 3' UTRs. The plots revealed a global maximum of complementarity to both 18S (Fig. 1A,B) and 28S (Fig. 1C,D) rRNAs specifically located at the first 50 nt in metazoans (Fig. 1A,C).

The global maximum of complementarity to rRNAs found for metazoans was not found in protozoan 3' UTRs (Fig. 1B, D). Instead, a local maximum centered around the 50th nucleotide was formed. Beyond the 50th nucleotide, the counts increased steadily toward the end of 3' UTR sequences; see further below.

### Distribution of rRNA–mRNA 3' UTR sequence complementarity in eukaryotic rRNAs

Based on the results described above, we further focused only on the first 50 nt of 3' UTRs. We asked how the complementarity of the first 50 nt of the 3' UTRs is distributed within the rRNA sequences. As the distribution of complementarity in 3' UTRs differed principally for metazoa and protozoa, the following analysis was performed separately for protozoans and metazoans to see if the difference observed for 3' UTR is also reflected in the distribution of complementarity in rRNAs.

We plotted the weighed counts of complementary segments between rRNAs and the first 50 nt of 3' UTRs at each nucleotide position in both 18S and 28S rRNAs of the individual species (Supplemental Fig. S1A,B, black curves). In rRNAs, the counts of complementary segments showed principally different distribution than in 3' UTRs (shown in Fig. 1). They did not form decreasing or increasing curves with either global or local maximum in rRNA sequences (Supplemental Fig. S1). Instead, the curves of their distribution exhibited several peaks, i.e., local maxima. These maxima identified the regions with enriched complementarity to 3' UTRs. However, although their complementarity was enriched, it was necessary to define a criterion that would distinguish biologically relevant regions of increased complementarity from random background. We did that by



**FIGURE 1.** Sequence complementarity between the mRNA 3' UTRs and 18S and 28S rRNAs within the metazoan (A,C) and protozoan (B,D) 3' UTRs. The sequence complementarity is shown as counts of complementary sequences (*y*-axis) at each nucleotide position (*x*-axis) in all 3' UTRs of all analyzed species. The counts were normalized to numbers of 3' UTR sequences that were long enough to include a given nucleotide position. The greater the nucleotide position in each diagram, the lesser the number of nucleotides available at that position for computation, because the number of 3' UTRs with the equal or bigger length relative to that position decreases. The minimal length of all analyzed 3' UTR sequences was 50 nt, and as such all tested 3' UTRs have their first 50 nt included in all calculations of sequence complementarity for the first 50 positions. In the case of the protozoan 3' UTRs, the noise caused by the small number of available sequences took effect for positions >300 (showed by increasingly dispersed character of the curve at positions >300 [B,D]) as they were generally a lot shorter than those of metazoans. Please note that the reason for approximately two times higher maximum of complementarity in 28S rRNA versus 18S rRNA is the length of the 28S rRNA, which is in metazoans ~2–2.5 longer than 18S rRNA. The counts of complementary segments were not normalized to the length of individual rRNAs since the purpose of this analysis was not a comparison between the two types of rRNA.

estimating their statistical significance using the density of complementarity function (see Materials and Methods for details).

To define the density of complementarity, we took advantage of the fact that long nucleotide sequences usually interact with each other through base-pairing between several shorter, continuous segments of reverse complementarity, separated from each other by a few nucleotides. Typically, these short segments are concentrated in both interacting sequences in base-pairing regions, thus increasing the density of complementarity (for definition, see Materials and Methods). The significant regions were defined as the regions

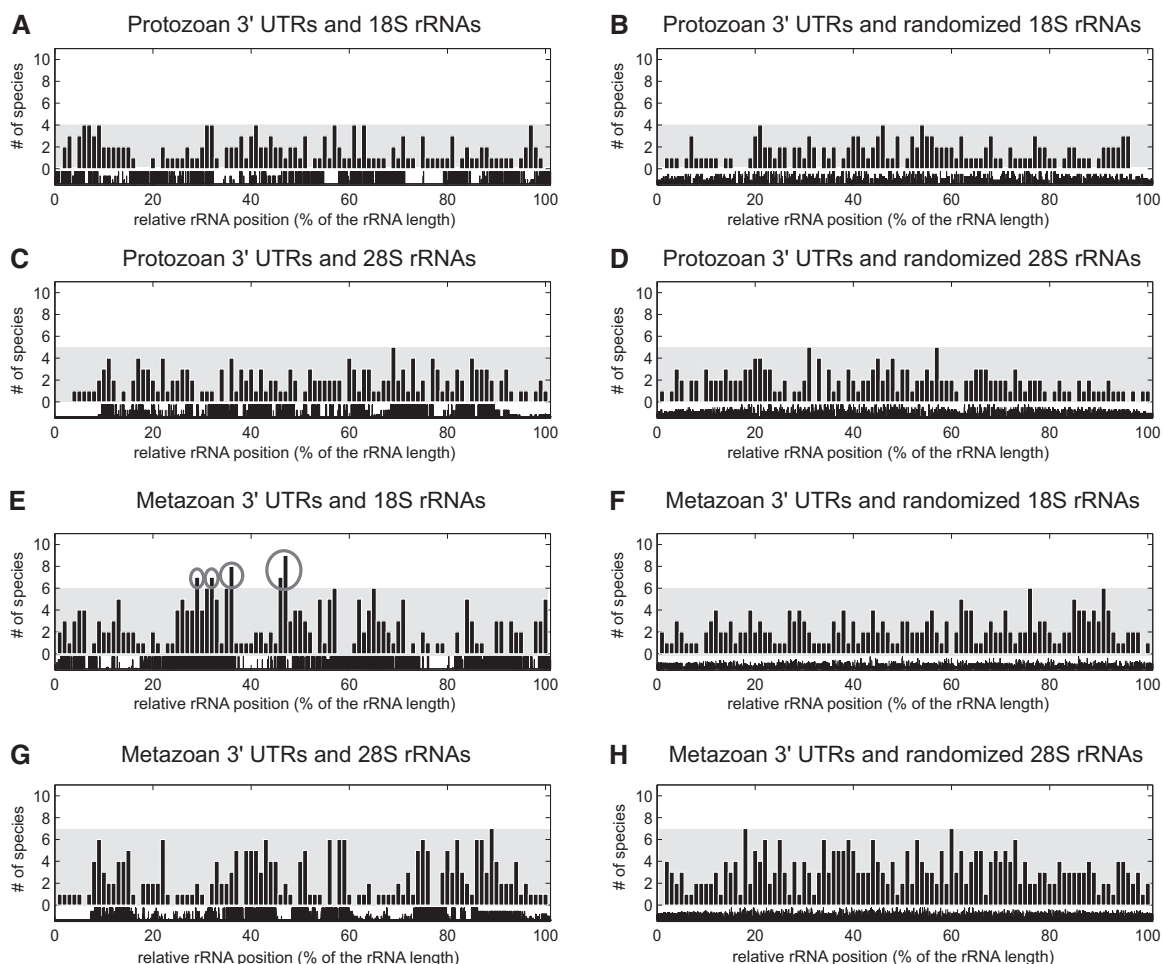
in which the density of complementarity increased probability of interaction significantly higher compared to the rest of the sequences. The existence of statistically significant regions indicates that the interactions in these regions are possible also in reality, and the location of these regions shows where the interaction might occur. Although this method lacks the accuracy of the thermodynamic computations (Dimitrov and Zuker 2004), it is fast and allowed us to evaluate complementarity and predict potential interactions for large numbers (thousands) of pairs of sequences, without losing potential to identify the regions of significant complementarity.

