Biochemical and Biophysical Research Communications 452 (2014) 66–71

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

CPuORF correlates with miRNA responsive elements on protein evolutionary rates

Wangxiong Hu [a,b,c,1], Tingzhang Wang [d,1], Yanmei Yang [e,1], Shu Zheng [b,c,*]

[a] College of Life Sciences, Zhejiang University, Hangzhou, Zhejiang 310058, China
[b] Cancer Institute, Second Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou, Zhejiang 310009, China
[c] Zhejiang-California International Nanosystems Institute, Zhejiang University, Hangzhou, Zhejiang 310058, China
[d] Zhejiang Institute of Microbiology, Hangzhou, Zhejiang 310012, China
[e] Key Laboratory of Reproductive and Genetics, Ministry of Education, Women's Hospital, Zhejiang University, Hangzhou, Zhejiang 310006, China

ARTICLE INFO

Article history:
Received 8 August 2014
Available online 19 August 2014

Keywords:
Correlation
Housekeeping genes
miRNA responsive elements
Upstream open reading frame

ABSTRACT

miRNA is increasingly being recognized as a key regulator of metabolism in animals. A wealth of evidence has suggested that miRNA mainly binds 3′ UTR of mRNA and modulates the cell activities via repressing the mRNA translation. However, as the translation initiates at 5′ UTR, *cis* elements like upstream open reading frame (uORF) resided in 5′ UTR may also affect the translation efficiency or elongation. In this study, we performed a systematic analysis of miRNA responsive elements (MREs) and uORF of the same transcript in three model organisms (human, mouse, and *Drosophila*). Intriguingly, we found that the 3′ UTR length grew with the complexity of species (human > mouse > *Drosophila*), in sharp contrast with the invariability of 5′ UTR. Additionally, MRE number correlated well with the 3′ UTR length, while uORF number showed a weak correlation with the 5′ UTR length. Further, we found that human genes with conserved peptide upstream open reading frame (CPuORF) tend to have more MREs and lower evolutionary rates, which provides new insights into the correlation between UTR properties and translational control in animals.

1. Introduction

The past decade has witnessed an explosion in the identification of microRNA (miRNA) as promising mediators in regulating animal versatile aspects [15]. miRNAs are short non-coding RNAs that range from 20 to 24 nucleotides (nt) and are derived from single-stranded precursors with stable hairpin structures [30,18]. It has long been proposed that miRNA responsive elements (MREs) in the majority of messenger RNAs (mRNAs) were located in the 3′ UTR in animals [24]. And miRNAs modulate the majority of mRNAs via base-pairing of the seed sequence to their targets, subsequently mediating translational inhibition and/or mRNA instability [1]. Despite intensive efforts, the molecular interplay between miRNA-mediated silencing complex (miRISC) and consequent translational repression remains unclear. Meijer et al. [21] propose that eIF4AII recruited by miRISCs may play an important role in translational repression. And the molecular mechanism underlying this is probably associate with hampering ribosome assembling and scanning in 5′ untranslated region (5′-UTR) [18,9,22]. Intriguingly, other *cis*-regulatory elements located in 5′-UTR like upstream open reading frame (uORF) may also affect ribosome scanning [26].

To date, uORF has been identified as a common feature in many eukaryotic mRNAs [32,14,13,3]. It often represses the accumulation of the protein level translated by primary mORF [26,20]. Thus, mutation that disrupt or introduce a uAUG would frequently accompany with serious biological consequences [31,23]. Both uORF and miRNA function at the translational level, which highlights the conceivable connection of translational regulation between these two elements. Aforementioned descriptions inspire us to explore the connection between uORF and miRNA in human (*Homo sapiens*), mouse (*Mus musculus*), and *Drosophila* (*Drosophila melanogaster*) for both well-annotated transcripts and miRNAs.

In this study, we systematically identified the uORFs and their corresponding MREs located in 5′ and 3′ UTR, respectively. Intriguingly, we found that MRE number correlated well with the 3′ UTR length, while uORF number showed a weak correlation with the 5′ UTR length. In parallel, we found that transcripts with conserved peptide upstream open reading frame (CPuORF) tend to have more MREs and evolve at a lower rate.