The  $q$ -value of the density of complementarity across the rRNA sequence is shown in Supplemental Figure S1 by the blue curve. The regions on rRNA sequences with the statistically significant density of complementarity ( $q \leq 0.05$ , green line in Supplemental Fig. S1) were those that we consider as biologically relevant (Supplemental Tables SI–IV). These regions have the potential to form interactions with the first 50 nt of 3' UTRs.

### Evolutionary conservation of the statistically significant sequence complementarity between rRNAs and mRNA 3' UTRs

The statistically significant regions formed patterns of 3' UTR–rRNA complementarity on the rRNA sequences that were localized in similar positions across all species

(Supplemental Fig. S1). To examine whether the similarity indicated evolutionary conservation, we aligned the patterns of individual species shown in Supplemental Figure S1A,B (see Materials and Methods). The patterns for protozoan and metazoan rRNAs were aligned separately to respect the observed differences between these two evolutionarily relatively distant groups. The aligned patterns are shown in Figure 2A, C,E,G. The heights of the individual bars in the plots represent the number of species with statistically significant complementarity at the given nucleotide position on the rRNA sequences. Since rRNAs from different species have naturally varying lengths, the positions of the bars on rRNAs were not expressed in the number of nucleotides but in percentages, where 100% equals the full length of the rRNA sequence. The number of nucleotides represented by each bar is given by the length of rRNAs divided by 100; e.g., for the metazoan



**FIGURE 2.** Evolutionary conservation of statistically significant sequence complementarity between rRNAs and the first 50 nt of 3' UTRs. Five protozoan 18S and 28S rRNAs (A–D), nine metazoan 18S rRNAs (E,F), and eight metazoan 28S rRNAs (G,H) were analyzed for evolutionary conservation. The horizontal axis of each panel shows relative positions in percent of the length of rRNAs. The vertical axis shows the number of species with statistically significant complementarity at a given position within the native (A,C,E,G) and representative randomized (B,D,F,H) rRNA sequences. The randomized rRNA sequences were used for statistical estimation of cut-off lines that separate a nonspecific computational and/or biological noise from evolutionarily conserved complementarity (see Materials and Methods). The cut-off level is shown as the gray zone in each panel. Only bars above the gray zone indicate evolutionarily conserved complementarity to 3' UTRs, indicated in the figure by light gray ellipses. Black profiles at the bottom of each panel indicate evolutionary conservation of rRNA sequences.

18S rRNAs (Fig. 2E), this number varied from 18 for the shortest 18S rRNA of *Ciona intestinalis* to 20 for the longest 18S rRNA of *Drosophila melanogaster*. Naturally, the profiles in Figure 2 included a nonspecific computational and biological bias due to which some bars in the profiles could have a high score, but carried no biologically relevant information. To control for these nonspecific signals, we estimated sequence specific cut-off values from the bar profiles of rRNAs with randomized nucleotide sequences (nucleotide composition of the sequences was preserved during randomization). The randomization was repeated 100 times for each rRNA to achieve robust statistical estimates (for details, see Materials and Methods). For these 100 randomized profiles for each and every rRNA, we estimated a height of the bars, above which the bars were expected to contain less than one false positive at the 0.05 level of probability. Then, we considered only the positions in the native rRNA sequences for which the complementarity was greater than that computed for the randomized data set—only these were considered as evolutionarily conserved with a potential to be biologically relevant.

The cut-off values for protozoan 18S and 28S rRNAs were set to four and five, respectively (Fig. 2B,D), and to six and seven for metazoan 18S and 28S rRNAs, respectively (Fig. 2F,H; note that Fig. 2B,D,F,H are representative plots chosen from the 100 randomized controls to illustrate the cut-off values). Comparing the profiles of native rRNAs to the corresponding controls showed no bars higher than the cut-off lines in both protozoan rRNAs and metazoan 28S rRNA, indicating that there were no complementary regions with statistically significant evolutionary conservation (cf. Fig. 2A and B, C and D, and G and H). In contrast, the metazoan 18S rRNAs showed four regions complementary to 3' UTRs with significant evolutionary conservation, indicated by five bars exceeding the corresponding cut-off line; one of these four regions was formed by two adjacent bars representing a continuous region of the 18S rRNA sequence. The regions included seven, seven, eight, and nine metazoan species (out of nine) (Fig. 2E). Together, these findings suggest that evolutionarily conserved, statistically significant complementarity between the first 50 nt of 3' UTRs and rRNA sequences exists only in metazoan 18S RNAs but not in metazoan 28S rRNA or both protozoan rRNAs.

### Evolutionary conservation of the pattern of the metazoan 18S-3' UTR complementarity in the 18S rRNA structure

Next, we wished to determine whether the distribution of the metazoan 18S-3' UTR complementarity conserved in the primary sequences was also conserved in secondary structures, because the first does not imply the other. To do that, we collected models of all 18S rRNA secondary structures that were available among the analyzed species, namely for *D. melanogaster*, *Xenopus laevis*, *R. norvegicus*, and *Homo sapiens*

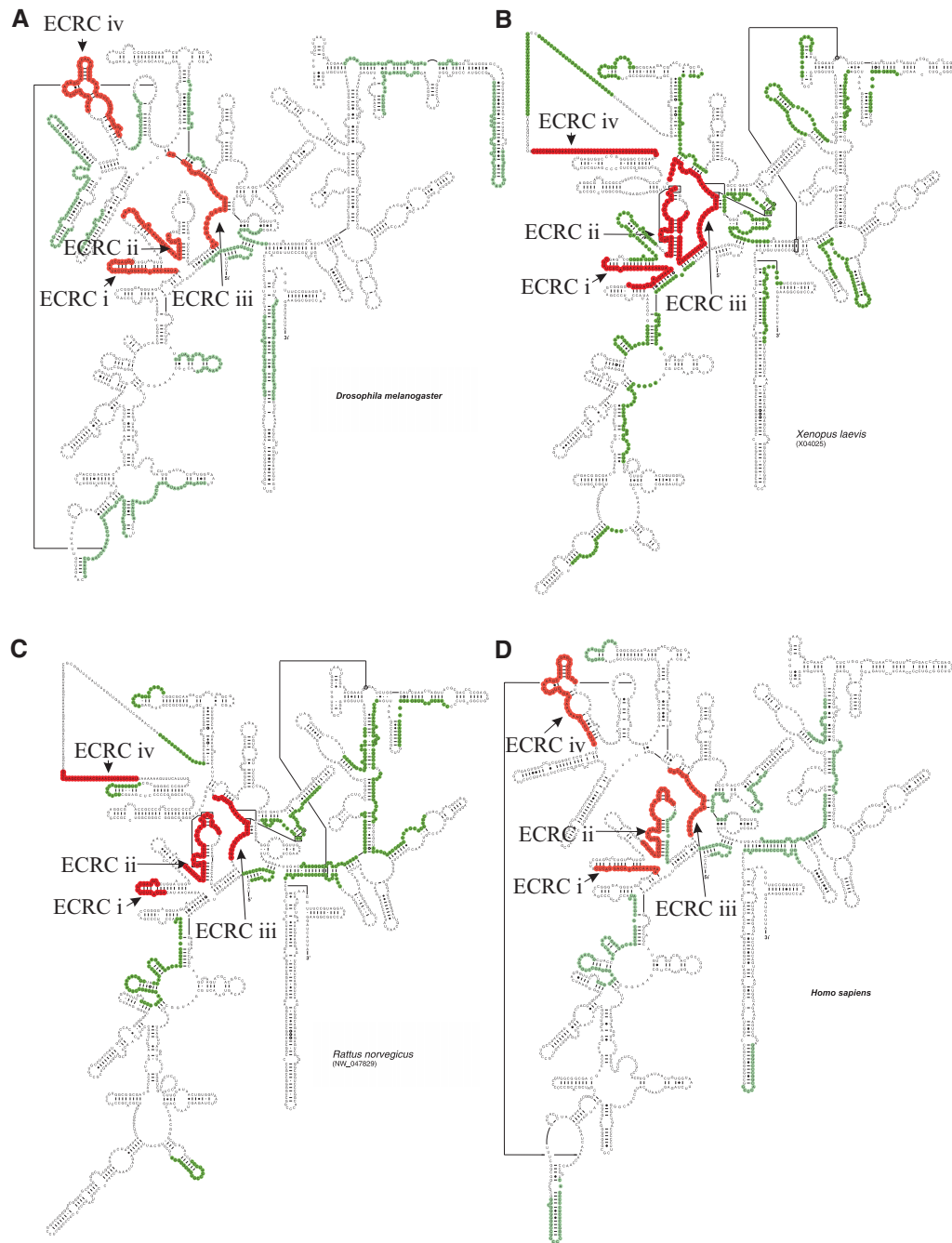
(Cannone et al. 2002), and projected all the regions of statistically significant complementarity (schematically shown in Supplemental Fig. S1A and listed in detail in Supplemental Table SII) onto their predicted secondary structures (Fig. 3), including the regions that did not appear to be evolutionarily conserved (Fig. 2A,C,E,G), for comparison. The majority of these regions was relatively randomly distributed over the secondary structures of the individual rRNAs, with the exception of only four regions of complementarity within nine metazoan 18S rRNAs with conserved sequence location. These locations clustered across the species at sites of similar secondary structure. We named them evolutionarily conserved regions of complementarity (ECRCs).

The ECRCs were conserved not only structurally but also in their primary sequences, as they occurred only in the strictly conserved specific segments of metazoan 18S rRNAs (see the conservation profile and multiple sequence alignment consensus at the bottom of Fig. 2E). Although the expansion segments (ESs) are generally known to form the least conserved parts of rRNAs (shown in, e.g., Ben-Shem et al. 2011), the only ECRC that was located in a ES (see the next section) fell into the only conserved part of the ES6 sequence (Fig. 2E).