* Corresponding author at: Cancer Institute, Second Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou, Zhejiang 310009, China.
E-mail address: zhengshu@zju.edu.cn (S. Zheng).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.bbrc.2014.08.050
0006-291X/

2. Materials and methods

2.1. Databases for sequences

The pre-miRNA and mature miRNA sequences were retrieved from miRBase (Release 20). RefSeq transcripts for human (hg38), mouse (mm10) and *Drosophila* (dm3) were downloaded from the University of California, Santa Cruz (UCSC) Genome Browser Database.

2.2. uORFs identification

First, sequences <300 nt were discarded. Transcripts without definitive CDS region were also excluded. In the case of different transcripts generated from a single gene locus by alternative splicing, only the longest transcript was retained for further analysis. Sequences precede CDS were regarded as 5′ UTR and followed by uORF searching by custom Perl scripts. That is, uORF should contain start codon + at least one amino acid codon + stop codon, with the start codon located in 5′ UTR and stop codon not identical to downstream mORF. We defined uORF number as the number of distinct uORFs within a transcript, as uORFs may overlap but not in the same frame.

Then mORFs were used to search homologs between inter-species (human and mouse) and intra-species (human) by blastp with *E*-value $\leqslant 1e^{-5}$ and *E*-value $\leqslant 1e^{-10}$, respectively. As for the identification of CPuORF, blastp with *E*-value $\leqslant 1e^{-3}$ were used. In addition, we only considered CPuORFs with at least 11 amino acids due to the difficulty in determining their homology in shorter sequences.

2.3. Target prediction of miRNAs

Targets of miRNAs were predicted by TargetScan [11], miRanda [2], and RNAhybrid [25] with default parameters. To gain high-confidence results, only targets overlapped by at least two algorithms were kept for further analysis. And no mismatch was allowed between seed region and its potential targets. The number of regulatory miRNA responsive elements (MREs) within the 3′ UTR of a transcript was counted by above methods.

2.4. Statistics

All statistics mentioned in this study were performed in *R*. The number of nucleotide substitutions per synonymous site (Ks) and the number of nucleotide substitutions per nonsynonymous site (Ka) were obtained from Ensemble Genome Broswer (Ensemble Genes 75) by using the homologous gene pairs between mouse and human [10]. Evolutionary rate (ω) for protein is defined as the Ka/Ks ratio, where $\omega < 1$, $\omega = 1$, and $\omega > 1$ indicate purifying, neutral, and positive selection, respectively.

3. Results and discussion

3.1. Abundance of uORFs and MREs in human, mouse, and Drosophila transcripts

Consistent with the observation in other studies [19,13,3], we found that uORFs were abundantly existed in these three model organisms (Fig. 1A). Actually 54%, 44%, and 54% of transcripts in human, mouse, and *Drosophila* contained at least one uORF, respectively (Fig. 1B).

As for the identification of miRNA binding sites among these three species, three different algorithms (TargetScan, miRanda, and RNAhybrid) were used to seek for potential targets within 3′ UTR. To reduce false positives, we only considered MREs that could be overlapped by at least two algorithms. Results showed that >90% transcripts were putatively targeted by miRNAs irrespective of species. In addition, a large part of the transcripts harbored both uORF and MRE, that is, 52%, 39%, and 53% in human, mouse, and *Drosophila*, respectively (Fig. 1B). This data indicated that considerable transcripts subject to a combinational regulation at the translational level [29].

3.2. MRE number correlates well with 3′ UTR length

Interestingly, we found that the 3′ UTR length grew with the complexity of species (human > mouse > *Drosophila*), with the mean length at 1474 nt, 1195 nt, and 533 nt, respectively (Fig. 2A). The length of 5′ UTR, however, retained basically unchanged during the long-term evolution across species (mean length: human – 255 nt, mouse – 199 nt, *Drosophila* – 282 nt, Fig. 2A). Further in-depth analysis of MRE properties revealed that MRE number strongly correlated with the 3′ UTR length regardless of species ($r^2 > 0.9$, linear regression; $\rho > 0.95$, Pearson correlation coefficient, Fig. 2B). We thus hypothesized that miRNAs co-evolved with their targets as the miRNA number (miRBase 20) grew with the 3′ UTR length among these three organisms. Nevertheless, the correlation of uORFs number and 5′ UTR length displayed a much less dependent manner (Fig. 2C), suggesting that there is little association between uORF number and species characteristics.

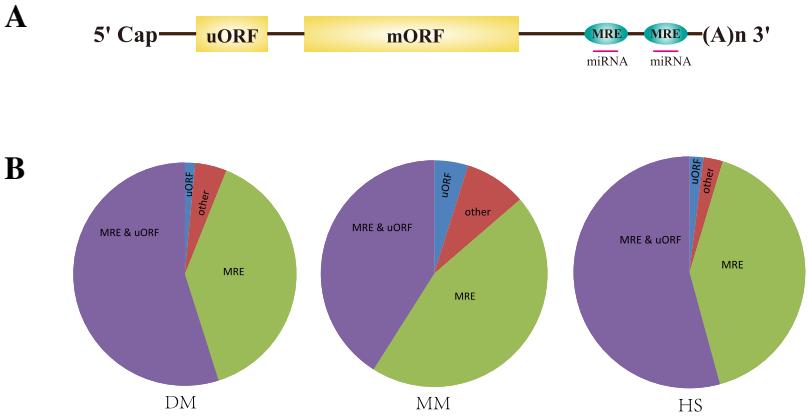

Fig. 1. Distribution of uORF and MRE in human, mouse, and *Drosophila* transcripts. (A) Schematic representation of mRNA transcript with uORF and MRE. uORF is defined by a start codon (AUG) in the 5′ UTR, at least one amino acid codon, and a stop codon preceding the end of downstream mORF. MRE is defined by seed pairing with the miRNA annotated in miRBase (Release 20). MREs overlapped by at least two softwares (TargetScan, miRanda, and RNAhybrid) were retained for further analysis. (B) Ratio of transcripts that harbored MRE-only, uORF-only, uORF & MRE, and others. DM, MM, and HS indicate *Drosophila*, mouse, and human, respectively.