The reasons why we found the statistical evidence of evolutionary conservation in metazoans, but not in protozoans, might be (i) the protozoan signal was weak and was under the threshold of our search, as we might not have enough protozoan data available; (ii) the ECRCs gained their function later on during evolution of metazoans, as a response to increasing complexity of life and a demand for more sophisticated regulatory mechanisms. Indeed, while neither statistical significance nor evolutionary conservation was detected in protozoa, the sequences of all four ECRCs were unambiguously discernible in primary sequences of all tested protozoan 18S rRNAs and also in yeasts (not included in the analysis), which suggests that the ECRCs started to develop in protozoans, although they did not give a sufficiently strong signal in our analysis. Whether the reason is the insufficient data or weak evolutionary conservation, it can be addressed by repeating the search with more protozoan sequences when available. Nevertheless, unlike in protozoa, the indicated sequence and structural evolutionary conservation of ECRCs in metazoa argue for their potential functional importance.

### Mapping ECRCs onto the 3D structure of the 40S ribosomal subunit

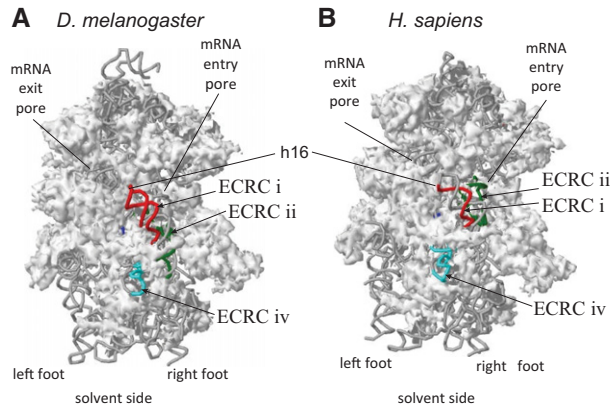
To be able to predict a hypothetical function of the ECRCs, we projected all four conserved ECRCs onto the recently solved tertiary structures of the *D. melanogaster* and *H. sapiens* small ribosomal subunits (Anger et al. 2013). As shown in Figure 4, three of the four ECRCs occurred within the surface-exposed segments on the solvent side of the 40S ribosome in the vicinity of the mRNA entry pore. Specifically, ECRC ii (shown in green in Fig. 4) occurred directly at the



**FIGURE 3.** Distribution of regions with statistically significant complementarity to 3' UTRs on available secondary structures of metazoan 18S rRNAs. Secondary structures of 18S rRNAs of *D. melanogaster* (A), *X. laevis* (B), *R. norvegicus* (C), and *H. sapiens* (D) are shown. All initially identified, statistically significant regions of complementarity between the first 50 nt of 3' UTRs and 18S rRNAs of these four species were projected onto their 2D structures and visualized in green. The regions that exhibit evolutionary conservation are highlighted in red and designated as ECRCs i–iv.

entry of the mRNA binding channel at positions 587–628 (*H. sapiens*) and 531–563 (*D. melanogaster*). Immediately following was the ECRC i (in red) mapping onto the helix 16 adjacent to the mRNA entry pore at positions 519–539 (*H. sapiens*) and 478–509 (*D. melanogaster*). Below these two, ECRC iv (in cyan) is embedded in the expansion segment ES6 at positions 816–843 (*H. sapiens*) and 848–867

(*D. melanogaster*). Lastly, ECRC iii, not visible in the projection in Figure 4, was located near the A site on the interface side of the 40S ribosome at positions 655–679 (*H. sapiens*) and 609–639 (*D. melanogaster*) (Supplemental Fig. S2, ECRC iii). The major part of this ECRC is most probably buried under the surface of the 40S ribosomal subunit. Structural and spatial conservation of the ECRCs together



**FIGURE 4.** Evolutionarily conserved regions of statistically significant sequence complementarity between 18S rRNA and 3' UTRs (ECRCs) cluster around the mRNA entry pore on the 40S ribosomal subunit. ECRCs were projected on the 3D structures of small ribosomal subunits of *D. melanogaster* (A) and *H. sapiens* (B). ECRCs are color-coded as follows: ECRC i in red, ECRC ii in green, and ECRC iv in cyan. ECRC iii is not visible in this solvent-exposed subunit projection but can be seen in Figure S2. Ribosomal proteins are colored in white.

with their specific occurrence only in metazoan 18S rRNA indicates that the 18S rRNA-3' UTR complementarity did not evolve accidentally; in other words, it further argues for its metazoan-specific regulatory role in translation, which we discuss below.

### Does the complementarity between 18S rRNA ECRCs and the first 50 nt of metazoan mRNA 3' UTRs selectively stabilize post-termination ribosomal complexes to facilitate ribosome recycling?

The fact that three out of four ECRCs occur right at or below the mRNA entry pore and are surface-exposed suggests that they could theoretically interact with the first 50 nt of mRNA 3' UTRs, where the global maximum of complementarity to 18S rRNAs was found (Fig. 1). Taking into account that the 80S footprint is ~30 nt long (Ingolia et al. 2009), it implies that this interaction could occur only shortly before or during the termination phase, when both ECRCs and the first 50 nt coexist within the steric reach of each other. It is important to note that when the ribosome terminates, approximately the first 9 nt of 3' UTR past the stop codon are already buried within the mRNA binding channel (Szamecz et al. 2008; Munzarová et al. 2011). However, the remaining 41 nt out of the first 50 nt that have not entered the mRNA entry pore yet should be available for contacting ECRCs i, ii, and iv (Fig. 4). The first 9 nt could instead be contacted by those sections of the remaining ECRC iii that might still be surface-exposed, since it occurs near the A site within the mRNA binding channel (Supplemental Fig. S2).