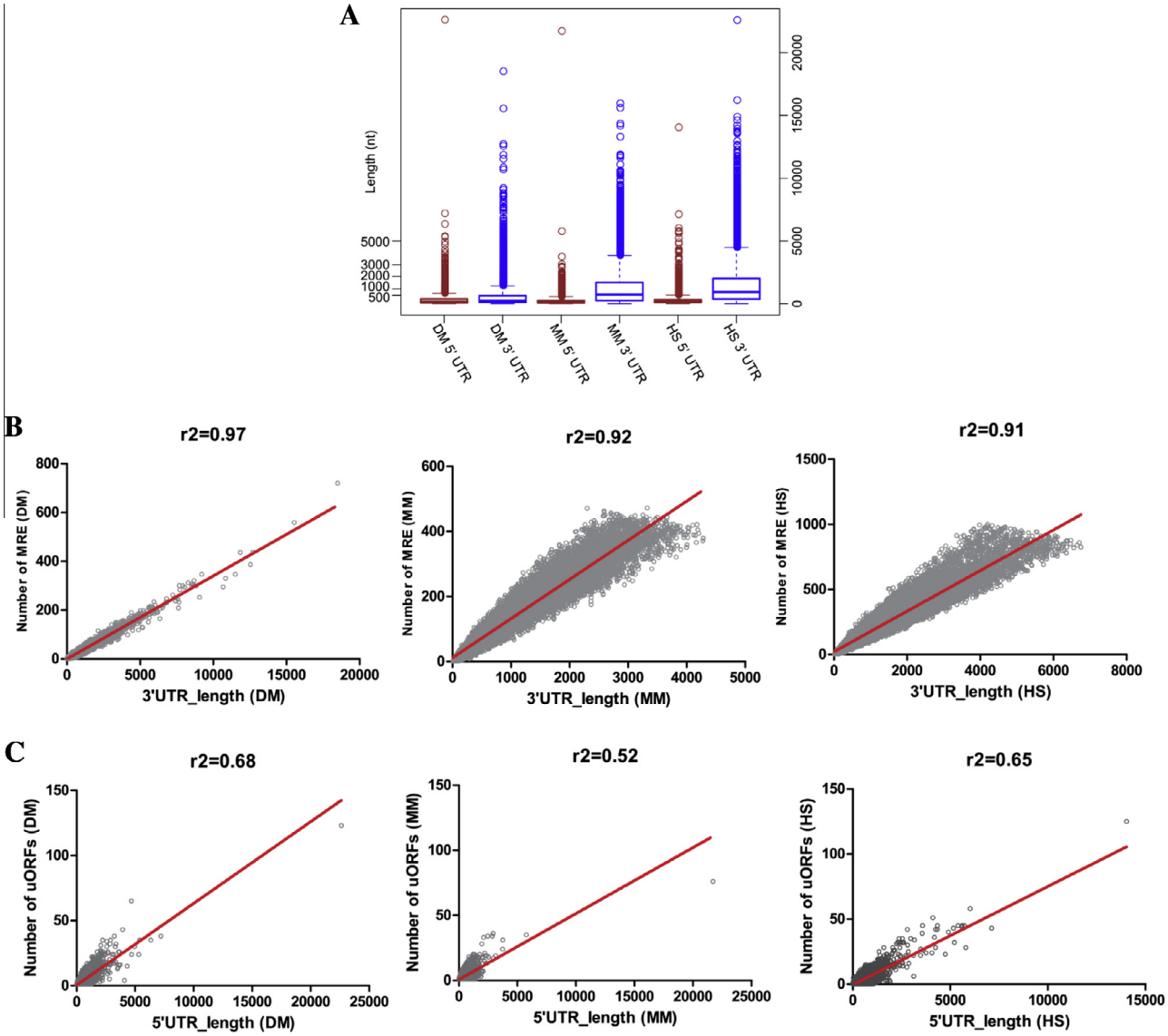

Fig. 2. Correlation of 3′ UTR length and MRE number as well as 5′ UTR length and uORF number. (A) Length distribution of 3′ UTR and 5′ UTR in *Drosophila*, mouse, and human. In striking contrast to invariability of 5′ UTR length, 3′ UTR length grew with the complexity of species (human > mouse > *Drosophila*). (B) Correlation of 3′ UTR length and MRE number in *Drosophila*, mouse, and human. (C) Correlation of 5′ UTR length and uORF number in *Drosophila*, mouse, and human. The red lines show the relation between the two quantities estimated by linear regression. (r^2 indicates goodness of fit). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Though both uORFs and MREs are abundant in these species, transcripts can mitigate their regulation via alternative splicing to shorten 3′ UTR or 5′ UTR in different tissues or specific conditions [27,17]. Gain and loss of uORF/MRE can thus play an essential role in modulating cellular regulatory networks.

3.3. MRE exhibit opposite correlation with protein evolutionary rates

In general, miRNA may tune their targets to make them achieve a more precise expression pattern [7]. Cheng et al. [5] have demonstrated that there is a significant negative correlation between the number of regulatory miRNAs and protein evolutionary rates in both human and mouse. Meanwhile, they find that another class of genes (housekeeping-HK genes, based on microarray data) possess shorter 3′ UTRs, slower evolutionary rates, and less number of regulatory miRNAs. They argue that HK genes are more likely to have shorter 3′ UTRs to avoid miRNA modulation. We thus divided the human genes into HK, tissue-specific (TS) and other gene groups by using RNA-seq [8] and microarray [4] data. Results showed that TS genes possessed shorter 3′ UTRs and much less MREs (Fig. 3A and B; $P < 0.001$, Mann Whitney test). Their evolutionary rate (ω), was higher than other gene groups that regulated by more miRNAs (Fig. 3C; $P < 0.001$, Mann Whitney test), namely a significant negative correlation between the number of regulatory miRNAs and protein evolutionary rates. To further check for the opposite correlation between MREs and protein evolutionary rates, we hypothesized that lower ω should be corresponding to more MREs in a linear manner. To support this assumption, we investigated the 3′ UTRs length and MRE number with different ω values. As expected, 3′ UTRs length and MRE number decreased linearly with ω increased (Fig. 3D and E). Together, these results suggested that MRE negatively correlated with protein evolutionary rates and there appear to be other means affecting the evolutionary rates of specific gene groups like HK genes. More recently, Chuang and Chiang [6] have proposed that promoter/gene body methylation, miRNA as well as TF regulation are all involved in determining the rate of mammalian protein evolution. As such, we further tested whether uORF number may impose an effect beyond MRE.