In theory, stable 18S rRNA-3' UTR base-pairing could, for example, stabilize post-termination complexes (post-TCs) on mRNA following the polypeptide release to facilitate ribo-

somal recycling and/or to prevent their migration far into 3' UTR. Indeed, recent results obtained using the mammalian *in vitro* reconstituted system showed that post-termination 80S ribosomes are not stably anchored at the stop codon and can migrate in both directions to codons that are cognate to the P-site deacylated tRNA (Skabkin et al. 2013). Such instability may lead to an undesirable accumulation of nonrecycled aberrant ribosomal complexes in the vicinity of the stop codon. In addition, post-termination ribosome migration could potentially promote aberrant reinitiation events, which would generally reduce translational fidelity. The predicted interaction between ECRCs and the first 50 nt of the 3' UTRs of metazoan mRNAs thus might come into play as an accessory stabilizing mechanism.

One could argue that partial base-pairing within the 18S rRNA itself in two of four ECRCs (i and ii) would make them unavailable for base-pairing with 3' UTRs. However, both of them are part of the h16-h18 region that is known to be involved in the dynamic structural changes of the 40S entry pore during initiation as well as termination phases (Passmore et al. 2007; Ben-Shem et al. 2011; des Georges et al. 2014; Hussain et al. 2014; Llacer et al. 2015). Hence, structural rearrangements that these two helices undergo during translation could easily make their sequences available for base-pairing with the 3' UTRs upon termination.

In any case, we hypothesize that the evolutionarily conserved spatial distribution of 18S rRNA ECRCs and the global maximum of complementarity matching the first 50 nt of metazoan 3' UTRs could underlie a mechanism that has evolved to help reduce the migration of post-termination ribosomes in an mRNA-specific manner, determined by the degree of complementarity between a given 3' UTR and the ECRCs. Notably, recent ribosome profiling experiments revealed that 80S ribosomes can gain access to 3' UTRs of mRNAs by a mechanism that does not involve ongoing translation and accumulate on them in the absence of the surveillance factor Dom34 (Guydosh and Green 2014). Interestingly, however, accumulation of ribosomes on 3' UTRs in the absence of Dom34 has been found to be restricted to only ~10% of yeast cellular mRNAs. Clear differences in ribosomal accumulation were especially apparent on mRNAs derived from duplicated ribosomal protein genes, in which the encoded proteins are nearly identical but 3' UTR sequences are divergent. These findings further support the idea that the nature of 3' UTRs has a significant impact on ribosomal movement and accumulation.

The question remains, why would the terminating ribosome with eRF1 bound in the A site require additional support to complete the translational cycle by splitting the post-TCs?

One reason could be that the peptide release step is coupled with ribosomal recycling and both are promoted by ABCE1/RLI1 (Pisarev et al. 2010; Shoemaker and Green 2011), which is half as abundant as eRF1 (Ghaemmghami et al. 2003). This implies that some termination events likely

occur without ABCE1/RLI1, which may lead to a delay between peptide release and recycling and allow for the aforementioned undesirable migration (Pisarev et al. 2010; Skabkin et al. 2013). Perhaps the base-pairing between the ECRCs and the mRNA 3' UTR reduces such migration to a certain degree and thereby facilitates recycling by allowing ABCE1/RLI1 sufficient time to bind to the post-TCs.

An alternative option is based on a recent structural observation showing that when the eRF3•GDP is ejected from the termination complex, eRF1 drastically changes its conformation such that the central domain with the GGQ motif extends toward the peptidyl transferase center (PTC). Binding of ABCE1/RLI1 appears to stabilize this fully extended active conformation of eRF1, thereby stimulating peptide release. Interestingly, the NTD of eRF1 with the conserved (TAS) NIKS motif—responsible for the stop codon recognition—appears to disengage the A site codon in this complex, indicating that codon engagement may not be required at this stage for peptide release (Preis et al. 2014). It may be required, however, for preventing migration of this stage post-TC away from the stop codon, implying that ECRCs thus may have evolved to partially compensate for the loss of this stabilization effect.

## Conclusion

Our statistical, computational analysis revealed that one of the types of DNA/RNA complementarity—between mRNAs and rRNAs, which is otherwise rather ubiquitous and abundant—has a nonrandom, highly specific character that provides it with a capability to form or at least be a part of a specific translational control mechanism. We found that specifically only the first 50 nt of metazoan 3' UTR sequences past the stop codon have the potential to base pair with several complementary regions of 18S rRNA (ECRCs) clustering around the mRNA entry channel of the 40S ribosome. Based on these findings, we proposed that the complementarity between ECRCs and the first 50 nt of 3' UTRs selectively stabilizes post-termination ribosomal complexes to facilitate ribosome recycling. In our previous study, we detected specific complementarity between 5' UTR sequences and several 18S rRNA surface-exposed sticky regions (SRs) proposed to modulate the rate and processivity of scanning of the 40S subunit through the mRNA leader (Panek et al. 2013). Taken together, we envisage the ribosome as a sophisticated macromolecular machine, which not only ensures that the coding sequences are read and turned into proteins, but which also turns the “message” of the 5' and 3' UTRs, depending on the degree of their complementarity to SRs or ECRCs, respectively, into a specific regulatory output at a given stage of protein synthesis. In a broader perspective, our work shows three principal facts: (i) a meaningful mRNA–rRNA complementarity occurrence analysis requires thorough statistics based on a large-scale sequence data set including many different mRNAs from a broad range of species, (ii) the comple-

mentarity pattern between 18S rRNA and mRNA's UTRs is nonrandom, creating specific mechanisms with defined functions in gene translation, and (iii) the UTR–rRNA complementarity may further expand the repertoire of regulatory elements/features “encoded” by the otherwise noncoding untranslated regions.