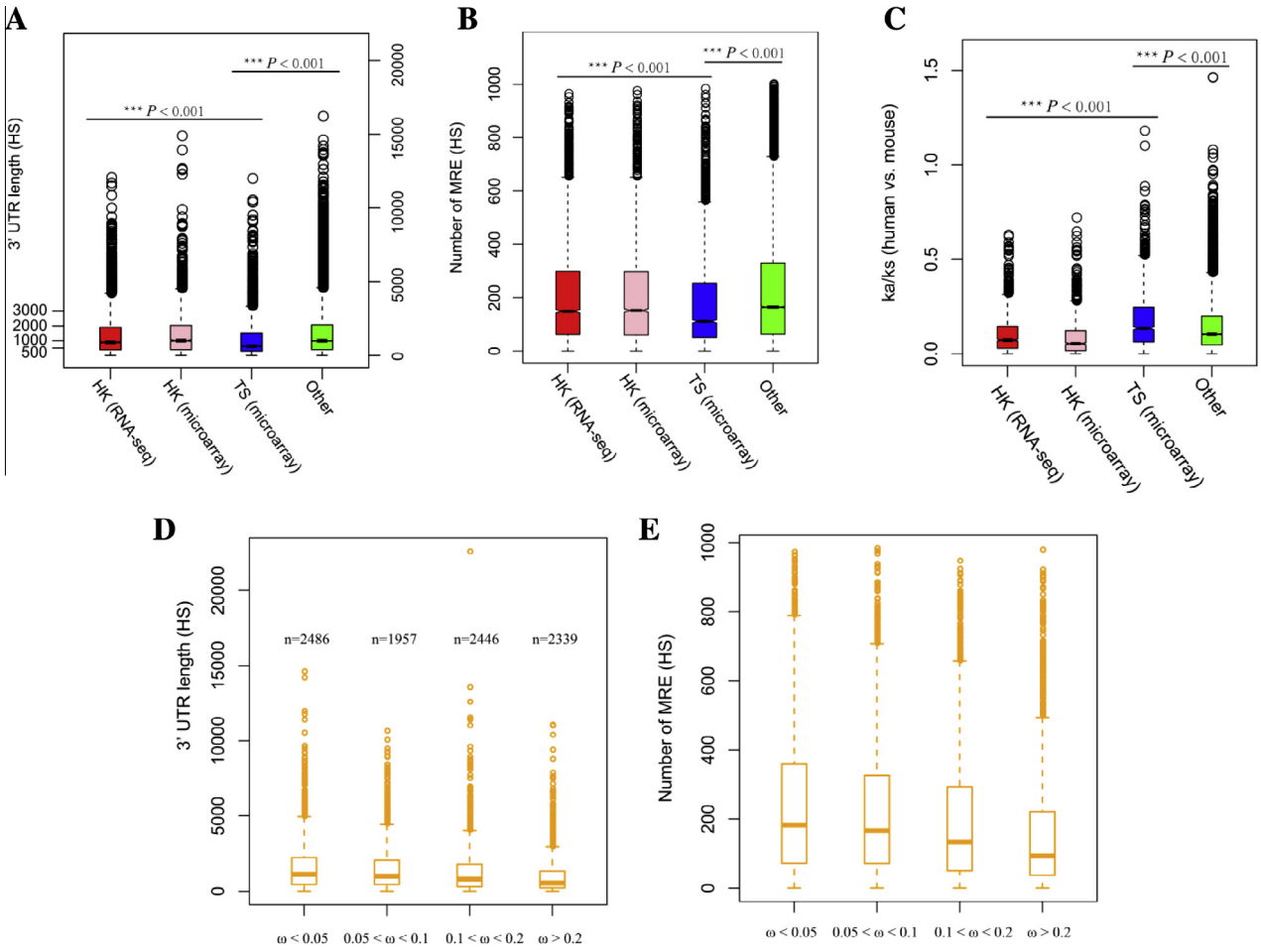

Fig. 3. Correlation of MRE number and protein evolutionary rates. To illuminate the interaction of MRE on protein evolution, we divided the genes into HK, TS and others. For (A–C), the relation between protein evolutionary rate and 3′ UTR length as well as MRE number. Evidently genes evolving slower (HK, Other) harbored more MREs and longer 3′ UTR length relative to TS genes. As for (D and E), comparison of protein evolutionary rate and 3′ UTR length as well as MRE number in four gene groups of similar size. Again, data showed that genes evolving slower harbored more MREs and longer 3′ UTR length. (***indicates $P < 0.001$, Mann Whitney test).