## MATERIALS AND METHODS

### Data sets

To ensure statistical robustness of our analysis, we collected sequences of mRNA 3' UTRs and rRNAs only from those species for which at least 500 sequences of 3' UTRs were available at the UTRdb database (Grillo et al. 2009) at the onset of our study; in particular from: *Homo sapiens*, *Caenorhabditis elegans*, *Ciona intestinalis*, *Gallus gallus*, *Danio rerio*, *Drosophila melanogaster*, *Bos taurus*, *Rattus norvegicus*, *Xenopus laevis*, *Tetrahymena thermophila*, *Plasmodium vivax*, *Trichomonas vaginalis*, *Trichoplax adhaerens*, and *Monosiga brevicollis*. These sequences were subsequently filtered for duplicates and ambiguity. Altogether, we obtained 2400 unique 3' UTR sequences longer than 50 nt for the first nine species and 1683, 556, 1226, 556, and 2025 sequences for the remaining five species, respectively. The rRNA sequences were collected from the Silva (Quast et al. 2012) and GenBank databases. The 28S rRNA for *G. gallus* was not available at the time of writing the manuscript and therefore the analyses for 28S rRNAs were restricted to 13 species only.

### Sequence complementarity

Our definition of sequence complementarity considered in our analysis derives from the following assumption: Two long sequences (of tens or even hundreds of nucleotides) interact through base-pairing of multiple relatively short (typically up to 10 nt) reverse complementary segments separated by gaps of an arbitrary number of nucleotides. Identification of regions with enriched abundance of such short segments of reverse complementarity between rRNAs and 3' UTRs by computational analysis would (i) indicate that rRNAs and 3' UTRs might interact in vivo and (ii) localize these potentially interacting regions on the 2D and 3D maps of rRNAs as well as within the primary sequence of the corresponding 3' UTRs.

To identify such regions, we computationally determined all segments within both 18S and 28S rRNA sequences that were (i) reverse complementary to 3' UTRs without gaps and (ii) at least 5 nt long. This minimal length limit was chosen due to the fact that the probability of finding random tetranucleotide segments that are complementary between two RNA molecules is too high to provide any biologically relevant information. The upper length limit was not set. If a shorter complementary segment was fully contained within a longer one, only the longer segment was used for analysis.

### Statistical significance of sequence complementarity between rRNAs and mRNAs

To identify statistically significant complementarity between rRNAs and mRNAs, we developed the following procedure: First, for each nucleotide *i* on an rRNA sequence, complementary segments (identified as described above) were counted within a 20-nt sliding



window centered on  $i$ . Second, using the counts of the complementary segments, the density of complementarity between the rRNA sequence  $s_R$  and the 3' UTR sequences  $s_U$  within the window  $W_i$  centered on  $i$  on the sequence  $s_R$  (an intensive quantity under a null model of random sequences) was calculated as follows:

$$d_i^{\text{within}}(s_R) = \frac{1}{N_U} \sum_U \frac{\text{number of complementary segments within } W_i}{\text{length of } W_i \times \text{length of } s_U}.$$

Also, the density of complementarity outside the  $W_i$ , i.e., in the rest of the sequence  $s_R$ , was computed:

$$d_i^{\text{outside}}(s_R) = \frac{1}{N_U} \sum_U \frac{\text{number of complementary segments outside } W_i}{(\text{length of } s_R - \text{length of } W_i) \times \text{length of } s_U}.$$

In both equations, we summed over all 3' UTR sequences  $s_U$  and normalized the density by their number  $N_U$ . The density profile of complementarity for ribosomal RNAs is depicted by black curves in Supplemental Figure S1.

The densities of complementarity were computed using the same number of 3' UTR sequences for each species to ensure similar statistical power. We used 556 3' UTR sequences for protozoans, and 2400 3' UTRs for metazoans. Since we had more than 556 3' UTR sequences for *T. vaginalis*, *T. thermophile*, and *M. brevicollis* and wished to utilize information held in all available sequences, we divided the full sets of these 3' UTRs into several subsets, each containing 556 sequences. Subsequently, we computed distributions of regions with statistically significant complementarity in rRNAs for all these subsets of the latter three protozoans and used them in our test for evolutionary conservation (see below) together with computed distributions of the remaining two protozoans with only 556 3' UTRs each. Individual distributions of complementarity of all subsets were highly similar for each organism and, therefore, produced the same evolutionary conservation pattern (data not shown). This finding strongly suggests that the selected number of sequences (556) in the data sets of all five protozoan species that are presented here can be considered as representative of the observed rRNA-3' UTR complementarity. Similarly, random subsets of metazoan 3' UTR sequences with 2400 entities each produced highly similar or identical distributions of complementarity within rRNA sequences for various subsets of 3' UTRs for individual species (data not shown).

The density profile of complementarity (Supplemental Fig. S1) made it possible to identify the regions of statistically significant enrichment of complementarity in a given rRNA sequence by comparing the densities of complementarity in ( $d_i^{\text{within}}$ ) and outside of ( $d_i^{\text{outside}}$ ) the window  $W_i$  using a one-sided Student's  $t$ -test (assuming equal variances), with the resulting  $P$ -values corrected for multiple testing by  $q$ -value (Storey 2003). Statistical significance was estimated at the level of 0.05. This way, all windows in which statistically significant density of complementary sequences was detected (i.e., all those with  $q_i < 0.05$ ) define regions of statistically significant rRNA-mRNA sequence complementarity. The  $q$ -value profiles are depicted by blue curves in Supplemental Figure S1. Typically, regions of statistically significant complementarity consist of several neighboring, often overlapping windows with a peak of the density of complementarity in the center (see, for example, Supplemental Fig. S1). The statistical analysis was performed independently for 18S and 28S rRNAs, producing two sets of  $P$ -values in which the  $q$ -values were evaluated.