We compared the uORF number distribution between HK and TS gene groups and observed no significant difference ($P > 0.05$, Mann Whitney test). Furthermore, the number of HK genes that harbored uORF or not, also exhibited no significant difference between HK and TS gene groups ($P > 0.05$, χ^2 test). Collectively, these results argued that selection was likely involved in shaping 5′ UTR properties since 5′ UTR length retained virtually unchanged across species and the uORF number exhibited a weak association with the 5′ UTR length. In this regard, we systematically identified the CPuORF between human and mouse. Strikingly, 2284 CPuORFs have been identified that covered 1968 genes in human.

3.4. Genes with CPuORF tend to have more MREs and lower evolutionary rates

To elucidate the underlying interplay between CPuORF and MRE, we looked into whether transcripts with CPuORF have more MREs or not. Intriguingly, we observed that transcripts harbored CPuORF tend to have longer 3′ UTR length and more MREs (Fig. 4A and B; $P < 0.001$, Mann Whitney test). It allowed us to hypothesize that CPuORF embedded in 5′ UTR had some association with miRNA activities. We also examined the correlations between CPuORF and downstream mORF evolutionary rates. Intriguingly, we found that proteins associated with CPuORF displayed significant lower evolutionary rates than others (median: 0.058 vs. 0.109; mean: 0.091 vs. 0.150), which suggested that CPuORF could contribute to the transcripts obtaining more MREs and evolving slower (Fig. 4C, $P < 0.001$, Mann Whitney test).

Retention of both CPuORF and MRE indicates that the gene suffers from a more strict translation repression or control. The underlying mechanism is on the basis of characterization of their molecular functions. Co-expressed genes usually share similar regulatory element, and probably perform related biological functions [12]. However, prior to our work, the function of genes with CPuORF has not been evaluated. We thus investigated the functions of human genes harbored CPuORF and found an enrichment in transcription activities by using DAVID [16]. This was in full agreement with the observation that genes coding for transcription factors have significantly longer 3′ UTRs than ribosomal genes [28]. Still, whether CPuORF complement functionally with MRE or alternatively, MRE complement CPuORF, is poorly understood. Given CPuORF is rather limited (<7%, human vs. mouse) and many more conserved MREs (>75%, human vs. mouse, based on seed sequences) was observed. It is tempting to believe that CPuORF complement functionally with MRE (Fig. 4D). As the assumption can be further corroborated by many more evolutionary conserved MREs that could be traced back to *Drosophila*, while the conservation of corresponding CPuORF was confined to mammals.

In this study, we simultaneously analyzed the correlation between CPuORF and MRE of the same transcripts. Actually, we cannot rule out the possibility that other mechanisms can modu-