## Evolutionary conservation of statistically significant sequence complementarity on rRNA

The regions of statistically significant sequence complementarity between rRNAs and 3' UTRs were tested for evolutionary conservation among all analyzed species. To do that, we first aligned rRNA sequences of all studied species using ClustalW (Blanco et al. 1994) and refined the resulting alignment using the SINA model of the rRNA multiple alignment (Pruesse et al. 2012). Alignments were carried out separately for protozoans and metazoans for the reasons explained in the second subsection of the Results section. Next, we identified the positions of nucleotides that formed regions of statistically significant complementarity (i.e., those with  $q$ -values  $< 0.05$ , where the blue curve dropped below the green line in the individual panels of Supplemental Fig. S1) in the individual species and mapped them onto the multiple sequence alignment of rRNA sequences. Because the nucleotide positions in the sequence alignment were aligned, the positions of nucleotides forming the regions of evolutionarily conserved statistically significant complementarity were aligned as well. Subsequently, we counted numbers of species having statistically significant regions of complementarity at the same position in the aligned rRNAs, and plotted these numbers against corresponding aligned positions in a form of a bar plot (Supplemental Fig. S1A, C, E, G). Hence, individual positions of the bars (on the horizontal axis) in our histograms indicate locations of statistically significant regions of complementarity with respect to their nucleotide positions in the given rRNA sequence. The height of the bars (on the vertical axis) then indicates how many species contain a statistically significant region of complementarity in a corresponding location in the rRNA sequence.

The level of a nonspecific biological and/or computational bias was estimated to identify biologically relevant bars in the obtained profiles defining evolutionarily conserved regions of complementarity. We repeated the whole procedure 100 times for each rRNA with randomly reshuffled rRNA sequences. Under this null model, 0.17%, 0.85%, 0.41%, and 0.16% of positions on randomized rRNA sequences had the height of their bars higher than four, five, six, and seven species for protozoan 18S rRNA, protozoan 28S rRNAs, metazoan 18S rRNA, and metazoan 28S rRNA, respectively, which set a cut-off line of biological relevance for each rRNA. By definition, the values of these cut-off lines indicate the limiting height of the bars above which less than one false positive could be expected. Thus, only the positions in the native sequences with bars higher than the cut-off lines can be considered as biologically relevant. We identified altogether five such positions (with  $P < 6 \times 10^{-5}$ , binomial distribution; i.e., highly significant) all of them occurring in the metazoan 18S rRNAs. Typical examples of the bar plots computed for randomized rRNAs are shown in Supplemental Figure S1B, D, F, H.

## SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

## ACKNOWLEDGMENTS

We are grateful to Alexandra Zinoviev and Tatyana V. Pestova for help and advice, Monica Yun Liu and Jiří Vohradský for critical

reading of the manuscript, and Jean-Paul Armache for providing us with optimized 2D structures of human and fly rRNAs. This work was supported by the Czech Science Foundation Grant P305/12/G034 and the Wellcome Trust Grant 090812/B/09/Z (both to L.S.V.), and the RVO Grant 68378050 (to M.K.).

**Author contributions:** J.P. initiated the study and performed the computational experiments and structural mapping. M.K. performed the statistics design. L.S.V. and A.H. performed the biological interpretation of the computational results. J.P., L.S.V., and M.K. wrote the manuscript.

Received January 25, 2016; accepted April 17, 2016.