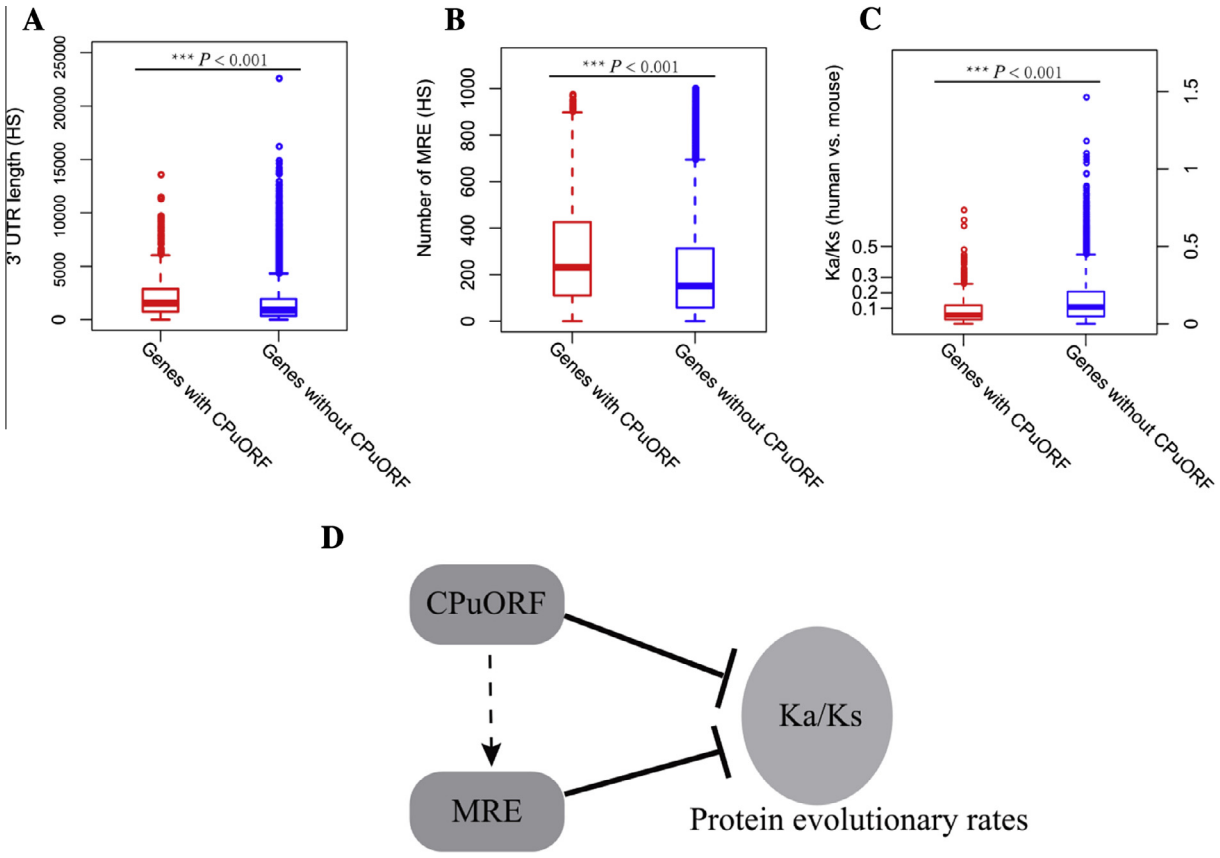

Fig. 4. Correlation of CPuORF and protein evolutionary rates. It was evident that genes with CPuORF harbored longer 3′ UTR length (A), more MREs (B), and evolving slower (C). (***indicates $P < 0.001$, Mann Whitney test). (D) Schematic representation of the regulatory effect of CPuORF and MRE upon protein evolutionary rates. The nail shapes represent negative regulation and the dashed arrow indicates positive regulation. This scenario suggests that these two factors have a mutual impact on protein evolutionary rates.

late the translation of mRNA. Additionally, a combination of different regulatory measures enables the organism to obtain a more robust network.

Acknowledgments

This study was supported by Project of the Science and Technology Department of Zhejiang Province of China under Grant Nos. 2014F30033 and 2014F50012.