## REFERENCES

- Anger AM, Armache JP, Berninghausen O, Habeck M, Subklewe M, Wilson DN, Beckmann R. 2013. Structures of the human and *Drosophila* 80S ribosome. *Nature* **497**: 80–85.
- Araujo PR, Yoon K, Ko D, Smith AD, Qiao M, Suresh U, Burns SC, Penalva LO. 2012. Before it gets started: regulating translation at the 5' UTR. *Comp Funct Genomics* **2012**: 475731.
- Barrett LW, Fletcher S, Wilton SD. 2012. Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cell Mol Life Sci* **69**: 3613–3634.
- Ben-Shem A, Garreau de Loubresse N, Melnikov S, Jenner L, Yusupova G, Yusupov M. 2011. The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science* **334**: 1524–1529.
- Blanco G, Rodicio MR, Puglia AM, Mendez C, Thompson CJ, Salas JA. 1994. Synthesis of ribosomal proteins during growth of *Streptomyces coelicolor*. *Mol Microbiol* **12**: 375–385.
- Cannone JJ, Subramanian S, Schnare MN, Collett JR, D'Souza LM, Du Y, Feng B, Lin N, Madabusi LV, Muller KM, et al. 2002. The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* **3**: 2.
- des Georges A, Hashem Y, Unbehaun A, Grassucci RA, Taylor D, Hellen CU, Pestova TV, Frank J. 2014. Structure of the mammalian ribosomal pre-termination complex associated with eRF1.eRF3.GDPNP. *Nucleic Acids Res* **42**: 3409–3418.
- Dimitrov RA, Zuker M. 2004. Prediction of hybridization and melting for double-stranded nucleic acids. *Biophys J* **87**: 215–226.
- Dresios J, Chappell SA, Zhou W, Mauro VP. 2006. An mRNA-rRNA base-pairing mechanism for translation initiation in eukaryotes. *Nat Struct Mol Biol* **13**: 30–34.
- Faghihi MA, Zhang M, Huang J, Modarresi F, Van der Brug MP, Nalls MA, Cookson MR, St-Laurent G III, Wahlestedt C. 2010. Evidence for natural antisense transcript-mediated inhibition of microRNA function. *Genome Biol* **11**: R56.
- Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, Weissman JS. 2003. Global analysis of protein expression in yeast. *Nature* **425**: 737–741.
- Grillo G, Turi A, Licciulli F, Mignone F, Liuni S, Banfi S, Gennarino VA, Horner DS, Pavesi G, Picardi E, et al. 2009. UTRdb and UTRsite (RELEASE 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. *Nucleic Acids Res* **38**: D75–D80.
- Guydosh NR, Green R. 2014. Dom34 rescues ribosomes in 3' untranslated regions. *Cell* **156**: 950–962.
- He Y, Vogelstein B, Velculescu VE, Papadopoulos N, Kinzler KW. 2008. The antisense transcriptomes of human cells. *Science* **322**: 1855–1857.
- Hussain T, Llacer JL, Fernandez IS, Munoz A, Martin-Marcos P, Savva CG, Lorsch JR, Hinnebusch AG, Ramakrishnan V. 2014. Structural changes enable start codon recognition by the eukaryotic translation initiation complex. *Cell* **159**: 597–607.
- Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. 2009. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**: 218–223.
- Kang MK, Han SJ. 2011. Post-transcriptional and post-translational regulation during mouse oocyte maturation. *BMB Rep* **44**: 147–157.
- Katayama S, Tomaru Y, Kasukawa T, Waki K, Nakanishi M, Nakamura M, Nishida H, Yap CC, Suzuki M, Kawai J, et al. 2005. Antisense transcription in the mammalian transcriptome. *Science* **309**: 1564–1566.
- Kosaka N, Yoshioka Y, Hagiwara K, Tominaga N, Katsuda T, Ochiya T. 2013. Trash or treasure: extracellular microRNAs and cell-to-cell communication. *Front Genet* **4**: 173.
- Llacer JL, Hussain T, Marler L, Aitken CE, Thakur A, Lorsch JR, Hinnebusch AG, Ramakrishnan V. 2015. Conformational differences between open and closed states of the eukaryotic translation initiation complex. *Mol Cell* **59**: 399–412.
- Luttermann C, Meyers G. 2009. The importance of inter- and intramolecular base pairing for translation reinitiation on a eukaryotic bicistronic mRNA. *Genes Dev* **23**: 331–344.
- Matveeva OV, Shabalina SA. 1993. Intermolecular mRNA-rRNA hybridization and the distribution of potential interaction regions in murine 18S rRNA. *Nucleic Acids Res* **21**: 1007–1011.
- Mauro VP, Edelman GM. 1997. rRNA-like sequences occur in diverse primary transcripts: implications for the control of gene expression. *Proc Natl Acad Sci* **94**: 422–427.
- Mauro VP, Edelman GM. 2002. The ribosome filter hypothesis. *Proc Natl Acad Sci* **99**: 12031–12036.
- Mauro VP, Edelman GM. 2007. The ribosome filter redux. *Cell Cycle* **6**: 2246–2251.
- Munzarová V, Pánek J, Gunišová S, Dányi I, Szamecz B, Valášek LS. 2011. Translation reinitiation relies on the interaction between eIF3a/TIF32 and progressively folded *cis*-acting mRNA elements preceding short uORFs. *PLoS Genet* **7**: e1002137.
- Nahvi A, Sudarsan N, Ebert MS, Zou X, Brown KL, Breaker RR. 2002. Genetic control by a metabolite binding mRNA. *Chem Biol* **9**: 1043.
- Namy O, Hatin I, Rousset JP. 2001. Impact of the six nucleotides downstream of the stop codon on translation termination. *EMBO Rep* **2**: 787–793.
- Namy O, Rousset JP, Naphtine S, Brierley I. 2004. Reprogrammed genetic decoding in cellular gene expression. *Mol Cell* **13**: 157–168.
- Pánek J, Kolar M, Vohradsky J, Shivaya Valasek L. 2013. An evolutionary conserved pattern of 18S rRNA sequence complementarity to mRNA 5' UTRs and its implications for eukaryotic gene translation regulation. *Nucleic Acids Res* **41**: 7625–7634.
- Passmore LA, Schmeing TM, Maag D, Applefield DJ, Acker MG, Algire MA, Lorsch JR, Ramakrishnan V. 2007. The eukaryotic translation initiation factors eIF1 and eIF1A induce an open conformation of the 40S ribosome. *Mol Cell* **26**: 41–50.
- Pisarev AV, Skabkin MA, Pisareva VP, Skabkina OV, Rakoton-drafara AM, Hentze MW, Hellen CUT, Pestova TV. 2010. The role of ABCE1 in eukaryotic posttermination ribosomal recycling. *Mol Cell* **37**: 196–210.
- Preis A, Heuer A, Barrio-Garcia C, Hauser A, Elyer DE, Berninghausen O, Green R, Becker T, Beckmann R. 2014. Cryoelectron microscopic structures of eukaryotic translation termination complexes containing eRF1-eRF3 or eRF1-ABCE1. *Cell Rep* **8**: 59–65.
- Pruesse E, Peplies J, Glockner FO. 2012. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**: 1823–1829.
- Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO. 2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**: D590–D596.
- Richardson LV, Richardson JP. 1996. Rho-dependent termination of transcription is governed primarily by the upstream Rho utilization (rut) sequences of a terminator. *J Biol Chem* **271**: 21597–21603.
- Shine J, Dalgarno L. 1975. Determinant of cistron specificity in bacterial ribosomes. *Nature* **254**: 34–38.

- Shoemaker CJ, Green R. 2011. Kinetic analysis reveals the ordered coupling of translation termination and ribosome recycling in yeast. *Proc Natl Acad Sci* **108**: E1392–E1398.
- Skabkin MA, Skabkina OV, Hellen CU, Pestova TV. 2013. Reinitiation and other unconventional posttermination events during eukaryotic translation. *Mol Cell* **51**: 249–264.
- Storey JD. 2003. The positive false discovery rate: a Bayesian interpretation and the q-value. *Ann Stat* **31**: 2013–2035.
- Szamecz B, Rutkai E, Cuchalova L, Munzarova V, Herrmannova A, Nielsen KH, Burela L, Hinnebusch AG, Valášek L. 2008. eIF3a cooperates with sequences 5' of uORF1 to promote resumption of scanning by post-termination ribosomes for reinitiation on GCN4 mRNA. *Genes Dev* **22**: 2414–2425.
- Tranque P, Hu MCY, Edelman GM, Mauro VP. 1998. rRNA complementarity within mRNAs: a possible basis for mRNA-ribosome interactions and translational control. *Proc Natl Acad Sci* **95**: 12238–12243.
- Wethmar K. 2014. The regulatory potential of upstream open reading frames in eukaryotic gene expression. *Wiley Interdiscip Rev RNA* **5**: 765–778.