References

[1] D.P. Bartel, MicroRNAs: target recognition and regulatory functions, Cell 136 (2) (2009) 215–233.
[2] D. Betel, A. Koppal, et al., Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites, Genome Biol. 11 (8) (2010) R90.
[3] S.E. Calvo, D.J. Pagliarini, et al., Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans, Proc. Natl. Acad. Sci. U.S.A. 106 (18) (2009) 7507–7512.
[4] C.W. Chang, W.C. Cheng, et al., Identification of human housekeeping genes and tissue-selective genes by microarray meta-analysis, PLoS ONE 6 (7) (2011) e22859.
[5] C. Cheng, N. Bhardwaj, et al., The relationship between the evolution of microRNA targets and the length of their UTRs, BMC Genomics 10 (2009) 431.
[6] T.J. Chuang, T.W. Chiang, Impacts of pre-transcriptional DNA methylation, transcriptional transcription factor and post-transcriptional microRNA regulations on protein evolutionary rate, Genome Biol. Evol. (2014).
[7] Q. Cui, Z. Yu, et al., MicroRNA regulation and interspecific variation of gene expression, Trends Genet. 23 (8) (2007) 372–375.
[8] E. Eisenberg, E.Y. Levanon, Human housekeeping genes, revisited, Trends Genet. 29 (10) (2013) 569–574.
[9] M.R. Fabian, N. Sonenberg, The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC, Nat. Struct. Mol. Biol. 19 (6) (2012) 586–593.
[10] P. Flicek, M.R. Amode, et al., Ensembl 2014, Nucleic Acids Res. 42 (Database issue) (2014) D749–D755.
[11] R.C. Friedman, K.K. Farh, et al., Most mammalian mRNAs are conserved targets of microRNAs, Genome Res. 19 (1) (2009) 92–105.
[12] M.P. Gustin, C.Z. Paultre, et al., Functional meta-analysis of double connectivity in gene coexpression networks in mammals, Physiol. Genomics 34 (1) (2008) 34–41.
[13] C.A. Hayden, G. Bosco, Comparative genomic analysis of novel conserved peptide upstream open reading frames in *Drosophila melanogaster* and other dipteran species, BMC Genomics 9 (2008) 61.
[14] C.A. Hayden, R.A. Jorgensen, Identification of novel conserved peptide uORF homology groups in Arabidopsis and rice reveals ancient eukaryotic origin of select groups and preferential association with transcription factor-encoding genes, BMC Biol. 5 (2007) 32.
[15] L. He, G.J. Hannon, MicroRNAs: small RNAs with a big role in gene regulation, Nat. Rev. Genet. 5 (7) (2004) 522–531.
[16] W. Huang da, B.T. Sherman, et al., Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, Nat. Protoc. 4 (1) (2009) 44–57.
[17] A.R. Kornblihtt, I.E. Schor, et al., Alternative splicing: a pivotal step between eukaryotic transcription and translation, Nat. Rev. Mol. Cell Biol. 14 (3) (2013) 153–165.
[18] J. Krol, I. Loedige, et al., The widespread regulation of microRNA biogenesis, function and decay, Nat. Rev. Genet. 11 (9) (2010) 597–610.
[19] M. Matsui, N. Yachie, et al., Bioinformatic analysis of post-transcriptional regulation by uORF in human and mouse, FEBS Lett. 581 (22) (2007) 4184–4188.
[20] J. Medenbach, M. Seiler, et al., Translational control via protein-regulated upstream open reading frames, Cell 145 (6) (2011) 902–913.
[21] H.A. Meijer, Y.W. Kong, et al., Translational repression and eIF4A2 activity are critical for microRNA-mediated gene regulation, Science 340 (6128) (2013) 82–85.
[22] F. Moretti, C. Kaiser, et al., PABP and the poly(A) tail augment microRNA repression by facilitated miRISC binding, Nat. Struct. Mol. Biol. 19 (6) (2012) 603–608.
[23] G. Occhi, D. Regazzo, et al., A novel mutation in the upstream open reading frame of the CDKN1B gene causes a MEN4 phenotype, PLoS Genet. 9 (3) (2013) e1003350.

[24] A.E. Pasquinelli, MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship, Nat. Rev. Genet. 13 (4) (2012) 271–282.
[25] M. Rehmsmeier, P. Steffen, et al., Fast and effective prediction of microRNA/target duplexes, RNA 10 (10) (2004) 1507–1517.
[26] M.S. Sachs, A.P. Geballe, Downstream control of upstream open reading frames, Genes Dev. 20 (8) (2006) 915–921.
[27] R. Sandberg, J.R. Neilson, et al., Proliferating cells express mRNAs with shortened 3′ untranslated regions and fewer microRNA target sites, Science 320 (5883) (2008) 1643–1647.
[28] A. Stark, J. Brennecke, et al., Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3′ UTR evolution, Cell 123 (6) (2005) 1133–1146.
[29] J.N. Vaughn, S.R. Ellingson, et al., Known and novel post-transcriptional regulatory sequences are conserved across plant families, RNA 18 (3) (2012) 368–384.
[30] O. Voinnet, Origin, biogenesis, and activity of plant microRNAs, Cell 136 (4) (2009) 669–687.
[31] Y. Wen, Y. Liu, et al., Loss-of-function mutations of an inhibitory upstream ORF in the human hairless transcript cause Marie Unna hereditary hypotrichosis, Nat. Genet. 41 (2) (2009) 228–233.
[32] Z. Zhang, F.S. Dietrich, Identification and characterization of upstream open reading frames (uORF) in the 5′ untranslated regions (UTR) of genes in *Saccharomyces cerevisiae*, Curr. Genet. 48 (2) (2005) 77–87.